ACTA UNIVERSITATIS CAROLINAE

AUC **GEOGRAPHICA**

60
2/2025

CHARLES UNIVERSITY • KAROLINUM PRESS

ISSN 0300-5402 (Print)
ISSN 2336-1980 (Online)

Sixty years on a journey of discovery: Editorial for the anniversary of AUC Geographica

Luděk Šídlo

Vice-Dean for the Geographical Section, Faculty of Science, Charles University

Dear readers,

It is with both pride and gratitude that we welcome you to the jubilee issue of our journal, which, incredibly, celebrates its sixtieth anniversary this year. Established in 1966, AUC Geographica has for decades served as a platform for sharing scientific knowledge and has become an indispensable part of the Czech and Slovak geography scene. This year provides an opportunity not only to celebrate but also to reflect on the journey the journal has travelled and the direction it will take in the future.

The beginnings of AUC Geographica date back to 1966 when the first issue was compiled from contributions from the Czech-Polish Geographical Seminar held at Charles University in Prague. The original intention was to showcase the diversity of geographical research and to provide a publishing space for domestic geographers – from renowned professors to novice researchers – and to support their international involvement. This approach has guided the journal since its early years when articles were published in various languages and copies were exchanged with foreign journals, which served to significantly enrich the resources of our geographical library. Subsequently, foreign authors began to contribute to the journal, which led to the transition to publishing in English from the 1990s.

The path from the first issues of the journal to the present has been full of changes. The original name "Acta Universitatis Carolinae, Geographica" was shortened to today's AUC Geographica in 2011. Concurrently, the journal was comprehensively redesigned, as described in detail by Vilímek (2011). However, this evolution primarily reflected the dynamics of the scientific field and efforts to remain relevant. The journal has managed to adapt to changing trends in geography – from traditional directions to modern methods and paradigms. Space remains open in the journal for a wide range of geographical topics – from physical geography, which examines the dynamics of natural processes, through human and social geography, which help us to understand societal changes, to the geography of regional development and demography and the study of advanced geoinformatics methods. These topics are often explored via a multidisciplinary approach, which is essential in terms of accumulating contemporary scientific knowledge.

Efforts towards international recognition and the ability to innovate culminated in the journal's inclusion in a number of prestigious international databases, including Scopus (listed since 1975) and the Web of Science (listed since 2017), in which the journal has gained a valuable impact factor since 2022. It ranks in the upper half of similar journals in the Central European region in terms of the citation count of its articles, with the number of citations increasing over time (Fig. 1). Moreover, articles that feature in the journal are widely cited not only in domestic periodicals but also in prestigious foreign publications (Fig. 2), all of which confirms the quality and relevance of the journal's content and demonstrates that the geographical research published stands up to global competition. The journal's open access format and the use of DOI significantly enhance its accessibility for the professional community.

On this special occasion, we would like to express our deep gratitude to everyone who has helped to ensure the journal's continuing success. We thank the authors for their original and inspiring contributions, the reviewers for their thorough and professional approach, which forms one of the pillars of the quality of the journal's output, and, naturally, past and present editorial boards and all those who were involved in the journal's management in past years (for more details, see Vilímek 2015). Their vision and determination laid the foundations upon which the journal stands today.

Looking to the future, the journal aims both to maintain its high quality and to continue to develop.

Šídlo, L. (2025): Sixty years on a journey of discovery: Editorial for the anniversary of AUC Geographica. AUC Geographica 60(2), 153–154
https://doi.org/10.14712/23361980.2025.17

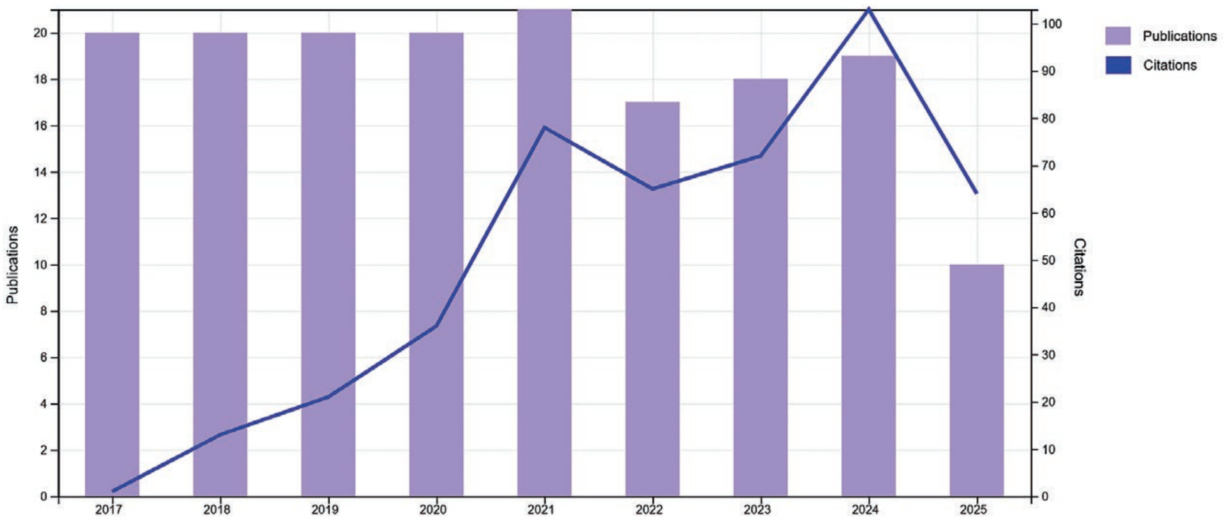

Fig. 1 Number of publications and citations in the AUC Geographica journal on the Web of Science (as of 8 September 2025).
Source: Web of Science Core Collection.

Fig. 2 Journals that most frequently cite articles from AUC Geographica published between 2015 and 2025 (as of 8 September 2025).
Note: The citations were tracked only for the "Article" document type. The 15 most frequent titles are displayed, which account for approximately 27% of the total number of citations. Only 16% of all the citations were from authors at Charles University, and 34% from Czech authors.
Source: Web of Science Core Collection.

We aim to focus on strengthening international cooperation, supporting multidisciplinary research, and reflecting global geographical challenges (e.g. Brodský et al. 2024). AUC Geographica must remain a platform for innovation and critical discussion for renowned geographers, as well as a publication that continues to support the younger generation of scientists and to offer them encouragement to enter the publishing process. The journal should remain a forum for the exchange of ideas and the foundation upon which the future of geographical science is built.

In conclusion, we wish AUC Geographica, one of the flagship journals published under the umbrella of Charles University, specifically the Faculty of Science, many more successful and inspiring years and we look forward to the next chapters of its continued successful development.

References

Brodský, L., Bollati, I. M., Nüsser, M. (2024): Preface. AUC Geographica 59(2), 161–162, https://doi.org/10.14712/23361980.2024.20.
Vilímek, V. (2011): Editorial. AUC Geographica 46(1), 5.
Vilímek, V. (2015): The 50th anniversary of AUC Geographica. Editorial AUC Geographica 50(1), 5, https://doi.org/10.14712/23361980.2015.82.

Demographic ageing through alternative indicators in the NUTS 2 regions of Czechia between 2003 and 2023

Filip Čábela, Luděk Šídlo*

Charles University, Faculty of Science, Department of Demography and Geodemography, Czechia
* Corresponding author: ludek.sidlo@natur.cuni.cz

ABSTRACT
The main objective of this article is to compare standard and alternative methodologies in measuring demographic ageing at the regional level in Czechia. We apply different indicators, such as the proportion of elderly people, the ageing index, the old-age dependency ratio, the pensioners to workers ratio, and real elderly dependency ratio, at the NUTS 2 regions of Czechia between 2003 and 2023. The concept of prospective age and linear interpolation method were also used in this article. The analysis revealed a continuing increase in the regional heterogeneity of Czechia in demographic ageing, as well as significant differences in the level of ageing between standard and alternative methodological approaches. When using alternative indicators, the North-west region can be considered the oldest, while the Prague region appears to be the youngest. It should also be noted that alternative indicators exhibit higher sensitivity to external mortality influences.

KEYWORDS
demographic ageing; NUTS 2 regions; Czechia; alternative indicators; prospective age

Received: 30 April 2025
Accepted: 13 August 2025
Published online: 9 September 2025

Čábela, F., Šídlo, L. (2025): Demographic ageing through alternative indicators in the NUTS 2 regions of Czechia between 2003 and 2023. AUC Geographica 60(2), 155–170
https://doi.org/10.14712/23361980.2025.13

1. Introduction

The number of older people has increased in recent years in almost every society in the world, (e.g. Gregory and Patuelli 2013; Sanderson and Scherbov 2013; Spijker and MacInnes 2013). This complex and multidimensional phenomenon is known as demographic ageing, a topic that has become one of the most intensely discussed issues of the 21st century. Although this phenomenon is unprecedented in human history, it should be perceived as a natural aspect of human development, which is closely linked to changes in the age and sex-age structures of the population. But why is the professional community so concerned with the study of demographic ageing and population structures? The study of the sex-age structure has the potential to assist experts in approximating the historical development of the population and to create a basis for estimating future developments. The early identification of potential social and economic challenges based on expected changes in the sex-age structure of the population plays an important role in terms of the design of policy measures, especially with concern to the social, health and economic systems of regions and the labour market (e.g. labour shortages, the participation of seniors etc.) (Siegel 2002).

The development of the population and demographic ageing are affected primarily by three basic factors, the first of which is the birth rate, i.e. fertility, and primarily unrealised fertility. The long-term decline in realised fertility, which has resulted in the stabilisation of the level of fertility at low values, has been attributed primarily to changes in family life and delayed parenthood. The concept of the second demographic transition (van de Kaa 1997; Lesthaeghe 2010) attempts to explain the declining fertility issue. However, the number of live births is related not only to the level of fertility but is also influenced by the number of women of reproductive age. The above-mentioned processes together influence demographic ageing from the bottom of the age pyramid upwards.

The second factor is referred to as the mortality component. Demographic ageing is influenced by increasing life expectancy both at birth and in old age. According to Omran (1971), the improvement in mortality rates is the result of a reduction in, or the disappearance of, mortality associated with certain causes of death, which is closely related to his theory of epidemiological transition. In contrast to this claim, the theory of morbidity expansion (Gruenberg 1977) claims that improving the health status of the population and increasing the quality of human life is not always associated with improving mortality rates. However, the increase in life expectancy leads in many ways to the deepening of the demographic ageing process, in which case we consider ageing from the top of the age pyramid downwards.

The last, and least stable, factor that impacts population ageing comprises migration, one of the important specific features of which concerns variations in its impact on the composition of the population in proportion to the size of the population of the monitored area. According to Eurostat (2025), immigrants have an average median age that is lower than the native population, which results in the "rejuvenation" of target regions, or in other words cohort turnover effect (the ability to compensate for the deficit of the working-age population) (Ghio et al. 2022). On the other hand, a corresponding decrease in the representation of, especially, working age persons can be expected in regions that experience emigration, which inevitably impacts the population ageing process. We can name it ageing (or rejuvenation) from the middle of the age pyramid. Today, migration is becoming, or has already become, an integral part of the reproduction dynamics of many European regions (Šídlo et al. 2020a) and it can be assumed that its importance in terms of human reproduction and population growth will continue to increase.

The initial age structure undeniably influences the trajectory of demographic ageing. Its imbalances, stemming from the historical development of the observed population, may either accelerate or slow down the process of demographic ageing (through the shifting of various cohorts into older ages). This process is referred to, as previously mentioned, as the cohort turnover effect (Ghio et al. 2022).

Most of the methods currently applied to analyse demographic ageing continue to be based on a set of multiple standard indicators (the ageing index, old-age dependency ratio, average age, etc.), which, however useful, provide only a limited and partially distorted view of the ageing issue. The amount of research on demographic ageing increased significantly in the first decade of the 21st century; however, despite the increasing interest in the topic, no significant changes have been applied to the approaches and analytical tools employed to measure the level of demographic ageing (Sanderson and Scherbov 2007).

Most of the analyses at the time (as well as mostly today) were based on the afore-mentioned set of indicators with the consideration of the number of years lived. The authors of standard (retrospective/classic) studies apply a fixed value of old age as determined according to the time of the lived (chronological) age (e.g. 65 years), which results in a partially limited and distorted view of the ageing issue. A further disadvantage of the use of these indicators concerns the fact that they remain unchanged in time and space, which results in non-objective comparisons of populations with different mortality rates and, potentially, in distorted conclusions (Sanderson and Scherbov 2007, 2013). Several authors have attempted to contribute to solving this problem, mainly over the last twenty years, by applying the so-called prospective age concept, according to which the age limit of old age is based on the same remaining life expectancy (see the following chapter). In summary, we can consider two basic approaches to the analysis of demographic ageing: standard (retrospective), which expresses the number of years that a given person has already lived, and alternative (including prospective indicators), which is understood as the number of years that a given person probably has left to live (Sanderson and Scherbov 2007).

This article outlines and applies both main approaches to demographic ageing and, further, seeks to complement them with potential alternative indicators (e.g. Spijker 2015), thus providing an interesting and unique approach to the development of population ageing with particular reference to the NUTS 2 regional level in Czechia between 2003 and 2023. The first aim of this article is to highlight the regional differences in the level of demographic ageing between the NUTS 2 regions of Czechia. The second aim is to analyse and compare progress in the level of demographic ageing by selected indicators (e.g. the proportion of elderly people, the ageing index, the old-age dependency ratio, the pensioners to workers ratio and the real elderly dependency ratio) of standard and alternative (including prospective indicators) methodologies during the studied period and simultaneously focusing on the effect of Covid-19 pandemic.

1.1 The concept of prospective age

Demographic ageing can be described as a complex phenomenon that affects almost every aspect of human life. The biggest "problem" in the analysis of demographic ageing concerns the determination of the boundary, threshold or limit of old age, which is related primarily to the continuous increase in life expectancy. In most cases (e.g. Lutz et al. 2008a, 2008b; Šídlo et al. 2020a), the age limit (following the standard/chronological approach associated with standard indicators) is set at 65 years. However, determining a fixed age limit based on the chronological age has a number of disadvantages. Differences in life expectancy at age 65 vary considerably over time and between populations; thus, the comparison of persons of the same chronological age across time and space is both problematic and even confusing (Sanderson and Scherbov 2008; Šídlo et al. 2019; Šprocha and Ďurček 2019), as confirmed by our views of the world around us. For example, it is becoming increasingly common to see people aged 65 and older living independent and active lifestyles, and a significant number of them prefer to remain active in the labour market.

The potential for adopting an alternative approach to the analysis and calculation of demographic ageing indicators was first mentioned by Norman B. Ryder in the 1970s in his paper on stable populations (1975). He suggested replacing the set age limit (65 years) with a tabular age (i.e. using life tables) at which a given population has 10 years left to live based on his view of chronological age, which loses its meaning for adults (in his opinion, it is valid only from birth to adulthood). Although it was a relatively revolutionary idea, his proposal came at a time when demographic ageing was not particularly relevant and was, therefore, of no real interest to either the professional or the lay public. Nevertheless, Ryder's ideas were gradually adopted by several other authors (e.g. Siegel and Davidson 1984; Seigel 1993), who proposed a further modification, i.e. changing the age limit to the age at which the tabular population has 15 years left to live.

However, Warren Sanderson and Sergei Scherbov (2005, 2007, 2008, 2010, 2013, 2016) made the most significant contribution to following up on Ryder's suggestions via the determination of new approaches to the analysis of demographic ageing via the concept of prospective age. The authors' consideration and subsequent modification of the afore-mentioned standard (retrospective) indicators led to the creation of a completely new approach applying prospective indicators, which consider information on changes in life expectancy over time. Bloom, Canning and Fink (2010) provided support for this approach, arguing that, due to increasing life expectancy, people are spending more time in retirement than before. Sanderson and Scherbov (2005) stated that it is important to take into account changes in life expectancy values over time, especially with concern to the older part of the population. Today's older populations differ from past older populations in terms of almost all the applied key characteristics. Put simply, this means that a person aged 65 today has different options and perspectives (e.g. concerning healthcare, disease prevention, the environment, education) than had a person of the same age in 1921, 1963 or 1998, or will have in 2050. According to Klapková et al. (2016), people live their last years of life at increasingly older ages, which acts to significantly affect their quality of life during these years. This appears to indicate that societies today should focus more on the number of years people have left to live than on the number of years they have already lived. Standard indicators based on the number of years lived do not take into account significant regional, national or gender-based differences (Šídlo et al. 2020b), whereas the concept of prospective age can be applied both to the analysis of changes over time and the analysis of differences between individual regions (Šprocha et al. 2018; Šprocha and Ďurček 2019; Gietel-Basten et al. 2020).

Just as the boundaries of entry into the final stage of life are changing, on the opposite side of the age spectrum the age of entry to work is also changing; today, the age of 15 no longer corresponds to reality in most cases due to the lengthening of the period of education. Participation in the labour market for people between the ages of 15 and 19 is relatively rare; therefore, the age of 20 is increasingly being considered and applied as the age of the commencement of economic activity, although this is still a not generalised approach.

A 2007 study by Sanderson and Scherbov assigned each person two ages (retrospective/chronological and prospective/remaining), which represented different aspects of old age, while complementing each other. The application of the dual age principle was not new and, indeed, had been applied in practice for some time for example in the healthcare sector, where concerning selected procedures related to the musculoskeletal system (total endoprostheses), doctors take into account whether a replacement will significantly increase the number of years of mobility of the patient (Klapková et al. 2016).

In recent decades, demographic ageing has increasingly been studied from the economic perspective, with the focus on its impacts on national economic systems (e.g. income growth or fiscal sustainability and stability) (Cuaresma et al. 2014; Bloom et al. 2010). Holzmann (2013), who examined the economic determinants of demographic ageing in depth, called for the reshaping of existing indicators and identified the key factors involved in the reform of pension systems.

The importance of the study of demographic ageing is confirmed by increasingly frequent concerns about the sustainability and stability of national social and pension systems, which are clearly justified. However, the majority of concerns are based on chronological (standard) indicators based on a fixed old age value (Klapková et al. 2016). Seniors are living longer today, and studies indicate that life expectancy will continue to increase in the future. Moreover, the health of older people is improving due to increasing numbers of seniors adopting an active ageing approach in both their leisure and work activities (Šídlo et al. 2020b).

The above examples highlight the importance of determining an alternative approach to the analysis of demographic ageing. In addition to Sanderson and Scherbov, the issue is addressed by Spijker (2023, 2022, 2015); Skirbekk et al. (2019); Riffe (2015); Spijker et al. (2014) and Spijker and MacInnes (2013). Their publications approach demographic ageing applying previously established standard, as well as alternative (including prospective) indicators, related to, for example, employment, the health status of the population and time-to-death (TTD). The main idea behind these new alternative indicators is to improve the added value of the indicators that are entered into the calculation process (related not only to the age structure, but for example, to those in paid employment [regardless of age], to all those who receive a pension, to the number of years spent in health or with a disability, or to economic indicators [e.g. GDP], etc.). One of the most important directions in terms of alternative indicators of demographic ageing concerns their connection with the influence of healthcare on the lives of individuals. The analysis of demographic ageing through alternative indicators (e.g. a constant prospective age) has the potential to provide a higher informative value (Čábela and Šídlo 2024); moreover, it also has the potential for use in other, previously unconsidered, areas.

2. Data and methods

This paper bases the analysis of demographic ageing on both standard (chronological/retrospective) indicators and their alternative equivalent (including prospective) indicators, which adapt to changes in life expectancy values. The basic and most important foundation stone for calculating most of the alternative indicators concerns the determination of the prospective age, which is understood as "*the age that is assigned to a given population in a given year based on the same remaining life expectancy in the reference year (and population)*" (Klapková et al. 2016, p. 131). According to Sanderson and Scherbov (2005), the prospective age can be defined as the number of years that a given person/population still has ahead of them according to the relevant mortality rates (i.e. the number of years of life remaining until the probable age of death). Spijker (2015, p. 6) describes prospective age as "*the age of a person in the standard life table who has the same remaining life expectancy as the person of interest*". The exact prospective age needed for the calculations can be estimated applying the linear interpolation method according to the following formula (see e.g. Ježek 2016, p. 20):

$$x = x_0 + (z - z_0)\frac{x_1 - x_0}{z_1 - z_0}$$

where the quantity x refers to the age and the quantity z denotes the remaining average life expectancy at age x. The designation x_0 subsequently denotes the age at which the remaining life expectancy is higher than the value that we wish to determine, while x_1 refers to the age (higher than x_0), when the remaining life expectancy is already lower than the specified value. Concerning the quantity z (the remaining life expectancy at the prospective age we wish to determine), the designation z_0 denotes the life expectancy at age x_0 and z_1 the remaining life expectancy at age x_1.

The so-called **Constant Prospective Age** was chosen for the analytical needs of the paper in accordance with similar studies, i.e. a remaining life expectancy (RLE) of 15 years **(*CPA RLE* 15–)**. This age is based on the fact that for each population in each year we are attempting to determine the age with a remaining life expectancy equal to 15 years (Sanderson and Scherbov 2013), for which the linear interpolation method can also be applied. The constant prospective age can be employed, for example, when calculating the proportion of elderly people in the selected population. One of the specific indicators concerns the **Proportion of the Population at Ages with Remaining Life Expectancy of 15 Years or Less (*Prop.RLE* 15–)**, for which the criterion 15 years of life remaining is equivalent to the standard indicator the **Proportion of the Population at Ages 65 Years or Above (*Prop*.65+)**; moreover, according to Sanderson and Scherbov (2008, 2010); or Lutz et al. (2008a, 2008b), the indicator sets an alternative age threshold of 65 years. The calculation formulae are as follows:

$$Prop.RLE\ 15- = \frac{P_{x_{RLE15-}}}{P} * 100$$

$$Prop.65+ = \frac{P_{65+}}{P} * 100$$

Further indicators considered in this paper comprise the **Prospective Ageing Index (*PAI*)** and its standard counterpart the **Ageing Index (*AI*)**. Both are calculated as the ratio of the number of people at older ages (specifically ages with a remaining life expectancy of 15 years or less – $P_{xRLE15-}$ or the number of people aged 65 and over – P_{65+}) to the number of people from birth to the completed age of 19 years (P_{0-19}). The corresponding formulae are as follows:

$$PAI = \frac{P_{x_{RLE15-}}}{P_{0-19}} * 100$$

$$AI = \frac{P_{65+}}{P_{0-19}} * 100$$

A further indicator used in the analysis of demographic ageing concerns the **Prospective Old-Age Dependency Ratio (*POADR*)**; this very specific indicator is calculated by combining the prospective and retrospective ages. It can be expressed as the relation between the number of people with a remaining life expectancy of 15 years or less to the number of people between the age of 20 and the considered prospective age limit (constant prospective age) (Sanderson and Scherbov 2008). The retrospective analogy of the *POADR* indicator comprises the standard **Old-Age Dependency Ratio (*OADR*)**. The calculation of the indicators can be illustrated as follows:

$$POADR = \frac{P_{x_{RLE15-}}}{P_{20-x_{RLE>15}}} * 100$$

$$OADR = \frac{P_{65+}}{P_{20-64}} * 100$$

where $P_{20-xRLE>15}$ defines the number of persons from 20 years of age to the constant prospective age.

Changes in the age structure significantly (e.g. the demographic dividend) affect economic growth (Bloom et al. 2010), where investment in the development of human capital and technological progress serve to increase the economic productivity of societies and regions (Spijker 2015). The "economic miracle" in East Asia from the mid-1960s to the early 1990s provides a typical example of the use of changes in the age structure (Bloom et al. 2010). The main impact of demographic ageing on the economy concerns the increasing number of older economically inactive (dependent) citizens compared to the economically active population. However, it is important to consider that the size of the economically active population does not depend only on age (Spijker and MacInnes 2013). Historically, it was influenced by, for example, lower female employment rates, which have increased dramatically over the last 50 years (the elimination of gender discrimination and the weakening of the "male breadwinner" paradigm) (Spijker 2015).

In addition, a significant part of the working-age population cannot be considered to be part of the labour force since they are not active in the labour market (students etc.), whereas conversely, in some countries persons over 65 years of age are considered to be economically active (Spijker 2015), which is related to personal preferences, the inability to live solely on the amount of pension received or, simply, the absence of a national old-age pension system. On the other end of the scale, certain high-income countries have statutory retirement ages of lower than 65 years (e.g. France and Japan), in which case the number of people receiving a pension is likely to exceed the number of people aged 65 and over (Bongaarts 2004).

Based on the above-mentioned factors, efforts have been made since the turn of the millennium to determine alternative demographic ageing indicators, e.g. Bongaarts (2004) considered the **Pensioners to Workers Ratio (*PWR*)** to be a more accurate indicator of the burden on the economically active population:

$$PWR = \frac{P^{PENS}}{P^{PE}} * 100$$

According to Spijker (2015), this ratio attains much higher values than the conventional *OADR*, especially in certain high-income countries (especially France and Italy). Interesting proposals include, for example, that only those in paid employment (P^{PE}) should pay for the health and well-being of older (P^{PENS}) people (not all people of working age) (Spijker and MacInnes 2013). Subsequently, an indicator known as the **Real Elderly Dependency Ratio (*REDR*)** was proposed whereby Spijker (2023) and Spijker and MacInnes (2013) applied their ideas in combination with the methodology considered by Sanderson and Scherbov (2007) that applied the constant prospective age method to determine the older part of the population. The calculation formula is as follows:

$$REDR = \frac{P_{x_{RLE15-}}}{P^{PE}} * 100$$

where the numerator ($P_{xRLE15-}$) refers to the number of men and women with a remaining life expectancy of 15 years or less and the denominator (P^{PE}) indicates the number of men and women in paid employment. One of the disadvantages of the *REDR* proposal concerns the fact that it considers all people in paid employment to be equally productive despite probable differences (e.g. in the hours worked) (Spijker 2015).

The authors of this article selected the NUTS 2 regions of Czechia for their relatively detailed analysis of the development of the chronological and alternative indicators of demographic ageing aimed at both determining and clarifying the development of regional differences over the period 2003–2023. NUTS regions were created so as to allow for the comparability of statistical data within the EU. Given the differences in the spatial characteristics of the NUTS 3 regions of Czechia, it was considered that more reliable outcomes would be determined via the comparison of the statistical data for the higher order NUTS 2 regions. It is evident that although demographic ageing is, in most cases, studied at the national level, populations do not age uniformly at the regional level. Since the aim of this article is to determine in which parts of Czechia the level of demographic decline is most advanced and in which, conversely, the population is ageing more slowly, it was deemed most appropriate to study the NUTS 2 regions.

The time period was chosen aimed at both capturing developments and changes essentially from the beginning of the new millennium to the present day and taking into account the impact of the Covid-19 pandemic on the considered indicators at the regional scale. The Public Database of the Czech Statistical Office (CZSO) and the Product Catalogue (Statistical Yearbooks of the Czech Republic 2004–2024) (CZSO 2024a; CZSO 2024b) served as the data bases. Specifically, detailed mortality tables for each year were employed (life expectancy at specific ages) together with the sex-age structure according to age units for the given years (number of persons at a specific age). The basic characteristics of the economic status of the population aged 15 and over were taken from a labour force sample survey (the number of persons in paid employment) together with selected indicators for the NUTS 2 and NUTS 3 regions for each given year (the number of pensioners).

Within the research, two main hypotheses were formulated. The first hypothesis posits the absence of homogenisation in the level of demographic ageing across the NUTS 2 regions of Czechia over time. The second hypothesis states that the values of alternative (including prospective) indicators – based on life table measures – are more sensitive to external mortality influences, such as various epidemics (e.g. Covid-19), thereby resulting in greater differentiation of these indicators over time.

3. Results

The study of the development of demographic ageing at the regional level in Czechia, specifically NUTS 2 regions, in the period 2003 and 2023 revealed significant inter-regional differences. In terms of the constant prospective age (*CPA RLE* 15–), differences were evident both between the genders and the NUTS 2 regions (Fig. 1). The constant prospective age values for men were significantly lower than for women within the individual regions, which is not surprising given the higher excess mortality of men at middle and older ages due to their more at-risk behaviour (Rochelle et al. 2015) and higher consumption of addictive substances (Clark and Peck 2012). Women, on the other hand, devote more attention to their health and more often undergo preventive health check-ups (Clark and Peck 2012). A further reason concerns genetic predisposition, e.g. women have higher levels of oestrogen in their blood, a hormone that protects against cardiovascular diseases (Rochelle et al. 2015). In addition, the research revealed greater interregional differences for the male part of the population than for the female part.

The lowest values of the indicator were recorded for both sexes in the North-west and Moravian-Silesian regions (Fig. 1), both of which exhibit lower than average life expectancy at birth and in older ages, higher unemployment rates and the highest levels of air pollution in Czechia due to the historically high presence of heavy industry. Conversely, the highest values were observed in the Prague region. These three regions recorded the lowest and highest life expectancy values throughout the studied period both at birth and at age 65, as well as at the constant prospective age. Moreover, it was observed that the age for determining the expected last 15 years of life was higher than the limit of 65 years for women in all the monitored regions throughout the period, and since 2007 it did not fall below 68 years in any of the regions. The constant prospective age for men did not reach values above 65 years throughout the period in any of the regions studied, with the exception of the Prague region in the first of the monitored years, i.e. 2003. During the first 10 monitored years, all the regions reached the value of 65 years for men, except for the afore-mentioned North-west region where this value was exceeded only 5 years later, in 2016.

The highest gender differences between the regions were observed for 2021, which was most likely linked to the Covid-19 pandemic, concerning which a paper by Dzúrová and Hulíková (2021) noted a higher level of excess mortality in the male part of the population from Covid-19 than in the female population. This was further confirmed by the higher decline in life expectancy values for men in Czechia during the Covid-19 pandemic than for women (Appendix 1).

One possible explanation for the regional differences relates to the higher quality of life in cities (the Prague region comprises only the city of Prague). Large cities and their hinterlands (the Central Bohemia region can be considered the Prague hinterland) have higher levels of employment in the tertiary sector than other parts of the country and lower levels of physically demanding professions. This is supported

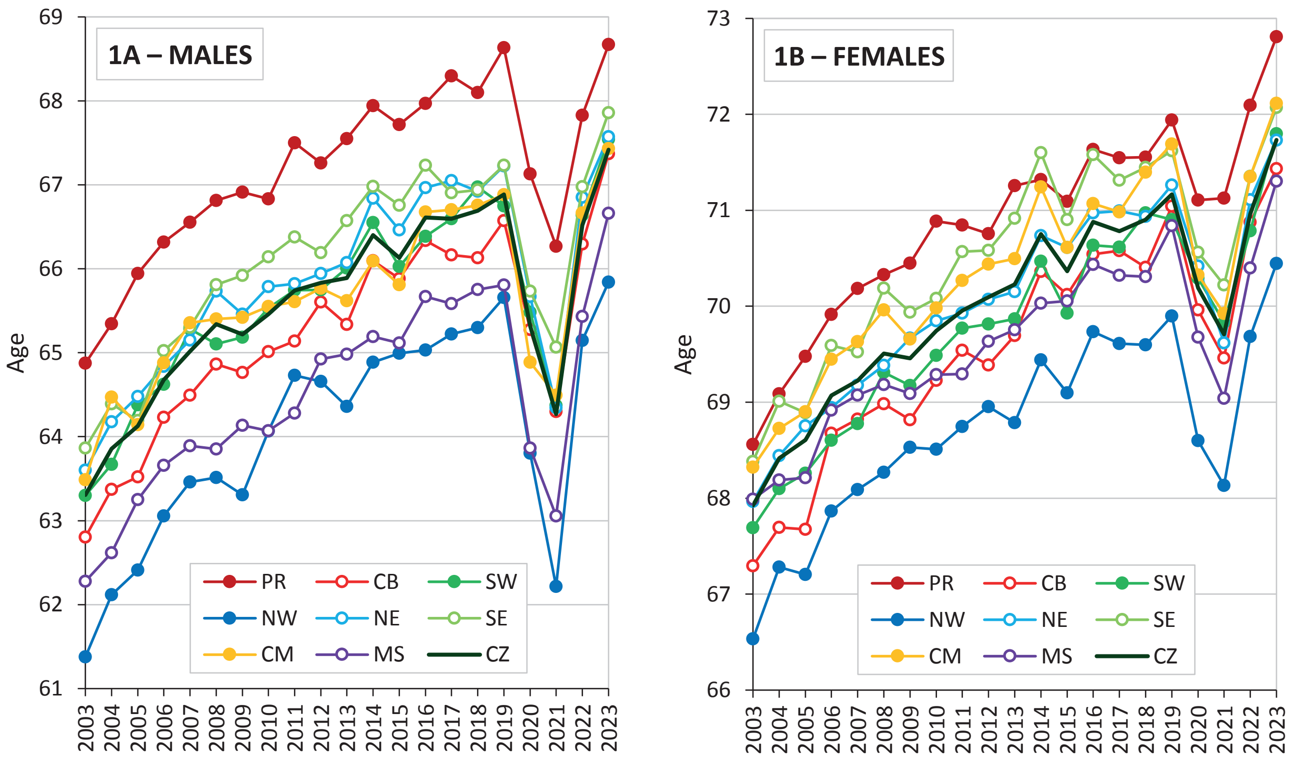

Fig. 1 Constant Prospective Age (*CPA RLE* 15–), Males (1A) and Females (1B), NUTS 2 Regions of Czechia, 2003–2023.
Note: PR = Prague region (CZ01), CB = Central Bohemia region (CZ02), SW = South-west region (CZ03), NW = North-west region (CZ04), NE = North-east region (CZ05), SE = South-east region (CZ06), CM = Central Moravia region (CZ07), MS = Moravian-Silesian region (CZ08), CZ = Czechia. Source: CZSO; authors' calculations.

by the fact that more people with higher levels of education who, according to Murtin et al. (2017) have higher average life expectancy rates, live in larger cities. In addition, Czechia routinely reports relatively higher differences in life expectancy values based on the level of education achieved than other European countries (Murtin et al. 2017).

The proportions of people in older age categories (the proportion of the population aged 65 and over – *Prop.*65+, and the proportion at ages with a remaining life expectancy of 15 years and less – *Prop.RLE* 15–) is one of the characteristics that is often used to quantify and analyse the population ageing process. Moreover, the values applied in the two approaches differ significantly (Appendix 2). The *Prop.RLE* 15– in the individual regions evinced lower values than the *Prop. 65+* throughout the monitored period. Values obtained for 2021 revealed the significant influence of the Covid-19 pandemic via the convergence of the values of the two approaches, which resulted in a decrease in the life expectancy values of older age groups in society and, thus, a decrease in the constant prospective age values.

The differences between the two approaches evinced a mainly increasing trend. The Prague region exhibited the greatest differences between the two approaches throughout the monitored period (with the exception of 2019 and 2023), while the North-west region evinced the lowest differences – even in 2003 the values of the standard proportion were lower than the prospective share in this region. The authors assume that the main reason for the North-west region achieving the lowest differencies concern the low values of the life expectancy (thus also low values of constant prospective age) in this region. The North-west region is also characterized by the relatively high unemployment rate (CZSO 2025), which is reflected in the emigration of younger, more educated people from this region.

Fig. 2 illustrates the shares of people with a remaining life expectancy of 15 years or less. The NUTS 2 regions did not experience particularly sharp increases in the values during the monitored period. This was due to the application of a variable age limit, which is the basis of the prospective age concept.

While the share of people aged 65 and over evinced almost linear growth in most Czech NUTS 2 regions between 2003 and 2023 (Fig. 2), the share of people with a remaining life expectancy of 15 years or less exhibited a sharp increase in values in 2020 and 2021. Moreover, the values of the two indicators converged significantly in these two years. This factor was attributed to the Covid-19 pandemic, which reduced life expectancy values at older ages (Dzúrová and Hulíková 2021).

Other indicators employed to analyse demographic ageing include the prospective ageing index and the prospective old-age dependency ratio, and its

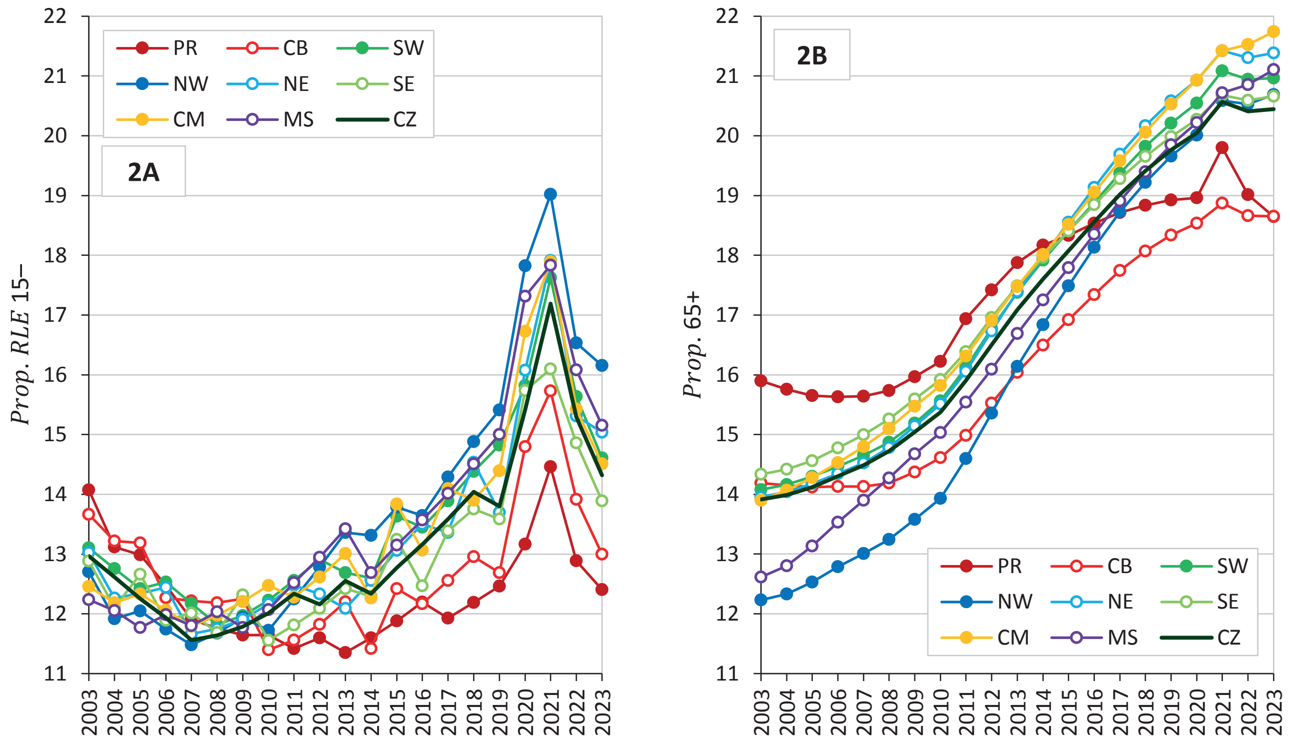

Fig. 2 Proportions of the Population at Ages with Remaining Life Expectancy of 15 years or less (*Prop.RLE* 15–) (2A), Proportions of the population aged 65 and over (*Prop*.65+) (2B), NUTS 2 regions of Czechia, 2003–2023.
Note: see Fig. 1. Source: CZSO; authors' calculations.

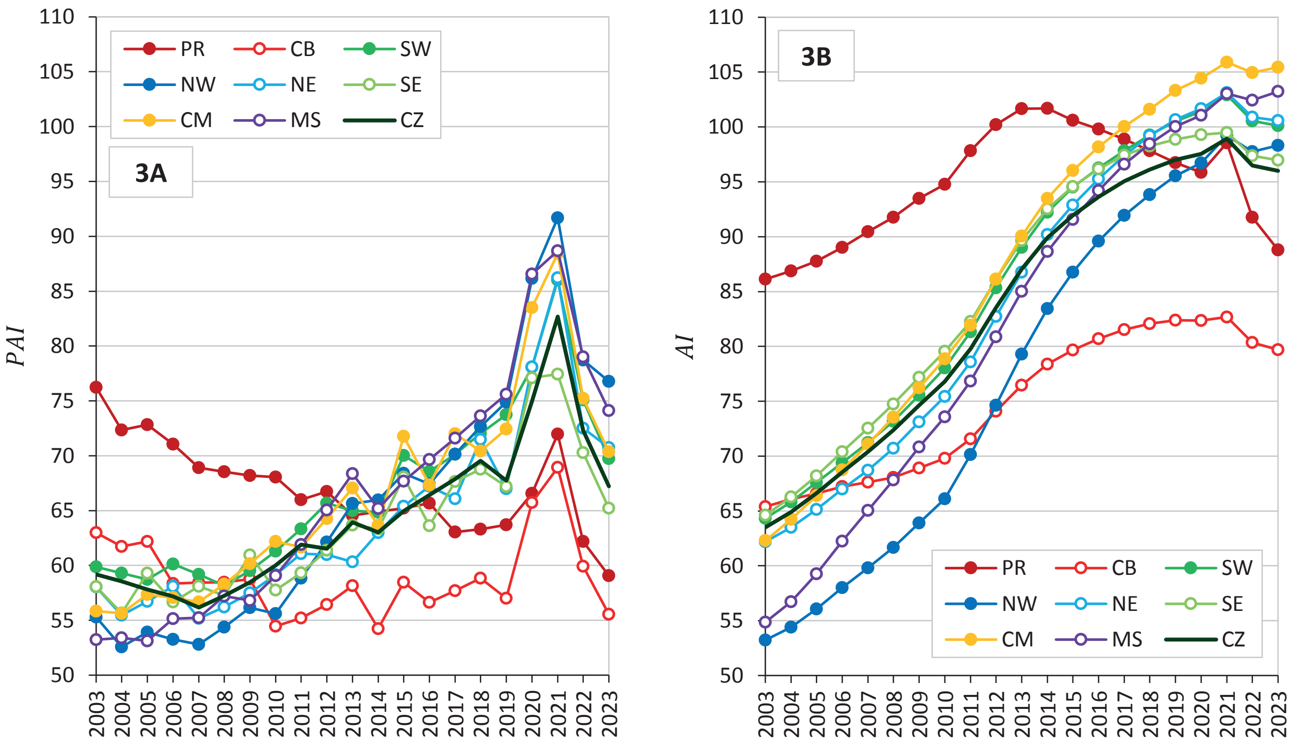

Fig. 3 Prospective Ageing Index (*PAI*) (3A) and the Ageing Index (*AI*) (3B), NUTS 2 regions of Czechia, 2003–2023.
Note: see Fig. 1.Source: CZSO; authors' calculations.

standard counterparts, i.e. the ageing index and the old-age dependency ratio.

Both pairs of indicators are closely related both to mortality at older ages and the fertility levels of people in the productive age group. Based on the time development of the prospective ageing index (Fig. 3), the NUTS 2 regions could be divided into two groups. The first group comprised the Prague and Central Bohemia regions, for which a decreasing trend in the development of values was observed. The second group consisted of the remaining six regions, for which the development trend was mainly of an increasing character. The standard ageing index values exhibited an increasing trend.

Once again, the highest differences were recorded in the NUTS 2 region of Prague, and in the last two years of the monitored period also in the Central Moravia region (Appendix 3). The lowest differences were observed in the North-west and Central Bohemia regions, as well as in the Moravian-Silesian region at the beginning of the monitored period. Although both the Prague and Central Bohemia regions were amongst the oldest regions in Czechia at the beginning of the monitored period (Fig. 3), they were amongst the youngest regions at the end of the period. The Central Bohemia region was a stagnant region with a decreasing population in the early 1990s; however, by 2010 it had become the most populous and in 2011 the youngest region in Czechia (CZSO 2023). One possible explanation for the downward trend in both regions concerns their mutual demographic, social and economic interconnection. For example, the Central Bohemia region is today considered a suburban area of the Prague region and according to Kashnitsky et al. (2020), mainly younger people of productive age move to the hinterlands of large cities, where they start a family, thus increasing the number of those in the pre-productive age category (0–19 years).

This change was confirmed by the CZSO (2023), which reported that the spread of suburbanisation from Prague to Central Bohemia is associated with intensive housing construction, especially in the immediate vicinity of Prague and in areas with good transport accessibility. According to Čábela (2023), the LAU 1 regions of Prague-West and Prague-East (the immediate hinterland of Prague) evince the lowest prospective ageing index values in Czechia. In addition, a lower proportion of people with a remaining life expectancy of 15 years and less has been recorded in both NUTS 2 regions since 2010. The decrease in the prospective ageing index in the Prague region can be attributed to both the highest life expectancy in Czechia and the permanently negative migration of people aged 60 and older from this region (CZSO 2012), as well as the immigration of younger people for study or career purposes.

Very similar trends can be observed in the development of both the prospective and standard old-age dependency ratios (Fig. 4). Again, significant increases are evident in the values of the standard indicator, while the increase in the prospective index is gradual with slight fluctuations at around 2015 caused by an increase in the intensity of mortality in Czechia (Kurkin et al. 2016) and at around 2020 and 2021 due to the Covid-19 pandemic (Dzúrová and Hulíková 2021). The differences between the values of the measurement approaches again increased throughout the period (Appendix 4), apart from the pandemic years 2020 and 2021. Concerning the prospective ageing index and the prospective old-age dependency ratio, a decrease in values can be observed in all the regions, which was most probably due to the end of the Covid-19 pandemic. A further explanation is provided by the wave of immigration since the outbreak of the war in Ukraine, when approximately 300 thousand people, mainly young mothers with small children, immigrated to Czechia; this served to increase the number of people in both the productive and pre-productive age categories (Štyglerová and Němečková 2023).

The pensioners to workers ratio comprises one of the more recent alternative indicators considered in the article. This indicator includes all pensioners regardless of whether they receive an old-age or other pension and all economically active residents regardless of age (from 15 years old). This indicator potentially provides a more accurate insight into the real situation in terms of dependency; moreover, it serves for the comparison of all dependent residents (= pensioners) in Czechia (except for the pre-productive age category) with those who contribute to the pension system (= workers). Fig. 5 illustrates the significant regional differences in the *PWR* values. The differences in the values between 2003, 2013 and 2023 are shown in Appendix 5.

The value of the indicator decreased between 2003 and 2023 in the Prague region due most likely to an increase in the number of workers by more than 70 thousand, while the number of pensioners remained essentially the same (CZSO 2024b). A further possible explanation is that many people rather than spending their pension years in the hustle of a large city, prefer to live in more peaceful rural areas. A decrease in this indicator was also evident for the Central Bohemia region, where, although the number of pensioners increased slightly, the number of workers increased by more than 120 thousand people in the studied period (CZSO 2024b). The *PWR* indicator values increased in all the other NUTS 2 regions between 2003 and 2023. However, it was interesting to note that between 2013 and 2023, all the regions evinced decreases in this indicator value (Appendix 5). Thus, overall increases in the *PWR* values of the NUTS 2 regions in the first half of the monitored period were followed by decreasing values in the second half of the period. The highest values achieved during the monitored period related to the NUTS 2 region Moravian-Silesian, which was most probably due to this region having the highest numbers of pensioners of all the regions and, conversely, one of the lowest proportions of workers. The high number of pensioners could be attributed to the historical predominance of the mining industry and the associated earlier retirement age of miners (Štefko 2018).

The real elderly dependency ratio indicator serves for the comparison of the number of people with a remaining life expectancy of 15 years or less to

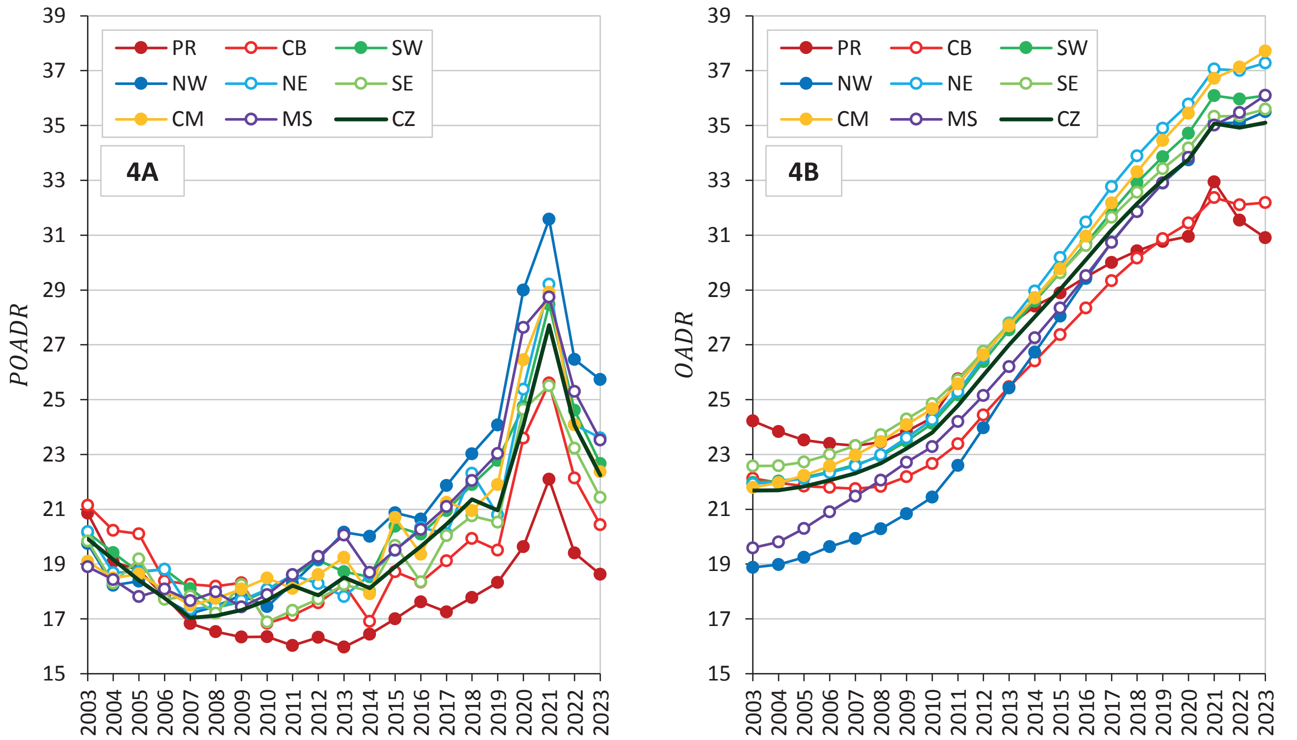

Fig. 4 Prospective Old-Age Dependency Ratio (*POADR*) (4A) and the Old-Age Dependency Ratio (*OADR*) (4B), NUTS 2 regions of Czechia, 2003–2023.
Note: see Fig. 1. Source: CZSO; authors' calculations.

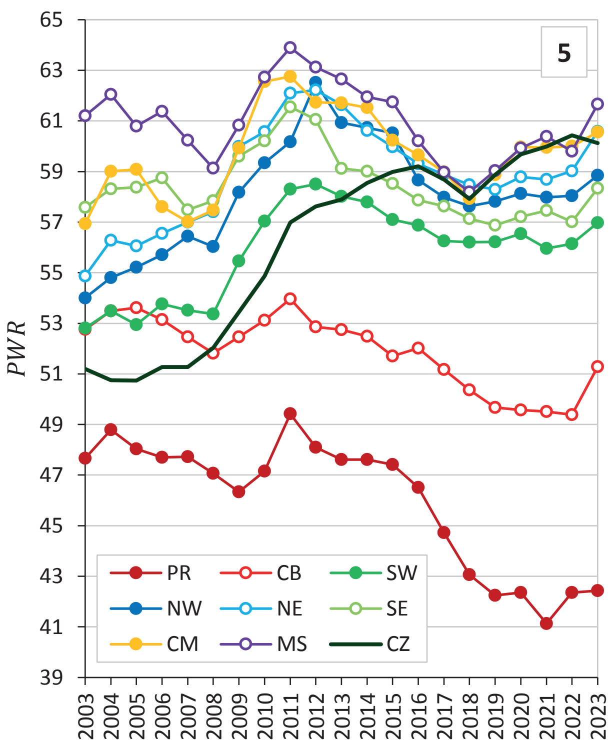

Fig. 5 Pensioners to Workers Ratio (*PWR*), NUTS 2 regions of Czechia, 2003–2023.
Note: see Fig. 1. Source: CZSO; authors' calculations.

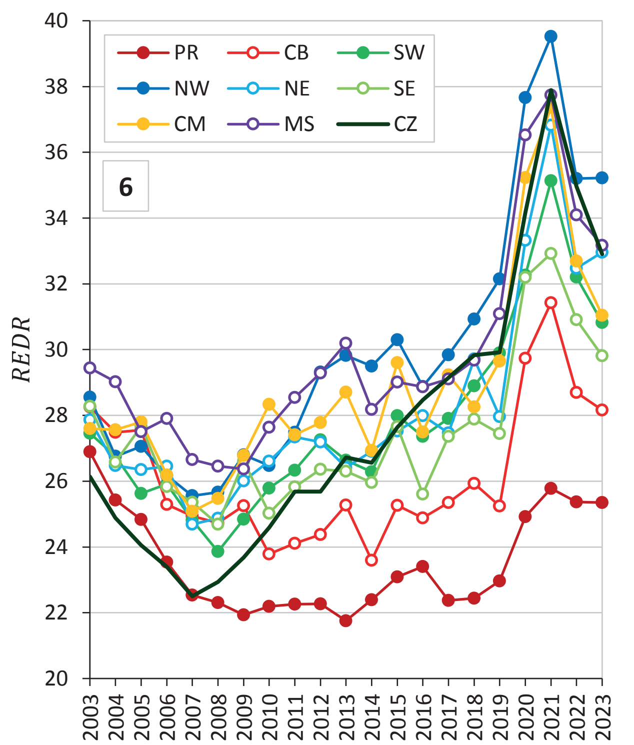

Fig. 6 Real Elderly Dependency Ratio (*REDR*), NUTS 2 regions of Czechia, 2003–2023.
Note: see Fig. 1. Source: CZSO; authors' calculations.

the number of people in paid employment. Spijker and MacInnes (2013) introduced this indicator on the grounds that only those who are employed and receive a salary should be considered as the population that finances the health and wellbeing of older people rather than all working-age people.

The development of the real elderly dependency ratio demographic ageing indicator during the monitored period was very similar to that of the prospective old-age dependency ratio indicator. This was not a coincidence since both indicators are very similar in terms of the calculation approach. The comparison of the *REDR* (Fig. 6) with the *POADR* (Fig. 4) revealed more significant differences between the individual regions for the *REDR* indicator. The highest *REDR* values were observed at the beginning of the monitored period for the Moravian-Silesian region and the lowest for the Prague region, whereas at the end of the period the highest values were observed for the North-west region and, again, the lowest for the Prague region. The North-west region has one of the highest unemployment rates in Czechia (CZSO 2025) and a relatively low quality of life. Appendix 6 shows that between 2003 and 2023, the value decreased only in the Prague and Central Bohemia regions, while the values increased in all the other regions.

Significantly high relative increases were recorded in all the regions in the period 2013–2023, which was probably at least partly due to the sharp increase in values during the Covid-19 pandemic; the indicator values had not returned to pre-Covid values at the end of the monitored period. The development of the *REDR* for the Prague region in 2020 and 2021 is interesting; all the other regions recorded a sharp increase in this indicator value, whereas the increase for the Prague region was notably less intense (as was the subsequent decrease). The decrease in the values after 2021 in all the regions was, again, due mainly to the end of the Covid-19 pandemic.

4. Discussion

Changes in the age structure of the population, as associated with the process of demographic ageing, exert a wide range of socioeconomic consequences. For example, they lead to an increased burden on the pension system, higher healthcare costs, greater demands in terms of caring for the elderly and labour shortages. Simply put, many countries are witnessing a situation in which in today's media age, the term demographic ageing is perceived negatively by the majority of society, a situation that is compounded by the fact that the extension of life expectancy at older ages does not automatically mean an extension of life expectancy in good health (e.g. Straka et al. 2024). This has resulted in an interesting paradox, i.e. one of the greatest achievements of the modern human era – increasing life expectancy – is generally perceived as an existential problem for the population (Spijker and MacInnes 2013). Although the potential impacts of ageing populations are significant in many areas, this phenomenon should not be viewed solely from the negative perspective. It should also be seen as a challenge for society, one that opens up new opportunities in the context of active ageing (e.g. Forster and Walker 2015), the participation of seniors in the labour market (e.g. Posthuma and Campion 2009) and research and development and innovation targeted at the specific needs of seniors.

The application of standard demographic ageing indicators based on chronological age is limited by the ability to compare these indicators across time and space. Alternative (including prospective) indicators, however, benefit from the fact that rather than employing fixed age limits, they consider changes in life expectancy values, i.e. variable age limits that better capture the real picture of the demographic ageing of society. For example, in many countries, the alternative principle for the measurement of ageing is used to set the retirement age (e.g. the Netherlands, Denmark and Greece, as well as Czechia – at the end of 2024, a pension reform was approved concerning the application of the alternative approach in the calculation of this age). However, the two approaches are interconnected and should not be applied in isolation. Together, they allow for the analysis of both the age dimension and the assessment of the ageing process of the population as a whole, which is also confirmed in the article by Šprocha and Ďurček (2018).

The comparison of development trends is better able to clarify ongoing changes and potential impacts on the age structure of the population. The increasing differences in the values of the indicators between the two approaches are the consequence of increasing life expectancy, which acts to increase the value of the constant prospective age (and, at the same time, diverges from the fixed age limit of 65 years). On the other hand, the convergence of the values of the indicators of the two approaches (e.g. as during the Covid-19 pandemic) reflects a decrease in life expectancy as a cross-sectional indicator in a given calendar year even though this impact is not particularly noticeable in the resulting representation of the number of people of senior age.

One of the disadvantages of applying alternative (including prospective) indicators of demographic ageing concerns their higher variability than more traditional indicators and susceptibility to minor fluctuations during the monitored period. Moreover, since their calculation is tied to cross-sectional mortality tables, these indicators fully reflect those changes that cause temporary fluctuations to current mortality rates. Despite these minor disadvantages, it is reasonable to assume that the application of a concept based on the number of years of life remaining to social, economic and healthcare systems has the potential to reduce state expenditure. Moreover, the authors recommend that these approaches be used to provide support primarily for those who really need it. At the same time, however, it is important to note that population ageing is associated solely with age. A certain degree of dependency in society is associated purely with pensioners or with the real number of people in paid employment (*PWR* and *REDR*). These alternative indicators provide proof that the demographic ageing of the population can be viewed in various ways, which influence the perception of population ageing.

However, since this article aimed to describe changes at the regional level, it is important to consider why the analysis of regional differences is crucial in terms of demographic ageing. National averages often mask significant heterogeneities between regions. However, analysis at the regional level allows for the more detailed understanding of the dynamics of ageing in different parts of a country, which often exhibit different fertility, migration and mortality trends. These differences may be due to historical, economic, sociocultural and geographical factors. For example, regions with a traditionally strong industrial sector may face specific challenges associated with the outflow of young labour, while rural areas may face more rapid population ageing due to the migration of young people to large cities. The distinctiveness of individual regions thus indicates the need for a specific, targeted approach to local policy intervention, e.g. regions with more progressive demographic ageing rates must adopt more active approaches to addressing, for instance, the construction of residential facilities for seniors or the support of other forms of social services for this group. It is important that, in cooperation with health insurance companies, such regions should optimise the network of healthcare providers so as to be able to meet the increasing demand for healthcare associated with advancing age and adapt local infrastructures, including public transport, and services to better meet the needs of an ageing population.

5. Conclusions

In today's world of easily accessible information, society is being influenced from all sides by opinions regarding the development and necessity of solving the issue of demographic ageing. Negative attitudes often prevail in the media, which is due largely to the standard (retrospective) approach to measuring the level of demographic ageing. Unfortunately, emphasis is often placed only on increasing numbers over time, without the explanation of their values, in terms of either the causes of, or connections between, rapid increases in standard indicator values.

The aims of this article were to both draw attention to this issue and compare standard and alternative approaches to measuring ageing, and to provide new insights into regional differences in terms of the development and pace of demographic ageing at the Czech NUTS 2 regional level. Although it may seem at first glance that these regions are relatively stable and mutually comparable, the results of the analysis indicate the opposite. Persistent differences were revealed between more economically developed regions and regions that are struggling with socioeconomic problems, which often have their roots in the transformation of the industrial base and related structural changes. A number of indicators even pointed to an increase in heterogeneity, thus supporting the first selected hypothesis. Accordingly, it is essential that society responds adequately to these changes via the introduction of policy measures at both the national and local government levels. The second hypothesis, suggesting that alternative indicators of demographic ageing are more sensitive to external mortality influences, was likewise confirmed (e.g. Covid-19 pandemic). Following a constructive approach to demographic ageing has the potential to uncover more positive responses to the various social challenges associated with this major 21st century phenomenon aimed at combating the negative perceptions that are currently prevalent in society.

Acknowledgments

The article was written with the financial support of institutional supports of research and creative activities at Charles University: Cooperatio Geography (COOP) research programme and the Specific Academic Research (SVV).

References

Bloom, D. E., Canning, D., Fink, G. (2010): Implications of Population Ageing for Economic Growth. Oxford Review of Economic Policy 26(4), 583–612, https://doi.org/10.1093/oxrep/grq038.

Bongaarts, J. (2004): Population aging and the rising cost of public pensions. Population and Development Review 30(1), 1–23, https://doi.org/10.1111/j.1728-4457.2004.00001.x.

Burcin, B., Šprocha, B., Šídlo, L. (2023): Vybrané aspekty úmrtnosti v Česku a na Slovensku v pandemickém roce 2020. Epidemiologie, mikrobiologie, imunologie 72(2), 99–110 (in Czech).

Clark, R., Peck, M. (2012): Examining the Gender Gap in Life Expectancy: A Cross-National Analysis, 1980–2005. Social Science Quarterly 93(3), 820–837, https://doi.org/10.1111/j.1540-6237.2012.00881.x.

Czaja, S. J., Boot, W. R., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., Sharit, J. (2006): Improving technology for older adults. Human Factors: The Journal of the Human Factors and Ergonomics Society 48(3), 625–663.

Cuaresma, J. C., Lábaj, M., Pružinský, P. (2014): Prospective Ageing and Economic Growth in Europe. The Journal of the Economics of Ageing 3, 50–57, https://doi.org/10.1016/j.jeoa.2014.05.003.

CZSO (Czech Statistical Office) (2012): Migrace v hl. m. Praze v letech 2001 až 2011 – Saldo migrace v Praze. Available online: https://csu.gov.cz/docs/107508/32905400-cb55-f8d5-3e3e-2e4ecc2d0a97/104035

-12k05.pdf?version=1.0 (accessed on 20 March 2025, in Czech).
CZSO (Czech Statistical Office) (2022a): Vloni zemřelo v Česku nejvíce lidí od konce války. Available online: https://csu.gov.cz/produkty/vloni-zemrelo-v-cesku-nejvice-lidi-od-konce-valky (accessed on 10 March 2025, in Czech).
CZSO (Czech Statistical Office) (2023): Základní tendence demografického, sociálního a ekonomického vývoje Středočeského kraje. Available online: https://csu.gov.cz/docs/107508/cd06ac63-add5-d884-7cdb-55bf7ad72b47/3301262401.pdf?version=1.7 (accessed on 19 March 2025, in Czech).
CZSO (Czech Statistical Office) (2024a): Veřejná databáze. Available online: https://vdb.czso.cz/vdbvo2/faces/cs/index.jsf?page=statistiky (accessed on 1 February 2025, in Czech).
CZSO (Czech Statistical Office) (2024b): Katalog produktů. Available online: https://csu.gov.cz/katalog-produktu?pocet=10&start=0&pouzeVydane=true&pouzePosledni=false&razeni=-datumVydani (accessed on 20 February 2025, in Czech).
CZSO (Czech Statistical Office) (2025): Nezaměstnanost v Ústeckém kraji k 31. 3. 2025 (podle MPSV). Available online: https://csu.gov.cz/ulk/nezamestnanost-v-usteckem-kraji-k-31-3-2025-podle-mpsv (accessed on 30 April 2025, in Czech).
Čábela, F. (2023): Aplikace prospektivních ukazatelů demografického stárnutí na okresní úrovni v Česku. Diploma Thesis. Charles University, Faculty of Science, Department of Deography and Geodemography. Supervizor: Luděk Šídlo. Available online: https://dspace.cuni.cz/handle/20.500.11956/181078 (accessed on 12 March 2025, in Czech).
Čábela, F., Šídlo, L. (2024): Demografické stárnutí v Česku mezi lety 2012–2022 prostřednictvím vybraných retrospektivních a prospektivních ukazatelů. Demografie 66(2), 154–165, https://doi.org/10.54694/dem.0341 (in Czech).
Dzúrová, D., Hulíková, K. (2021): Pandemie Covid-19 a její dopad na úmrtnost v Česku. 50th Conference of Czech Demographic Society, Ostrava, 15.–17. September 2021. Available online: https://www.czechdemography.cz/res/archive/008/000963.pdf (accessed on 11 March 2025, in Czech).
Eurostat (2025): EU population diversity by citizenship and country of birth. Statistics Explained, Lucemburk. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_population_diversity_by_citizenship_and_country_of_birth (accessed on 1 March 2025).
Gietel-Basten, S., Saucedo, S. E. G., Scherbov, S. (2020): Prospective measures of aging for Central and South America. PLoS One 15(7), https://doi.org/10.1371/journal.pone.0236280.
Foster, L., Walker, A. (2015): Active and successful aging: a European policy perspective. Gerontologist 55(1), 83–90, https://doi.org/10.1093/geront/gnu028.
Ghio, D., Goujon, A., Natale, F. (2022): Assessing the demographic impact of migration on the working-age population across European territories. Demographic Research 46(9), 261–272, https://doi.org/10.4054/DemRes.2022.46.9.
Gregory, T., Patuelli, R. (2013): Regional Age Structure, Human Capital and Innovation – Is Demographic Ageing Increasing Regional Disparities? Discussion Paper 13-057, 1–31, http://dx.doi.org/10.2139/ssrn.2316259.
Gruenberg, E. M. (1977): The Failures of Success. Milbank Q 2005 83(4), 779–800, https://doi.org/10.1111/j.1468-0009.2005.00400.x.
Holzmann, R. (2013): A Provocative Perspective on Population Aging and Old-Age Financial Protection. IZA Discussion Paper Series, 7571. Available online: https://ssrn.com/abstract=2318752 (accessed on 5 March 2025).
Ježek, J. (2016): Geostatistika a prostorová interpolace. Praha: Nakladatelství Karolinum (in Czech).
Kashnitsky, I., De Beer, J., van Wissen, L. (2020): Unequally ageing regions of Europe: Exploring the role of urbanization. Population Studies 75(2), 221–237, https://doi.org/10.1080/00324728.2020.1788130.
Klapková, M., Šídlo, L., Šprocha, B. (2016): Koncept prospektivního věku a jeho aplikace na vybrané ukazatele demografického stárnutí. Demografie 58(2), 129–141 (in Czech).
Koukalová, J. (2023): Population development in the Czech Republic in 2022. Demografie 65(4), 221–239, https://doi.org/10.54694/dem.0336.
Kurkin, R., Němečková, M., Štyglerová, T. (2016): Population Development in the Czech Republic in 2015. Demografie 58(4), s. 299–319.
Lesthaeghe, R. (2010): The unfolding story of the second demographic transition. Population and Development Review 36(2), 211–251, https://doi.org/10.1111/j.1728-4457.2010.00328.x.
Lutz, W., Sanderson, W. C., Scherbov, S. (2008a): Global and regional population ageing: how certain are we of its dimensions? Population Ageing 1, 75–97, https://doi.org/10.1007/s12062-009-9005-5.
Lutz, W., Sanderson, W. C., Scherbov, S. (2008b): The Comming Acceleration of Global Population Ageing. Nature 451, 716–719, https://doi.org/10.1038/nature06516.
Murtin, F., Mackenbach, J., Jasilionis, D., d'Ercole, M. M. (2017): Inequalities in longetivity education in OECD countries: Insights from new OECD estimates. OECD, https://dx.doi.org/10.1787/6b64d9cf-en.
Omran, A. R. (1971): The epidemiological transition: a theory of epidemiology of population change. Milbank Memorial Fund Quarterly 49(4), 509–538, https://doi.org/10.1111/j.1468-0009.2005.00398.x.
Posthuma, R. A., Campion, M. A. (2009): Age Stereotypes in The Workplace: Common Stereotypes, Moderators, and Future Research Directions. Journal of Management 35, 158–188, https://doi.org/10.1177/0149206308318617.
Riffe, T. (2015): The force of mortality by life lived is the force of increment by life left in stationary populations. Demographic Research 32, 827–834, https://doi.org/10.4054/DemRes.2015.32.29.
Rochelle, T., Yeung, D. K. Y., Bond, M. H., Li, L. M. W. (2015): Predictors of the gender gap in life expectancy across 54 nations. Psychology, Health & Medicine 20(2), 129–138, https://doi.org/10.1080/13548506.2014.936884.
Ryder, N. B. (1975): Noteson Stationary Populations. Population Index 41(1), 3–28, https://doi.org/10.2307/2734140.

Sanderson, W. C., Scherbov, S. (2005): Average remaining lifetimes can increase as human populations age. Nature 435(7043), 811–813, https://doi.org/10.1038/nature03593.

Sanderson, W. C., Scherbov, S. (2007): A new perspective on population aging. Demographic Research 16, 27–58, https://doi.org/10.4054/DemRes.2007.16.2.

Sanderson, W. C., Scherbov, S. (2008): Rethinking Age and Aging. Population Bulletin 63(4), 3–16. Available online: https://www.prb.org/wp-content/uploads/2021/01/Population-Bulletin-2008-63.4-aging.pdf (accessed on 3 March 2025).

Sanderson,W. C., Scherbov, S. (2010): Remeasuring Aging. Science 329(5 997), 1287–1288, https://doi.org/10.1126/science.1193647.

Sanderson, W. C., Scherbov, S. (2013): The Characteristics Approach to the Measurement of Population Aging. Population and Development Review 39(4), 673–685, https://doi.org/10.1111/j.1728-4457.2013.00633.x.

Sanderson, W. C., Scherbov, A. (2016): A Unifying Framework for the Study of Population Aging. Vienna Yearbook of Population Research 14, 7–39, https://doi.org/10.1553/populationyearbook2016s007.

Siegel, J. S., Davidson, M. (1984): Demographic and Socioeconomic Aspects of Aging in the United States. Current population reports. Available online: https://www2.census.gov/library/publications/1984/demographics/p23-138.pdf (accessed on 14 March 2025).

Siegel, J. S. (1993): A Generation of Change: a Profile of America's Older Population. New York: Russell Sage Foundation, 1993, xxxvi, 647 p.

Siegel, J. S. (2002): Applied demography. Applications to Business, Government, Law, and Public Policy, San Diego: Academic Press.

Skirbekk, V. F., Staudinger, U. M., Cohen, J. E. (2019): How to Measure Population Aging? The Answer Is Less than Obvious: A Review. Gerontology 65(2), 136–144, https://doi.org/10.1159/000494025.

Spijker, J., MacInnes, J. (2013): Population Ageing: the Time Bomb that Isn't? British Medical Journal 347: 6598, https://doi.org/10.1136/bmj.f6598.

Spijker, J., Riffe, T., MacInnes, J. (2014): Incorporating Time-To-Death (TTD) in health-based population ageing measurements. Available online: https://www.oeaw.ac.at/fileadmin/subsites/Institute/VID/PDF/Conferences/2014/p1-Spijker.pdf (accessed on 9 March 2025).

Spijker, J. (2015): Alternative indicators of population ageing: an inventory. Working papers, Vienna Institute of Demography, Vienna, https://doi.org/10.13140/RG.2.1.2115.8884.

Spijker, J. (2022): The future of ageing in Latin America and the Caribbean from different methodological perspectives. Papeles de Poblacion 28(114), 15–36, http://dx.doi.org/10.22185/24487147.2022.114.20.

Spijker, J. (2023): Combining remaining life expectancy and time to death as a measure of old-age dependency related to health care needs. International Journal of Health Economics and Management 23(2), 173–187, https://doi.org/10.1007/s10754-022-09328-7.

Straka, J., Šídlo, L., Kulhánová, I. (2024). Trends in healthy life years between 2005 and 2019 in 31 European countries: the compression or expansion of morbidity? International Journal of Public Health 69: 1607574, https://doi.org/10.3389/ijph.2024.1607574.

Šídlo, L., Šprocha, B., Ďurček, P. (2020a): A retrospective and prospective view of current and future population ageing in the European Union 28 countries. Moravian Geographical Reports 28(3), 187–207, https://doi.org/10.2478/mgr-2020-0014.

Šídlo, L., Šprocha, B., Ďurček, P. (2020b): Prospective dimension of population ageing and potential use in pension security in V4 countries. Ekonomický časopis 68(6), 601–621. Available online: https://www.sav.sk/?lang=en&doc=journal-list&part=article_response_page&journal_article_no=18540 (accessed on 7 March 2025).

Šídlo, L., Šprocha, B., Klapková, M. (2019): Regional differences in population aging in Europe viewed through prospective indicators. Erkunde 73(3), 225–240, https://doi.org/10.3112/erdkunde.2019.03.06.

Šprocha, B., Ďurček, P. (2018): Prospektívny prístup pri regionálnej analýze procesu populačného starnutia na Slovensku. Geographica Cassoviensis XII., 2, 212–223 (in Slovak).

Šprocha, B., Ďurček, P. (2019): Starnutie populácie Slovenska v čase a priestore. Bratislava: Prognostický ústav, 96 s. Available online: https://www.prog.sav.sk/wp-content/uploads/Starnutie-populacie-SR.pdf. (accessed on 12 July 2025, in Slovak).

Šprocha, B., Šídlo, L., Klapková, M., Ďurček, P. (2018): Nové prístupy k hodnoteniu procesu populačného starnutia a ich aplikácia v prípade Slovenska a Európy. Geografický časopis 70(4), 351–371, https://doi.org/10.31577/geogrcas.2018.70.4.19 (in Slovak).

Štefko, M. (2018): Kauza hornické důchody v reformě z roku 2016, odůvodněnost zvláštního důchodového věku. Časopis pro právní vědu a praxi, 26(1), 87–100, https://doi.org/10.5817/CPVP2018-1-6 (in Czech).

Štyglerová, T., Němečková, M. (2023): Odraz války na Ukrajině v demografické statistice Česka. 52th conference of Czech Demographic Society, Hradec Králové, 24.–26. May 2023. Available online: https://www.czechdemography.cz/res/archive/011/001312.pdf (accessed on 11 March 2025, in Czech).

van de Kaa, D. J. (1997): Options and sequences: Europe's demographic patterns. Journal of the Australian Population Association 14(1), 1–29. Available online: https://www.jstor.org/stable/41110439 (accessed on 2 March 2025).

Appendix 1 Life Expectancy at Age 65 (e_{65}) and the Constant Prospective Age (*CPA RLE* 15–), Males and Females, Czechia, 2003–2023.

Year	Males (M)		Females (F)		Difference F vs. M	
	e_{65}	CPA RLE 15–	e_{65}	CPA RLE 15–	e_{65}	CPA RLE 15–
2003	13.90	63.31	17.29	67.92	3.40	4.62
2004	14.25	63.86	17.68	68.42	3.43	4.56
2005	14.43	64.13	17.83	68.61	3.41	4.48
2006	14.79	64.68	18.21	69.07	3.42	4.39
2007	15.01	65.02	18.34	69.22	3.33	4.21
2008	15.22	65.34	18.57	69.51	3.35	4.17
2009	15.14	65.22	18.53	69.46	3.39	4.24
2010	15.29	65.46	18.75	69.75	3.45	4.29
2011	15.48	65.74	18.91	69.95	3.43	4.21
2012	15.55	65.83	19.04	70.09	3.49	4.26
2013	15.59	65.89	19.12	70.23	3.53	4.34
2014	15.93	66.40	19.58	70.75	3.64	4.35
2015	15.76	66.13	19.26	70.37	3.51	4.23
2016	16.09	66.61	19.70	70.88	3.61	4.27
2017	16.09	66.60	19.62	70.78	3.53	4.19
2018	16.14	66.69	19.71	70.90	3.57	4.21
2019	16.29	66.89	19.94	71.17	3.65	4.28
2020	15.22	65.32	19.17	70.19	3.95	4.87
2021	14.51	64.27	18.65	69.71	4.14	5.44
2022	16.05	66.52	19.81	70.96	3.76	4.44
2023	16.70	67.42	20.43	71.73	3.73	4.32
2023–2003 abs.	+2.80	+4.11	+3.14	+3.81	+0.34	-0.30
2023–2013 abs.	+1.11	+1.53	+1.31	+1.51	+0.20	-0.02
2013–2003 abs.	+1.69	+2.59	+1.82	+2.30	+0.14	-0.28
2023–2003 rel.	+20.1%	+6.5%	+18.1%	+5.6%	+10.0%	-6.6%
2023–2013 rel.	+7.1%	+2.3%	+6.9%	+2.1%	+5.7%	-0.5%
2013–2003 rel.	+12.2%	+4.1%	+10.6%	+3.4%	+4.0%	-6.1%
Maximum	16.70	67.42	20.43	71.73	4.14	5.44
Minimum	13.90	63.31	17.29	67.92	3.33	4.17
Range	2.80	4.11	3.14	3.81	0.81	1.27

Source: CZSO; authors' calculations.

Appendix 2 Differences between the *Proportion 65+* and the *Proportion RLE 15–*, NUTS 2 regions of Czechia (in percentage points).

Region/Year	PR	CB	SW	NW	NE	SE	CM	MS
2003	1.82	0.52	0.98	−0.47	0.93	1.45	1.44	0.38
2008	3.98	2.00	3.05	1.57	3.03	3.57	3.13	2.24
2013	6.52	3.84	4.69	2.78	5.30	5.07	4.47	3.27
2018	6.65	5.11	5.44	4.34	5.64	5.90	6.15	4.89
2021	5.34	3.14	3.46	1.56	3.51	4.58	3.53	2.89
2023	6.24	5.66	6.36	4.53	6.35	6.77	7.23	5.95

Source: CZSO; authors' calculations.

Appendix 3 Differences between the Ageing Index (*AI*) and the Prospective Ageing Index (*PAI*), NUTS 2 regions of Czechia.

Region/Year	PR	CB	SW	NW	NE	SE	CM	MS
2003	9.88	2.42	4.46	−2.06	4.15	6.53	6.46	1.64
2008	23.23	9.60	15.02	7.30	14.51	17.49	15.24	10.62
2013	37.10	18.31	24.02	13.66	26.45	25.98	23.03	16.63
2018	34.52	23.22	27.24	21.17	27.73	29.49	31.18	24.82
2021	26.58	13.76	16.90	7.54	16.90	22.05	17.46	14.35
2023	29.73	24.17	30.37	21.53	29.85	31.77	35.06	29.12

Source: CZSO; authors' calculations.

Appendix 4 Differences between the Old-Age Dependency Ratio (*OADR*) and the Prospective Old-Age Dependency Ratio (*POADR*), NUTS 2 regions of Czechia.

Region/Year	PR	CB	SW	NW	NE	SE	CM	MS
2003	3.36	0.99	1.83	−0.87	1.76	2.73	2.69	0.70
2008	6.92	3.64	5.53	2.82	5.54	6.51	5.73	4.08
2013	11.72	7.21	8.82	5.26	9.98	9.51	8.44	6.16
2018	12.65	10.24	11.01	8.84	11.59	11.80	12.36	9.80
2021	10.85	6.76	7.61	3.51	7.85	9.83	7.81	6.28
2023	12.27	11.76	13.43	9.78	13.68	14.16	15.35	12.58

Source: CZSO; authors' calculations.

Appendix 5 Differences in the Pensioners to Workers Ratio (*PWR*), NUTS 2 regions of Czechia, 2023–2003, 2023–2013, 2013–2003.

Region/Year	PR	CB	SW	NW	NE	SE	CM	MS
2013–2003	−0.05	−0.02	5.21	6.92	6.77	1.54	4.77	1.45
2023–2013	−5.18	−1.46	−1.04	−2.09	−1.06	−0.78	−1.14	−0.99
2023–2003	−5.23	−1.48	4.17	4.84	5.71	0.75	3.62	0.45
2013–2003 rel. (%)	−0.10	−0.04	9.86	12.82	12.33	2.67	8.37	2.37
2023–2013 rel. (%)	−10.88	−2.77	−1.80	−3.43	−1.72	−1.32	−1.85	−1.59
2023–2003 rel. (%)	−10.97	−2.80	7.89	8.95	10.40	1.31	6.36	0.74

Source: CZSO; authors' calculations.

Appendix 6 Differences in the Real Elderly Dependency Ratio (*REDR*), NUTS 2 regions of Czechia, 2023–2003, 2023–2013, 2013–2003.

Region/Year	PR	CB	SW	NW	NE	SE	CM	MS
2013–2003	−5.14	−2.99	−0.82	1.27	−1.45	−1.98	1.11	0.75
2023–2013	3.59	2.89	4.18	5.39	6.55	3.51	2.33	2.97
2023–2003	−1.55	−0.10	3.36	6.66	5.10	1.54	3.45	3.72
2013–2003 rel. (%)	−19.10	−10.58	−3.00	4.45	−5.21	−6.99	4.03	2.56
2023–2013 rel. (%)	16.51	11.45	15.71	18.09	24.79	13.36	8.13	9.84
2023–2003 rel. (%)	−5.75	−0.34	12.24	23.34	18.29	5.44	12.49	12.65

Source: CZSO; authors' calculations.

Abbreviations

EU = European Union
PR = Prague region (CZ01)
CB = Central Bohemia region (CZ02)
SW = South-west region (CZ03)
NW = North-west region (CZ04)
NE = North-east region (CZ05)
SE = South-east region (CZ06)
CM = Central Moravia region (CZ07)
MS = Moravian-Silesian region (CZ08)

Review of glacial lake inventories in the Sikkim Himalaya

Madhukar Srigyan*

Charles University, Faculty of Sciences, Department of Physical Geography and Geoecology, Czechia
* Corresponding author: srigyanm@natur.cuni.cz

ABSTRACT
This article aims to review the existing research associated with the preparation of glacial lake inventories in the Sikkim Himalaya, to understand their evolution over the past two decades and analyse them. The number and areal extent of glacial lakes has been increasing in the Sikkim region due to climate change and global warming in the Himalayas. This makes glacial lakes vulnerable to the occurrence of Glacial Lake Outburst Floods (GLOF). This review analyses the findings from 30 research studies of glacial lakes, including key parameters, such as temporal coverage, criteria for lake delineation (including minimum size thresholds), lake typology criteria, remote sensing data sources, and methodologies employed for glacial lake identification and also the GLOF analysis. The review highlights inconsistencies such as variations in data sources, automatic or semi-automatic mapping technique, minimum size criteria and need for expert interpretations. The study finds that Normalised Difference Water Index (NDWI) based identification of glacial lake is the most preferred technique with visual interpretation. It is also observed that the source of water for a glacial lake is the prominent typological criterion for both glacial lake mapping and GLOF analysis. This study highlights the possibility of preparing a comprehensive inventory of glacial lakes having at least 0.0036 km^2 area from 1962. Such inventory may further improve our understanding of glacial lake dynamics and facilitate an effective GLOF risk analysis in the Sikkim region.

KEYWORDS
Glacial lake inventory; Glacial lake outburst flood (GLOF); satellite data; typology of glacial lakes; proglacial lakes

Received: 2 June 2025
Accepted: 22 August 2025
Published online: 17 September 2025

Srigyan, M. (2025): Review of glacial lake inventories in the Sikkim Himalaya. AUC Geographica 60(2), 171–185
https://doi.org/10.14712/23361980.2025.14

1. Introduction

A lake is termed a glacial lake if and only if its origin is related to glacier action. Formally, glacial lakes are defined as the lakes whose origin is due to glacial processes, unlike other lakes, formed due to processes like tectonic, volcanic, landslide, fluviatile, and aeolian activities and organic material, which are usually found in the high-altitude regions of the mountain environment (Iturrizaga 2014). Several processes like glacier erosion, transport or accumulation, lead to the formation of the glacial lakes. Proglacial lakes are one of the prominent glacial lakes, which are formed mainly due to deglaciation, leaving an empty space to be filled as a lake or the merging of several supraglacial lakes at the terminus of the glacier, or glacial melt being collected at the depression created during the previous glacial extent maxima (ICE Age) (Huss et al. 2017). The high-altitude regions of the Himalayas and other glaciated mountain ranges, such as the Alps or Andes, are abundant with proglacial lakes. Under the influence of a warming climate and due to special geological and geomorphological preconditions (Emmer and Cochachin 2013), these proglacial lakes may become vulnerable to Glacial Lake Outburst Flood (GLOF). Therefore, understanding the past and present status of these proglacial lakes is the key to studying the process of their origin and evolution. The preparation of an inventory of glacial lakes in the region of Sikkim Himalaya is the prominent step to such a study.

Inventory is defined as an ensemble of information containing geographic location, area, length, orientation, elevation, and classification of the entity (World Glacier Inventory). Local geology, geomorphology and existing climatic conditions affect the formation and evolution of the glacial lakes, and therefore, they are crucial to be studied while preparing an inventory of glacial lakes. The Eastern Himalaya is a seismically active zone (Das et al. 2024), and the rate of temperature rise in its high-altitude (3–7 km) and low-altitude (0–3 km) regions is significantly higher in the last 45 years (Desinayak et al. 2023). This, combined with high-altitude (above 3000 m a.s.l.) proglacial lakes in the Sikkim Himalaya, makes them vulnerable to GLOF. Under the influence of climate change, the health of the glaciers is drastically deteriorating with the increase in the extent and number of proglacial lakes across the region (Mohanty and Maiti et al. 2021; Gaikwad et al. 2025). Kovanda (2024) analysed 2939 GLOFs (occurred across the world since the beginning of the 20th century) and highlighted that GLOFs occur most frequently in summer due to increased melting, with glacial-dammed lakes primarily affected by positive temperatures and moraine/bedrock-dammed lakes by heavy precipitation, especially during rainy seasons in regions like the Central Andes. A GLOF event can have a catastrophic impact including the loss of life, damage to the local ecology and property, and destruction of hydropower plants (Chowdhury et al. 2025). Preparation of inventories is a crucial first step to understand the present status of the proglacial lakes in the Sikkim Himalaya, and therefore, researchers have been actively studying this area and developing several inventories over several years (Tab. 1).

With the development of remote-sensing technology in recent decades, the monitoring of these glacial lakes and associated glaciers has become much easier. Further, the availability of several free satellite datasets has incentivised researchers to map such remote glacial lakes, which would be difficult to monitor, as they are difficult to access physically. To date, several inventories of glacial lakes in the Himalayan region have been prepared using several satellite images (Ives et al. 2010; Campbell and Pradesh 2005; NWIA/SAC/ISRO 2010 and 2012; Govindha Raj et al. 2013; Worni et al. 2013; Schmidt et al. 2020; Das et al. 2024; Mohanty and Maiti 2021; Chen et al. 2021; Gaikwad et al. 2025; Song et al. 2025). These inventories provide important data for several research objectives, including GLOF analysis (Aggarwal et al. 2017; Mohanty and Maiti 2021). Several factors that vary include mapping methodology, the use of satellite images, expert knowledge or interpretation capability, and their purpose, like GLOF analysis (Mohanty and Maiti 2021) or glacial lake extent change monitoring (Banerjee and Bhuiyan 2023). The time series is the most prominent criterion on which most of these datasets vary (Mohanty et al. 2023; Kaushik et al. 2024; Gaikwad et al. 2025). Therefore, the main aim of this study is to understand the evolution of glacial lake inventory in the Sikkim region over the past 25 years and simultaneously assess their quality and limitations.

2. Study area

Sikkim lies in the northeastern region of India. Geographically the region lies above sea level from ~280 m up to 8586 m (Kumar and Sharma 2023) and ~ 40% of the total state area is cryosphere, consisting of glaciers, glacial lakes and permafrost (Kumar and Sharma 2023) (Fig. 1). The region has nearly 100 glaciers, having different sizes, among which Zemu Glacier (~26 km) is the largest glacier (Kumar et al. 2020). Teesta River is a major tributary of Brahmaputra River and originates in the northeastern corner of Sikkim at Chhombo Chhu glacial lake and Khangchung Chho glacial lake, which are situated at an elevation of 5280 m. The Teesta river system comprises several 7 tributaries; among them, Rangit is the major tributary (Abdul Hakeem et al. 2018; Sharma et al. 2019). These rivers are fed with mostly precipitation from the southwest monsoon, along with snow and glacier melt. Several dams have been built on these rivers for hydropower generation and irrigation purposes. With the change in climatic conditions, the retreat of glaciers and potential occurrences of GLOF events pose a threat to these river systems.

Sikkim Himalaya has young folded mountains of the Himalayan system with rugged terrain and steep slopes (~43% area). Lithology of the area broadly suggests three major rock constituents, which are half-schistose, gneissose and Precambrian rocks like schists and phyllites (NWIA/SAC/ISRO 2010). The slopes of this region are very prone to erosion and weathering due to the presence of schists and phyllites. Such a possibility of soil erosion makes the area susceptible to landslides. The region is also sensitive to seismic activity (seismic active zone IV), which is high in terms of earthquake intensity (Aggarwal et al. 2017; Gaikwad et al. 2025), as it is situated in the young Himalayan mountain.

Sikkim's climate varies dramatically with altitude, mostly influenced by the southwest monsoon, having nearly 35–40% snow cover even in summer with limited influence of westerlies and easterlies (Kumar et al. 2020; Basnett et al. 2013). The precipitation season largely extends from May to October, covering summer and monsoon. Heavy monsoon rains usually happen in July and August, which may trigger landslides in the Sikkim region. While most inhabited areas enjoy a temperate climate with mild summers and cool winters, the high-altitude north experiences prolonged snowfall and freezing temperatures. Kumar et al. 2020 suggest accelerated warming in the last two decades by analysing the minimum temperature.

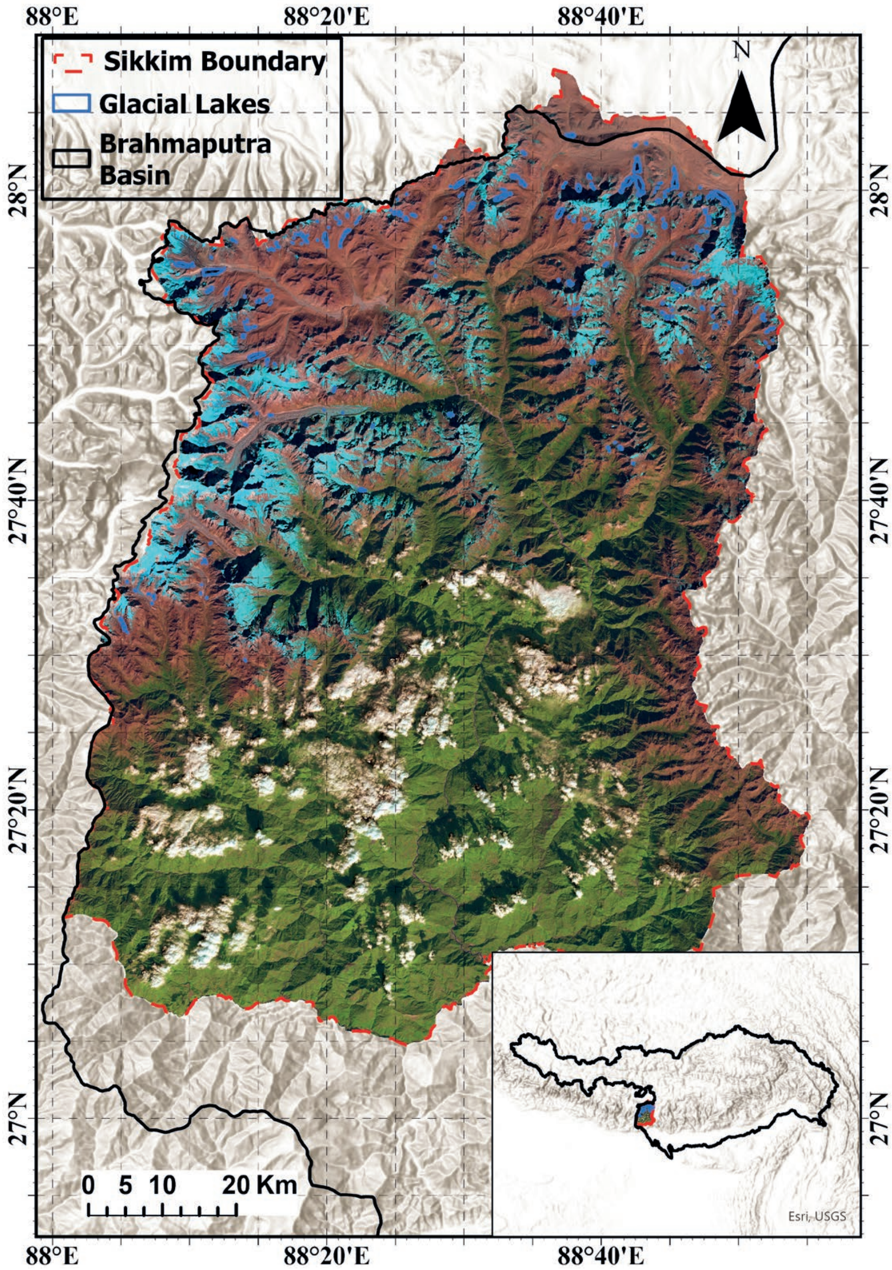

Fig. 1 The study area is situated in the Sikkim Indian Himalayan region, with the large number of glacial lakes, distributed across the glaciated region, which is displayed on the background image of False colour composite of Landsat data.

Precipitation does not show any particular trend from 1961 to 2017 (Kumar et al. 2020).

3. Methodology and data

Several studies account for glacial lake monitoring to understand their spatial and temporal variation, and therefore, preparation of an inventory is critical to the study. A basin-based research focused towards GLOF analysis also ends up preparing an inventory of the glacial lakes, including all the parameters associated with an inventory. Therefore, in the current study, the main focus was to identify the studies related to glacial lakes carried out from the early 2000s and filtered them on the criterion like the location of the study, i.e., Sikkim region or Teesta basin region, inventory preparation, glacial lake change study and GLOF analysis. With the rise of the threat of GLOF occurrence under the influence of climate change, the study of glacial lake monitoring of Sikkim region was initially picked up by some big organisations like International Centre for Integrated Mountain Development (ICIMOD) (Ives et al. 2010), Aisa-Pacific Network (APN) (Campbell and Pradesh 2005) and Space Applications Centre (SAC) (NWIA/SAC/ISRO 2010 and 2012) during the early 2000s. In later decades of the 2000s, mapping and monitoring of glacial lakes have been widely carried out by several researchers across the Himalayas, including Sikkim region (Tab. 1). The compiled research includes categorisation on the basis of mapping strategies like the use of satellite images, strategies for mapping involving automatic, manual or semi-automatic methodology, the typology of the glacial lakes for glacial lake extent change study or GLOF analysis, along with the methodology followed for GLOF analysis. Consideration of the minimum area for identifying glacial lakes and further considering them for GLOF analysis is also analysed in this study. The study highlights the factors affecting the GLOF analysis utilised by these researchers. Since temporal variation of the glacial lake extent significantly affects GLOF analysis and therefore has also been included in this study along with the years of mapping (Tab. 1).

4. Review of lake inventories

The current study thoroughly investigates the previous 30 such studies (Tab. 1) and highlights a comprehensive overview of these studies focusing on glacial lake inventories and GLOF analysis as well as glacial lake extent change in the Sikkim Himalaya. Nearly half (~14) of the inventories have been prepared for the world (Shugar et al. 2020; Song et al. 2025), High Mountain Asia (HMA) (Chen et al. 2021; Zhang et al.

Tab. 1 Since the early 2000s, the researchers have been preparing a glacial lake inventory focusing on the Sikkim region or the whole/partial Himalayan region for analysing glacial lake extent change or GLOF analysis. [*Publication: Red colour is for across the Indian Himalayas, and the rest are limited to only Sikkim region, Purpose of the study: Broad purposes of the study are (i) Inventory (Cyan), (ii) GLOF Analysis (Red) and (iii) Lake monitoring and analysis (Purple), Area: Inventories with minimum area threshold (Cyan or otherwise), Broadly the Typology is categorised in five classes based on various criteria (legend similar to Fig. 4).*] * lake created after the occurrence of a landslide based blocking of the flowing stream, ** land that is temporarily or permanently covered with water, *** depression filled with water and has less area, and **** glacial lake with no direct attachment to the glacier terminus.

Publication	Time series	No. of lakes	Purpose of the study	Area (km^2)	Studied lakes and their typology (Category of lakes)
Ives et al. 2010	2000	266	Inventory	0.10000	Glacial lakes
Campbell and Pradesh 2005	2000–2003	266	Inventory	0.10000	Glacial lakes: Erosion, Valley trough, Cirque, Blocked*, Moraine Dammed (Lateral Moraine and End Moraine Dammed lakes), and Supraglacial lakes.
NWIA/SAC/ISRO 2010	2005	359	Inventory	0.02250	Wetlands**: High altitude Wetlands (elevation > 3000m) 3000–4000m: 10 lakes 4000–5000m: 130 lakes >5000m: 119 lakes
NWIA/SAC/ISRO 2012	2006–2008	534	Inventory	0.02250	Wetlands**: High altitude Wetlands (elevation > 3000m) 3000–4000m: 6 lakes 4000–5000m: 323 lakes >5000m: 205 lakes
Govindha Raj et al. 2013	2005–2010	320	Inventory	0.10000	Glacial lakes: Moraine-dammed lakes, Blocked, Valley
Worni et al. 2013	2000–2002	8	Inventory	0.01000	Glacial lakes: Moraine-dammed lakes; Ice-dammed lakes
Zhang et al. 2015	1990, 2000, 2010	98	Inventory and GLOF analysis	0.00270	Glacial Lakes: Glacier-fed and Non-glacier-fed
Abdul Hakeem et al. 2018	2000, 2007, 2014	644	Inventory	0.00100	Glacial lakes: cirque lake, debris dammed lake, end moraine dammed lake, lateral moraine dammed lake, other lake fed by glacial melt, supraglacial lake and water body***.

Publication	Time series	No. of lakes	Purpose of the study	Area (km²)	Studied lakes and their typology (Category of lakes)
Aggarwal et al. 2 017	1975, 1985, 1995, 2005, 2015	1104	Inventory and GLOF analysis	0.00100	Glacial lakes: Moraine-dammed, bedrock-dammed, combined dam, ice-dammed lakes, and non-specified lakes. Classified on the basis of Quantitative characteristics (Latitude, longitude, lake level elevation, lake area, Lake depth, lake volume, dam geometry and freeboard measurements) and Qualitative characteristics (lake type, watersheds, and direct connection with glacier).
Debnath et al. 2018	1988, 2001, 2014	68	Inventory	0.00300	Glacial lakes: Supraglacial, Ice scoured, Moraine-dammed, Proglacial lake and Tarn
Shukla et al. 2018	1975–2017	466	Inventory	0.00360	Glacial lakes: Supraglacial, pro/peri glacial lakes in contact with glacier, pro/peri glacial lakes NOT in contact with glacier and other glacial lakes
Maharjan et al. 2018. (ICIMOD)	2004–2007	401	Inventory	0.00300	Glacial lakes: 1) Moraine dammed lake: (a) End moraine (b) Lateral moraine (c) Other moraine 2) Ice dammed lake: (a) Supraglacial lake (b) Glacier Ice dammed lake 3) Bedrock dammed lake: (a) Cirque (b) Other bedrock-dammed lakes 4) Others
Garg et al. 2019	1991–2017	17	Glacier monitoring and analysis	0.00360	Glacial Lakes: 9 Proglacial lakes and 8 Supraglacial lakes
Shugar et al. 2020	1990–2018	48	Inventory	0.05000	Glacial lakes: area and elevation based classification
Wang et al. 2020	1990–2018	253	Inventory	0.00540	Glacial lakes: Glacier-fed and non-glacier-fed-lakes, supraglacial lakes, ice-contacted lakes and ice-uncontacted lakes****
Mohanty and Maiti 2021	1990–2019	165	GLOF analysis	0.01010	Glacial lakes: Glacier-fed, connected, moraine-dammed lake
Chanda and Biswas 2021	1990, 2000, 2017	282	Inventory and GLOF analysis	0.00005	Glacial lakes: Distance from the glacier, Growth of lakes, Elevation and Slope
Chen et al. 2021	2008–2017	242	Inventory	0.00810	Glacial lakes: Classified according to their position relative to the parent glacier or their formation mechanisms. Proglacial lakes, Supraglacial lakes, Unconnected lakes, Ice marginal lakes
Islam and Patel 2022	2000–2018	354	Inventory and GLOF analysis	0.00100	Glacial lakes: Area based classification
Verma and Ramsankaran 2022	2020	419	Inventory	0.01000	Glacial lakes
Banerjee and Bhuiyan 2023	1987–2020	406	Inventory and analysis	0.01000	Erosional lake, Supraglacial lake, Proglacial lake and Blocked lake, Area
Agarwal et al. 2023	1975–2017	14	Inventory and analysis	0.00500	Proglacial lakes; Area based categories: small: <0.2 sq. km, medium: 0.2–1 sq. km and large: >1 sq. km
Mohanty et al. 2023	1990, 2000, 2010, 2015, 2020	382	Inventory	0.01000	Glacial lakes: Glacier-fed, connected, moraine-dammed lakes
Kumar and Sharma 2023	1988–2018	12	Glacier monitoring and analysis	None	Glacial Lakes: Supraglacial lakes and proglacial lakes
Zhang et al. 2023	1990–2020	208	Glacier mass balance analysis	0.00360	Glacial lakes: Proglacial lakes, Supraglacial lakes and Non-ice contact glacial lakes
Kaushik et al. 2024	2015, 2017, 2020	231	Inventory and GLOF analysis	0.01000	Glacial lakes: Moraine dammed lake
Das et al. 2024	1990–2023	47	GLOF analysis	0.01000	Glacial lakes including Proglacial lakes, Lake area and volume
Gaikwad et al. 2025	1990, 2000, 2010, 2020	440	Inventory and GLOF analysis	0.02000	Glacial lakes: Based on dam formations: moraine-dammed, bedrock-dammed, ice-dammed, and other lake types
Song et al. 2025	1984–2020	244	Inventory	0.01000	Glacier Fed or Non-glacier Fed
Kumar et al. 2025	1990, 2000, 2010, 2020	363	Glacier Monitoring and analysis	0.01000	Moraine-dammed: End moraine-dammed lakes [M(e)], Lateral moraine-dammed lakes [M(l)], and other moraine-dammed lakes [M(o)], Bedrock-dammed lakes were divided into two categories: cirque lakes [B(c)] and other bedrock-dammed lakes [B(o)], and ice-dammed lakes: Supraglacial (S) and dam med by tributary glaciers [I(v)]

2023, etc.) or the partial Himalayan region (Ives et al. 2010, NWIA/SAC/ISRO 2012, Mohanty et al. 2023), including the Sikkim region. These inventories include all types of glacial lakes, including supraglacial as well as proglacial lakes, and they are categorised as per the user's requirement and their understanding.

4.1 According to time span

Time series analysis of the inventories prepared by several researchers highlights that a few inventories have been prepared since 1975 (Shukla et al. 2018; Agarwal et al. 2023). However, the majority of the inventories have been prepared from around 1990 (Zhang et al. 2015; Garg et al. 2019; Mohanty and Maiti 2021; Chanda and Biswas 2021; Zhang et al. 2023; Das et al. 2024; Gaikwad et al. 2025), including a few from 1988 (Kumar and Sharma 2023; Debnath et al. 2018) and 1987 (Banerjee and Bhuiyan 2023), covering a duration of nearly 30 years. In recent years, researchers have tried to look for the latest 20 years of the lake's extent change and therefore several inventories were prepared based on the early 2000s satellite data (Ives et al. 2010; Campbell and Pradesh 2005; Worni et al. 2013; Abdul Hakeem et al. 2018; Maharjan et al. 2018; Islam and Patel 2022) (Fig. 2).

The largest span for studying the Sikkim region was carried out for 40 years (Aggarwal et al. 2017). Another study by Agarwal et al. (2023) covers a temporal span of 42 years and spatially the whole of the Himalayas but has only 14 lakes in the Sikkim region. Three inventories have been prepared on the basis of a single year, each having a reference year as 2000, 2005 and 2020. Some inventories were prepared on the basis of a few years (i.e., combining 2–3 years) of data, the majority in the early 2000s. Several inventories were also prepared for the purpose of glacial lake extent change, covering nearly 5 to 10 year interval, among them, the majority were prepared later in the 2000s. Chowdhury et al. (2021) have studied only 4 glacial lakes of Sikkim region covering a time span from 1962 to 2018 for the purpose of GLOF analysis. Racoviteanu et al. (2015) studied the change in glaciers from 1962 to 2006 for Sikkim region, which suggests the possibility of mapping the whole Sikkim region and analysing over a large temporal range.

Several studies were carried out for a periodical analysis of the glacial lake extent change, inventory preparation or GLOF analysis, like Aggarwal et al. (2017), Mohanty et al. (2023) and Gaikwad et al. (2025), who carried out a decadal inventory; Debnath et al. (2018) with a period of 13 years; and Kaushik et

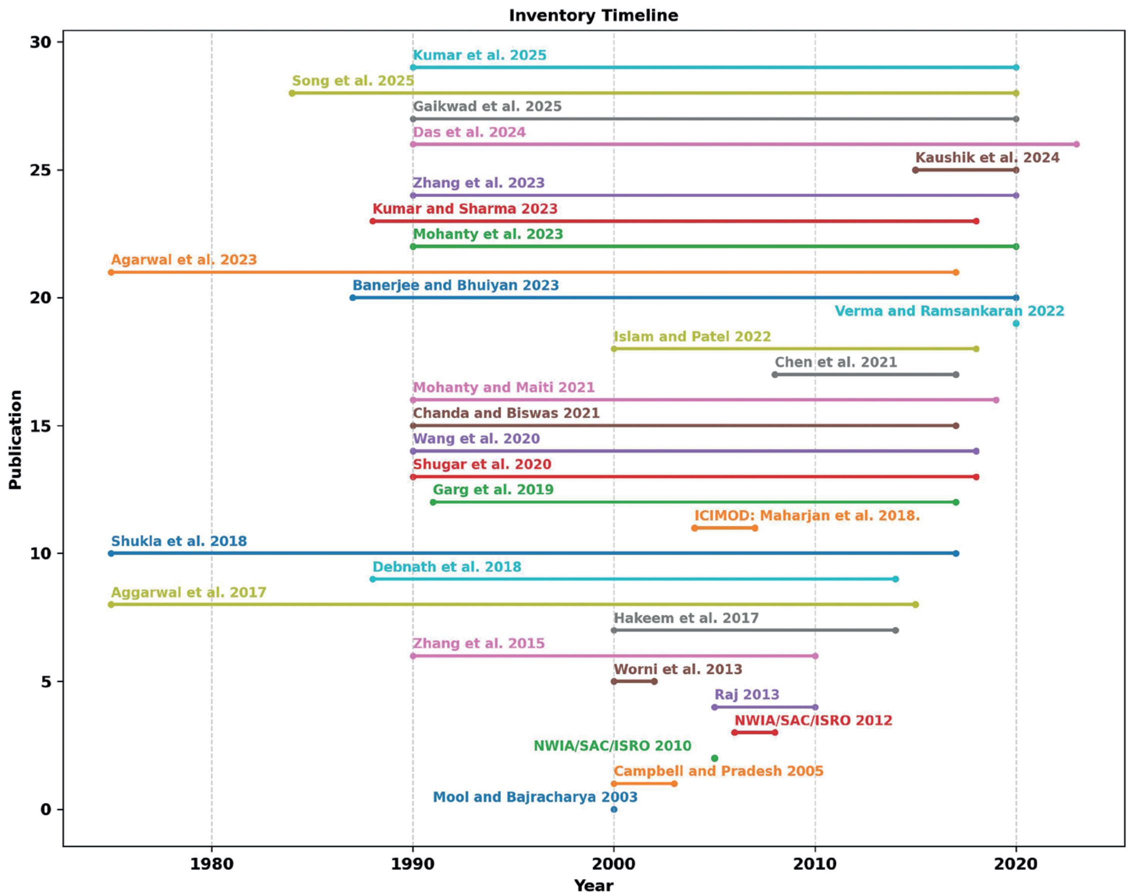

Fig. 2 The inventories were prepared by various researchers for monitoring and GLOF analysis of glacial lakes using satellite datasets, earliest from 1975 to recent, in the Sikkim Himalayan region.

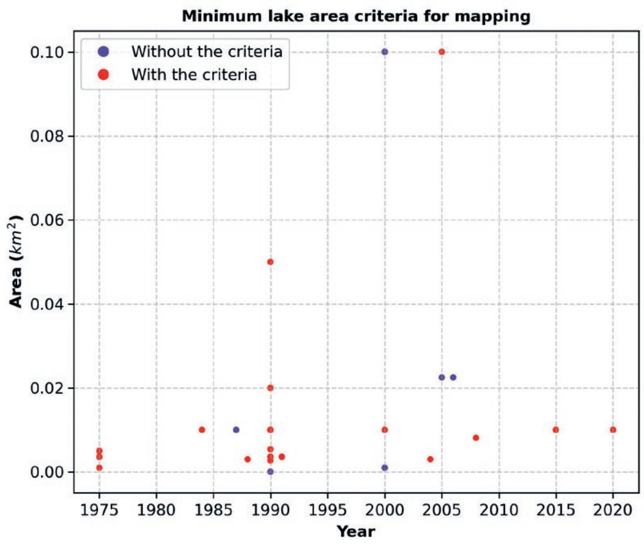

Fig. 3 The variation of the different minimum area criteria adopted by researchers for glacial lake mapping.

al. (2024) with an interval of nearly two years. Hazra and Krishna (2022) studied just one lake, Shako-Cho lake for its GLOF analysis but highlighted the possibility of finding a suitable dataset for mapping for the years 2000, 2005, 2009, 2015 and 2018. Shugar et al. (2020) and Song et al. (2025) have provided a global dataset of glacial lakes mapped from 1990 to 2018 and from 1984 to 2020, respectively.

4.2 Size of lakes and mapping uncertainty

Glacial lakes are of different sizes due to their varying formation stages, and among them, supraglacial lakes are mainly smaller than the proglacial lakes. In this study, we found that generally, the inventories are prepared with a minimum area-based criteria to map the glacial lakes (Tab. 1). It has also been observed that there is a large variation in the minimum area criteria, like 0.0036 km^2 (Garg et al. 2019), 0.0054 km^2 (Wang et al. 2020), 0.0081 km^2 (Chen et al. 2021) and 0.01 km^2 (Mohanty and Maiti 2021), in spite of utilising the same reference Landsat series data. The Landsat series is the most common satellite sensor used by the researchers for glacial lake mapping, irrespective of the purpose of their study (Tab. 1). Majorly, the researchers prefer the threshold to map glacial lakes to be <= 0.01 km^2 (Fig. 3). Researchers who have prepared their inventory not limited to Sikkim region follow a minimum area criteria, which range from 0.0027 km^2 (3 adjoining pixels of 30m) to 0.1 km^2, except two inventories (NWIA/SAC/ISRO 2010 and 2012).

4.3 Typology of lakes

In this study, the typology of the glacial lakes is usually based on the lakes' position relative to the parent glacier, which are namely, supraglacial lakes, proglacial lakes, unconnected lakes and ice marginal lakes and based on dam formation, which are namely, moraine-dammed, bedrock-dammed, and ice-dammed (Tab. 2) (Maharjan et al. 2018). Mapping of glacial lakes in the Sikkim region has been mainly carried out for the purpose of preparation of inventory (19 studies) and GLOF analysis (8 studies) (Fig. 4). The studies with the purpose of inventory have often defined the typology of the glacial lakes on the basis of the process of formation (Campbell and Pradesh 2005; Govindha Raj et al. 2013; Worni et al. 2013; Abdul Hakeem et al. 2018; Debnath et al. 2018; Maharjan et al. 2018) followed by "water source" to the lake (Shukla et al. 2018; Wang et al. 2020; Mohanty et al. 2023; Song et al. 2025) (Fig. 3). Similarly, in the case of GLOF analysis, the main typology criterion is the "process of formation" (Aggarwal et al. 2017; Kaushik et al. 2024; Gaikwad et al. 2025), followed by "water source" (Zhang et al. 2015; Mohanty and Maiti 2021), and then the relative position of the glacial lake with the nearest glacier (Das et al. 2024). Lakes are often classified on the basis of physical characteristics (like elevation or area of the lake) for purposes like preparing a glacial lake inventory as well as GLOF analysis (Tab. 1), but here two early inventories adopted physical characteristics like elevation and area of the lake as the prominent typology criteria (NWIA/SAC/ISRO 2010 and 2012). Two inventory preparation works (Ives et al. 2010; Verma and Ramsankaran 2022) did not follow any typology while mapping the glacial lakes. Erosional processes in the bedrock often lead to the formation of piedmont and cirque lakes where

Tab. 2 The typology of glacial lakes based on various criteria (modified based on Maharjan et al. 2018).

Typology criteria	Lake type	Lake sub-type
Process of formation	Moraine dammed lake	End-moraine dammed lake
		Lateral moraine dammed lake
		Other moraine dammed lake
	Ice dammed lake	Supraglacial lake
		Glacier Ice dammed lake
	Bedrock dammed lake	Cirque lake
		Other bedrock-dammed lake
	Combined lakes	
Source of water to lake	Glacier Fed	
	Not Glacier Fed	
Position relative to glacier	Supraglacial lake	
	Proglacial lake	
Physical characteristics of the lake	Area: Small, medium and large glacial lakes Altitude: High and low altitude glacial lakes	

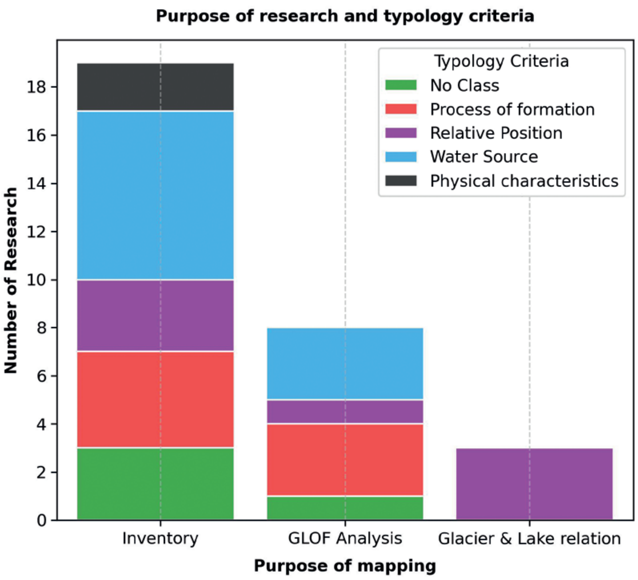

Fig. 4 The use of typology criteria for categorising glacial lakes vary significantly based on the purpose of mapping undertaken by various researchers.

glaciers were earlier present, which is also considered for the classification of the glacial lakes (Campbell and Pradesh 2005; Banerjee and Bhuiyan 2023). GLOF analysis many a time utilises multiple typologies to understand the vulnerability of lakes (Chen et al. 2021; Gaikwad et al. 2025).

4.4 Data selection and methodology of glacial lake identification

The selection of data for mapping is crucial for mapping precise glacial lake extent, which largely depends on the snow and cloud-free conditions, which usually happen at the end of the ablation period, i.e., September to December in the eastern part of the Himalaya (Mohanty et al. 2023). However, there is a small variation in the selection of reference data, which further leads to variation in the lake area. For example, the glacial lake extent for the year 1990 varies in the selection of date as well as numbers because Wang et al. (2020) selected the date roughly in between June and November, 1990; Zhang et al. (2015) considered 5th November, 1990; Zhang et al. (2023) considered 23rd September, 1992 and Mohanty et al. (2023) further added some lakes polygons over the mapped lakes by Zhang et al. (2015) using a satellite dataset from September to December of the year 1990, and Kumar et al. (2025) considered the ablation season of 1990 as the reference layer (Fig. 6). Preparation of an inventory requires precision in the delineation of the lake boundary, which can be achieved by several methodologies. The Normalised Difference Water Index (NDWI) has been quite often utilised for mapping of glacial lakes; however, the final lake extent is only achieved after manual correction using visual interpretation. Band ratio has also been utilised for highlighting the water bodies and further manually corrected for analysis (Campbell and Pradesh 2005; NWIA/SAC/ISRO 2010, 2012; Worni et al. 2013, etc.). This can be further analysed for the purpose of glacial lake monitoring/extent change analysis as well as GLOF analysis. The utilisation of manual or semi-automatic techniques are entirely dependent on the dataset used and the researchers expertise. The inventories utilise the earliest available satellite datasets, which are Hexagon KH-9 and Landsat 4 (MSS sensor). Corona, Hexagon KH-9 data have grayscale images and therefore it is only possible to map the boundary of the lakes with visual interpretation (Chowdhury et al. 2021; Schmidt et al. 2020). The Landsat series has one of the largest time spans (from 1982 to ongoing) of the data acquired over the study area and is therefore widely utilised for such research studies (Tab. 1) with a consistency of spatial resolution of 30 m.

4.5 GLOF analysis: adopted methodology and parameters

Glacial lakes mapped for either inventory or GLOF analysis always have a few general associated parameters, like the location of the lake (latitude and longitude), altitude, area and perimeter (Ives et al. 2010; Zhang et al. 2015; Chanda and Biswas 2021; Gaikwad et al. 2025) (Tab. 3). GLOF analysis is the most important criterion for mapping or preparing an inventory of these glacial lakes in Sikkim region. Tab. 3 summarises the uses of various parameters in GLOF analysis. Broadly, the various parameters which are affecting the occurrence of GLOF include morphological parameters of the lake and associated glacier, the lake's dam characteristics, physical conditions of the surrounding area in terms of occurrence of landslides, snow avalanches, and seismic activity, lake area change, the lake's proximity to the glacier, the lake's shape index highlighting the range of ellipticity from irregular shape to a circle (Kaushik et al. 2024), topographic potential suggesting the potential of occurrence of ice and/or rock avalanches (slope > 30°) (Kaushik et al. 2024), and downstream slope. Morphological parameters like area, perimeter, length and area change of glaciers and glacial lakes, along with glacier dams, are derived from optical data sets. The elevation, slope, aspect of the lake and glacier, and freeboard dimension of the dam, including height, are derived using a Digital Elevation Model (DEM) like SRTM, ASTER, Cartosat and ALOS PULSAR.

Analytical Hierarchy Process (AHP) is one of the most widely utilised methodologies for assigning the weights of the aforementioned parameters following another methodology, Fuzzy Analytical Hierarchy Process (FAHP). Islam and Patel (2022), and Worni et al. (2013) utilised the weights derived from other researchers for GLOF analysis. Threshold-based techniques have also been popular to identify potentially

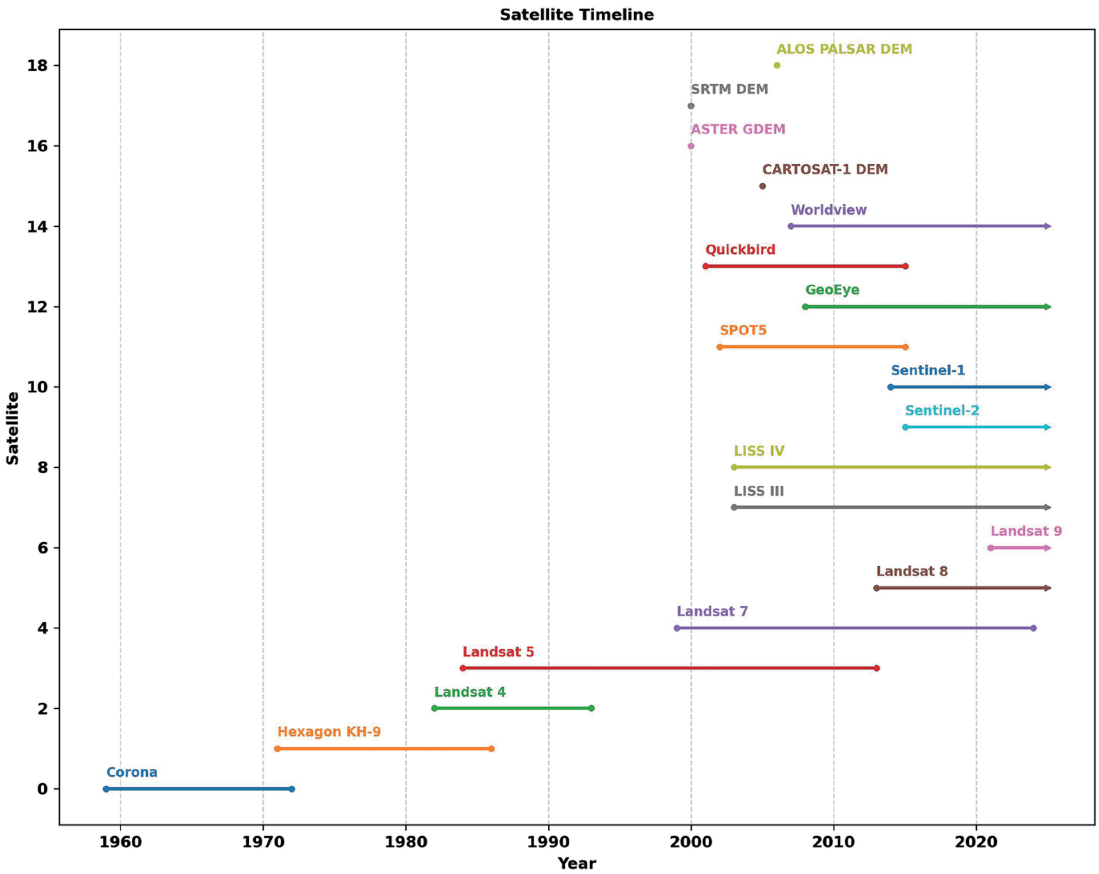

Fig. 5 Availability of satellite data for mapping of glacial lakes in the Sikkim region over past 60 years. Ongoing satellite missions are marked with an arrow.

Tab. 3 GLOF analysis and associated parameters utilised in the Sikkim region.

Publication	Methodology for GLOF Analysis	Parameters
Campbell and Pradesh 2005	Potentially dangerous lakes identification	Moraine dams, Rise of lake water level, Activity of supraglacial lakes, Position of lakes relative to glaciers, Glacial lake size, Dam conditions, Conditions of Associated Mother Glacier, Physical conditions of the surrounding area
Worni et al. 2013	Risk analysis using weight assignment	Dam type, Dam geometry, Freeboard and Potential for lake outburst impacts
Aggarwal et al. 2017	AHP	Area, Elevation, Length, Width, Volume, Type of the lake, Orientation of the lake, Freeboard level, Dam width-height ratio, Activity of the lake, Drainage type of the lake, Slope of the moraine walls, Growth in the lake area, Distance from the parent glacier, Distance of settlement, Geomorphology, like physical conditions of surroundings, and Seismic and tectonic factors
Chanda and Biswas 2021	Threshold based categorisation to identify hazardous lakes	Minimum elevation, Slope threshold, Distance from connecting stream or outlet, Distance from settlement area, Distance from the hydroelectric power projects, Growth of lakes, Distance from the glacier
Mohanty and Maiti 2021	AHP and Object Based Classification method	Lake size, Area change rate, Volume, Elevation, Proximal distance, Lake type, Lake aspect, Glacier calving frontal width, Glacier snout steepness, Avalanches, Landslide, Earthquake, Lake freeboard, Dam steepness, Dam height
Islam and Patel 2022	Weight assignment	Change in lake area, Slope difference between glacial snout and lake outlet, Connection/Proximity to Parent Glacier, Connection with river channel, Height of Dam, Availability of rocks for avalanches/rockfall, Distance from the basin outlet
Kaushik et al. 2024	AHP	Rate of lake change, Lake area, Downstream Slope, Total watershed area, Shape index, Topographical potential, Rate of lake change
Das et al. 2024	AHP and Fuzzy AHP	Glacial Lake Volume, Seismic Activity, Glacial Lake Area, Elevation, Avalanche, Rockfall, Slope, Distance from River, Rainfall
Gaikwad et al. 2025	FAHP	Volume of lake, Area change rate, Area of the mother glacier, Slope between lake and snout of the mother glacier, Distance between glacier and lake, Average Steep Lakefront Area angle

dangerous/hazardous lakes using several morphological and terrain parameters (Campbell and Pradesh 2005; Chanda and Biswas 2021).

5. Discussion

5.1 According to time span

There are several satellite data available now, and therefore, more periodic analysis are possible. The decadal changes from 1990 till 2020 are very common, but in order to understand the evolution of the glacial lakes, the year of formation is very crucial for the study. Such studies have widely exploited the Landsat series data (Tab. 1). The uses of high-resolution images are necessary to identify the various parameters of the lake dam and further improve the accuracy of the lake's extent. Similarly, DEM data is also important to derive the lake's dam parameters like slope, elevation, etc., and SRTM has been widely utilised for such studies. The sensors with high spatial resolution are advantageous for mapping small glacial lakes. Since several Corona (KH-4) images have been declassified and are available for Sikkim region, the time span for monitoring these glacial lakes can be increased. Chowdhury et al. (2021) have performed GLOF analysis for four glacial lakes in the Sikkim region and utilised the available 1962 data. Such analysis with a longer time span enhances the GLOF analysis. Another benefit of KH-4 data is that it is collected in high resolution, and therefore smaller glacial lakes can also be identified, and their evolution over recent years can also be effectively analysed. Schmidt et al. (2020) utilised the Corona images of 1969 to prepare an inventory of glacial lakes in the Ladakh region (Trans-Himalayan region), highlighting their further utility in GLOF analysis.

5.2 Size of lakes and mapping uncertainty

With the formation of glacial lakes and further undergoing evolution, the size can vary from a few square meters to several thousand square meters. The mapping of glacial lakes highly depends on the spatial resolution of the available satellite data. Gardelle et al. (2010) suggest that the minimum area of a glacial lake can be mapped using a satellite image which covers at least 4 pixels and satisfies the condition of the presence of at least one pure water pixel (i.e., a pixel whose reflectance is homogeneous and represents only one endmember). Several research studies carried out after 2011 have still been using either 9 pixels as a minimum size criteria, or some have no threshold, suggesting inconsistency in such a wide range of prepared datasets (Chen et al. 2021). Chen et al. (2021) suggest that at least 9 pixels should be considered for a lake polygon to define while keeping the area error under 50%. There is a variation in extent and/or number of mapped glacial lakes for the year 1990 by Zhang et al. 2015, Wang et al. (2020), Mohanty et al. (2021), and Zhang et al. (2023), as their date of reference satellite data may vary. Another reason for such variation may be due to the minimum size criteria. Zhang et al. (2023) establish that the maximum area of a lake can be up to 200 km^2.

While mapping the glacial lake, the mapping uncertainty is induced. The estimation of mapping uncertainty is dependent on the methodology followed along with the pixel's spatial resolution and other qualities of the available satellite data. Theoretically, the maximum area error in a glacial lake boundary is the half area of the boundary pixels because pure lake water body pixels are usually surrounded by mixed pixels (Zhang et al. 2015, Debnath et al. 2018, Shukla et al. 2018, Wang et al. 2020, Agarwal et al. 2023, Mohanty et al. 2023). The mapping error is inherent and specific to the spatial resolution of the satellite image and may be estimated using equation 1 (Zhang et al. 2015).

$$\%\ Mapping\ uncertainty = \frac{P * R * R * 100}{R * A * 2} \qquad (1)$$

where,
P is the perimeter and A is the Area of the lake,
R is the spatial resolution of the satellite image.

The factor of 0.678 is sometimes multiplied by (1) to estimate the area uncertainty of an individual lake within one standard deviation (σ) (Song et al. 2025). A few of the researchers, like Gaikwad et al. (2025), Banerjee and Bhuiyan (2023), Chen et al. (2021), and Govindha Raj et al. (2013), have considered an uncertainty of 1 pixel on the edges of the glacial lake as the mapping uncertainty.

Considering the availability of high-resolution satellite images in recent times, the minimum size criteria can vary for the preparation of an inventory. However, while doing long-term time series analysis, the inventories are compared with one another, which may or may not be developed on the same resolution of the satellite images. While performing a glacial extent change analysis over a period, each glacial lake is considered as an entity and further analysed accordingly (Shukla et al. 2018). Considering the recent availability of higher resolution images, the precision of mapping is increasing but has a basic limitation in performing a time series analysis, as we have the majority of coarser resolution data. The comparative analysis should always include the error associated with mapping accuracy for both the conditions, same resolution based mapping and different resolution-based mapping.

5.3 Typology of lakes

Researchers choose the typology of lakes based on the purpose of analysis. A wide range of regional or local scale glacial lake extent change analysis or GLOF analysis requires preparation of an inventory. Such inventories may vary in attributing the various typologies of lakes, considering the expertise of researchers as well as the purpose of their limited study, which may include prominently the process of formation and physical characteristics (Tab. 2). GLOF based analysis specifically requires a few other important typologies of lakes to better understand the process, like the relative position of the glacial lake with the nearest glacier, the source of the water, dam type, etc.

Although such studies on glacial lakes have been carried out for many years, there is a lack of a standardised typology of lakes. Specifically, under the current warming climatic scenario, the increase in the occurrence of GLOF across the world requires inventories built with the necessary typology, which further helps in understanding the evolution of the lakes. The typology of a lake is based on the process of formation, position relative to glacier, source of water to the lake and physical characteristics of the lake, which explains the process of evolution of the glacial lakes, formation of the lakes and future of those lakes under various GLOF scenarios. Same type of lake can be categorised on the basis of various typologies of lakes, like proglacial lakes, which can be formed due to the damming of lateral moraines, and their source can be the glacier melt (Shukla et al. 2018; Wang et al. 2020). Similarly, cirque lakes are erosional lakes which are bounded by a bedrock dam and have the least vulnerability in terms of dam breaching (Campbell and Pradesh 2005; Abdul Hakeem et al. 2018; Maharjan et al. 2018). Considering such parameters simultaneously will be beneficial for GLOF analysis. Supraglacial lakes, which are mostly ice dammed lakes and occur on the glacier, are becoming crucial for GLOF analysis, and under current warming conditions, these are expanding at a large rate in the Himalayas (Mohanty and Maiti 2021). In some cases, several supraglacial lakes at the terminus of the glacier merge to form a proglacial lake; therefore, they are crucial in understanding the evolution of such proglacial lakes. Tab. 2 includes the major typology of lakes, highlighting the strength of the dams, which is an important criterion for GLOF analysis (Kaushik et al. 2024; Das et al. 2024; Gaikwad et al. 2025; Mohanty and Maiti 2021).

5.4 Data selection and methodology for identification

Correct identification and delineation of glacial lakes depend on the spatial resolution of the satellite data, along with snow and cloud-free conditions. Sikkim experiences such favourable conditions at the end of ablation, usually during September–December (Mohanty et al. 2023). Satellite-based investigation helps in the derivation of these parameters, but limits their accuracy due to their spatial resolution. LANDSAT provides the longest time series of available data for mapping the lake extent. The availability of KH-4 images at high resolution (~2.5 m) (Chowdhury et al. 2021) over Sikkim region provides an opportunity to make an inventory of glacial lakes with the largest possible time span (~60 years). DEM-based derived parameters like slope, aspect and elevation information for a glacier or lake's dam are crucial for GLOF analysis. Fujita et al. 2008 assessed the utility of SRTM and ASTER DEMs for glacial lake studies and established that relative accuracy is worse over moraines and hill slopes due to the narrow ridges and steep slopes. Despite of having several limitations, these DEMs provide reference information from 2000, and ASTER DEMs derived from optical stereo pair images may provide temporal information but have limitations due to the presence of clouds (Fujita et al. 2008). ALOS-PALSAR DEM and Cartosat DEM (Govindha Raj et al. 2013; Debnath et al. 2018; Das et al. 2024) are also gaining popularity in the GLOF analysis because of better spatial resolution. Similarly, the morphological parameters related to dimensional measurement, like area, length or proximity to the glacier, etc., also include mapping error due to mapping uncertainty along the boundary pixels (Chand and Sharma 2015). The availability of high-resolution images or DEMs helps in minimising such uncertainty and therefore various researchers have started using them (Mohanty et al. 2023; Das et al. 2024). However, freely available data for both optical and DEM are more popular among the researchers than the paid ones, in spite of having finer spatial resolution. Collectively, the dataset with longer time series and better spatial

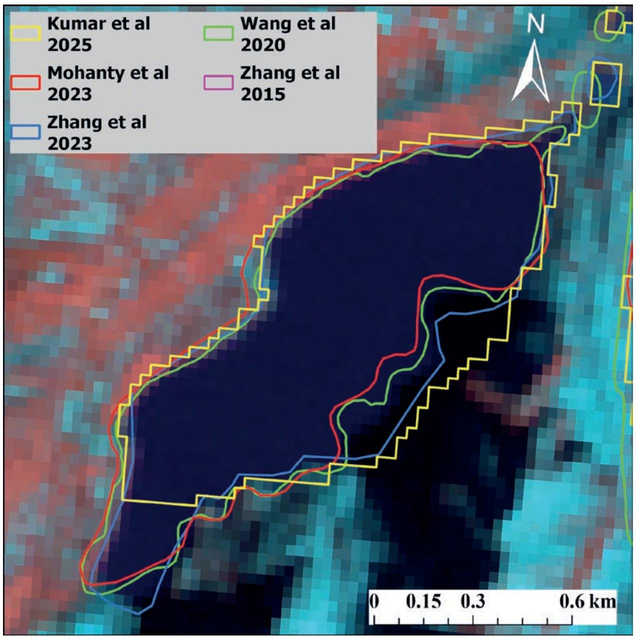

Fig. 6 The variation in glacial lake outline for the year 1990 corresponding to different researches are shown with False colour composite of Landsat 05, dated 20th Oct, 1990, where, Mohanty et al. (2023) have taken Zhang et al. (2015) as the reference layer (Not visible in the image).

Tab. 4 Selection of time and mapping criteria for defining the outline of the glacial lake boundary may lead to change in the variation of their number.

Publication	Time of mapping	Number of lakes	Mapping criteria (km^2)
Zhang et al. 2015	Nov, 1990	80	0.0027
Wang et al. 2020	Nov, 1990	229	0.0054
Zhang et al. 2023	Sep, 1992	192	0.0036
Mohanty et al. 2023	Nov, 1990	95	0.0100
Kumar et al. 2025	Ablation period, 1990	330	0.0036

resolution will be helpful in understanding the evolution of glacial lakes in the Sikkim region.

The NDWI-based identification is the most popular and established methodology to detect a water body (Gaikwad et al. 2025). However, shadow pixels of the image may also be categorised as water pixels, therefore limiting the capacity of the threshold-based water body delineation technique (Chen et al. 2021). Choosing a threshold for the identification of water pixels may also vary region to region, and therefore a general consensus has not been established among researchers. Such complications lead to the requirement of manual quality assessment (Zhang et al. 2023; Chen et al. 2021). The selection of different years' data (see section 4.4) and, at the same time, the mapping criteria like minimum lake area vary as 0.001, 0.0027, 0.0036, 0.0054 and 0.01km^2 for Aggarwal et al. (2017), Zhang et al. (2015), Zhang et al. (2023), Wang et al. (2020) and Mohanty et al. (2021), respectively, which influence the number of outlined lakes (Tab. 4) and their extent (Fig. 6), which may have been due to the aforementioned reasons.

5.5 GLOF analysis: adopted methodology and parameters

The recent occurrence of GLOF at South Lhonak Lake (Zhang et al. 2024; Sattar et al. 2025) has highlighted the necessity to investigate the prior research related to GLOF analysis methodology and associated parameters. Multiple-criteria decision-making (MCDM) methods like AHP and FAHP are the most common among such researchers for weight assignment to the various parameters affecting GLOF (Mohanty and Maiti 2021; Kaushik et al. 2024; Das et al. 2024; Gaikwad et al. 2025). Expert knowledge is a key factor in assigning the relative importance among the parameters for which weights are estimated using the aforementioned methods. It is very popular because it is most efficient and easy to replicate without much computational requirement (Gaikwad et al. 2025). Huggel et al. (2004) characterise the various parameters with qualitative probability, but it is still limited to the geography (Swiss Alps). However, further research advancement over recent years has improved the parameters selection for GLOF analysis (Emmer and Vilímek 2014; Emmer et al. 2022). The studies related to the Sikkim area include the parameters for GLOF analysis: lake and glacier area, area change, expansion rate, watershed area, downstream slope, shape index, slope, height of the lake outlet dam, rockfall or avalanche occurrences, proximity to glacier, and proximity to basin outlet. Satellite-based investigation helps in the derivation of these parameters without physically visiting the area but limits their accuracy due to the spatial resolution of the images (Section 5.4).

The parameter selection for GLOF analysis by researchers across the high mountain regions of the world is limited to remote sensing based derived parameters due to glacial lake inaccessibility. Emmer et al. 2022 have extensively worked on the parameters associated with GLOF analysis for the Andes and the Cordillera Blanca, Peru (Emmer et al. 2016). Emmer and Cochachin (2013) have studied in detail the failure mechanism of moraine dams in North America and the Himalayas, highlighting almost no difference in the failure mechanism except for the season of frequent occurrence of GLOF. All the researchers working in the Himalayas marginally vary in choosing the parameters for the GLOF analysis, like Worni et al. (2013), who do not consider the glacier's parameters and the watershed area, and similarly Gaikwad et al. 2025 also do not consider the watershed area. The interdependency of glacial lake vulnerability with the associated glacier is noticed by every researcher, and therefore proximity to the glacier, glacier and lake morphometric parameters, along with lake dam parameters, have been widely utilised for the assessment of GLOF analysis across the world (Emmer and Vilímek 2014).

6. Conclusion

Overall, the existing research on the glacial lakes in the Sikkim Himalaya demonstrates a wide range of inventory creation and significant enough research on GLOF analysis. These studies have been mostly dependent on the availability of satellite data due to the inaccessibility of the glacial lakes for physical mapping. The most reliable Landsat image with the longest time series has been widely utilised by the researchers. NDWI-based glacial lake identification is commonly implemented for the identification of glacial lakes, but their precision is further improved using manual correction. The typology criteria significantly vary for various purposes of research, where "sources of water" for the lake is the most common. Morphological, topographical, and environmental parameters associated with glaciers, glacial lakes and the surrounding area affect the GLOF analysis, and AHP is commonly utilised for estimation of their weights. An inventory with consistent mapping standards over a longer temporal coverage (possibly from 1962) over the Sikkim region can be prepared, which will be a better basis for the assessment of glacial lake vulnerability under the influence of continuous climate change.

References

Abdul Hakeem, K., Abirami, S., Rao, V. V., Diwakar, P. G., Dadhwal, V. K. (2018): Updated Inventory of Glacial Lakes in Teesta Basin Using Remote Sensing Data for Use in GLOF Risk Assessment. Journal of the Indian Society of Remote Sensing 46(3), 463–470, https://doi.org/10.1007/s12524-017-0699-1.

Agarwal, V., van Wyk de Vries, M., Haritashya, U. K., Garg, S., Kargel, J. S., Chen, Y. J., Shugar, D. H. (2023): Long-term analysis of glaciers and glacier lakes in the Central and Eastern Himalaya. Science of the Total Environment 898: 165598, https://doi.org/10.1016/j.scitotenv.2023.165598.

Aggarwal, S., Rai, S. C., Thakur, P. K., Emmer, A. (2017): Inventory and recently increasing GLOF susceptibility of glacial lakes in Sikkim, Eastern Himalaya. Geomorphology 295, 39–54, https://doi.org/10.1016/j.geomorph.2017.06.014.

Banerjee, P., Bhuiyan, C. (2023): Glacial lakes of Sikkim Himalaya: their dynamics, trends, and likely fate – a timeseries analysis through cloud-based geocomputing, and machine learning. Geomatics, Natural Hazards and Risk 14(1), https://doi.org/10.1080/19475705.2023.2286903.

Basnett, S., Kulkarni, A. V., Bolch, T. (2013): The influence of debris cover and glacial lakes on the recession of glaciers in Sikkim Himalaya, India. Journal of Glaciology 59(218), 1035–1046, https://doi.org/10.3189/2013JoG12J184.

Campbell, J. G., Pradesh, H. (2005): Inventory of glaciers, glacial lakes and the identification of potential glacial lake outburst floods (GLOFs) affected by global warming in the mountains of India, Pakistan and China/Tibet Autonomous Region. Available online: https://www.apn-gcr.org/wp-content/uploads/2020/09/d71dea72bc764245642e7d047c095463.pdf (accessed on 5 April 2025).

Chanda, A., Biswas, B. (2021): Glacial Lake Outflow Hazard and Risk Probability in Sikkim. In Rai, P. K., Singh, P., Mishra, V. N. (Eds.), Recent Technologies for Disaster Management and Risk Reduction. Earth and Environmental Sciences Library. Cham: Springer, https://doi.org/10.1007/978-3-030-76116-5_1.

Chen, F., Zhang, M., Guo, H., Allen, S., Kargel, J. S., Haritashya, U. K., Watson, C. S. (2021): Annual 30 m dataset for glacial lakes in High Mountain Asia from 2008 to 2017. Earth System Science Data 13(2), 741–766, https://doi.org/10.5194/essd-13-741-2021.

Chowdhury, A., Kroczek, T., De, S. K., Vilímek, V., Sharma, M. C., Debnath, M. (2021): Glacial lake evolution (1962–2018) and outburst susceptibility of gurudongmar lake complex in the Tista basin, Sikkim Himalaya (India). Water 13(24): 3565, https://doi.org/10.3390/w13243565.

Chowdhury, A., Begam, S., Kroczek, T., Vilimek, V., Sharma, M. C., De, S. K. (2025): Erosional dam breach modelling for multi-glacial lake outburst process and downstream flood exposure assessment in the Eastern Himalayas. Journal of Hydrology 649: 132417, https://doi.org/10.1016/j.jhydrol.2024.132417.

Das, A., Singh, S. K., Kanga, S., Sajan, B., Meraj, G., Kumar, P. (2024): Glacial Lake Outburst Flood Susceptibility Mapping in Sikkim: A Comparison of AHP and Fuzzy AHP Models. Climate 12(11): 173, https://doi.org/10.3390/cli12110173.

Debnath, M., Syiemlieh, H. J., Sharma, M. C., Kumar, R., Chowdhury, A., Lal, U. (2018): Glacial lake dynamics and lake surface temperature assessment along the Kangchengayo-Pauhunri Massif, Sikkim Himalaya, 1988–2014. Remote Sensing Applications: Society and Environment 9, 26–41, https://doi.org/10.1016/j.rsase.2017.11.002.

Desinayak, N., Prasad, A. K., Vinod, A., Mishra, S., Shukla, A., Nayak, S. (2023): Rise in Mid-Tropospheric Temperature Trend (MSU/AMSU 1978–2022) over the Tibet and Eastern Himalayas. Applied Sciences 13(16): 9088, https://doi.org/10.3390/app13169088.

Emmer, A., Wood, J. L., Cook, S. J., Harrison, S., Wilson, R., Diaz-Moreno, A., Reynolds, J. M., Torres, J. C., Yarleque, C., Mergili, M., Jara, H. W. et al. (2022): 160 glacial lake outburst floods (GLOFs) across the Tropical Andes since the Little Ice Age. Global and Planetary Change 208: 103722, https://doi.org/10.1016/j.gloplacha.2021.103722.

Emmer, A., Cochachin, A. (2013): The causes and mechanisms of moraine-dammed lake failures in the Cordillera Blanca, North American Cordillera, and Himalayas. AUC Geographica 48(2), 5–15, https://doi.org/10.14712/23361980.2014.23.

Emmer, A., Vilímek, V. (2014): New method for assessing the susceptibility of glacial lakes to outburst floods in the Cordillera Blanca, Peru. Hydrology and Earth System Sciences 18(9), 3461–3479, https://doi.org/10.5194/hess-18-3461-2014.

Emmer, A., Klimeš, J., Mergili, M., Vilímek, V., Cochachin, A. (2016): 882 lakes of the Cordillera Blanca: An inventory, classification, evolution and assessment of susceptibility to outburst floods. Catena 147, 269–279, https://doi.org/10.1016/j.catena.2016.07.032.

Fujita, K., Suzuki, R., Nuimura, T., Sakai, A. (2008): Performance of ASTER and SRTM DEMs, and their potential for assessing glacial lakes in the Lunana region, Bhutan Himalaya. Journal of Glaciology, 54(185), 220–228, https://doi.org/10.3189/002214308784886162.

Gaikwad, D., Tiwari, R. K., Kumar, M., Guha, S. (2025): Hazard, and risk modelling of glacial lakes in the Sikkim Himalaya: Integrating qualitative and quantitative approaches for hazard assessment. Geomorphology 471: 109577, https://doi.org/10.1016/j.geomorph.2024.109577.

Gardelle, J., Arnaud, Y., Berthier, E., (2010): Contrasted evolution of glacial lakes along the Hindu Kush Himalaya mountain range between 1990 and 2009. Global and Planetary Change 75(1–2), 47–55, https://doi.org/10.1016/j.gloplacha.2010.10.003.

Garg, P. K., Shukla, A., Jasrotia, A. S. (2019): On the strongly imbalanced state of glaciers in the Sikkim, eastern Himalaya, India. Science of the Total Environment 691, 16–35, https://doi.org/10.1016/j.scitotenv.2019.07.086.

Govindha Raj, B. K., Kumar, V. K., Remya, S. N. (2013): Remote sensing-based inventory of glacial lakes in Sikkim Himalaya: Semi-automated approach using satellite data. Geomatics, Natural Hazards and Risk 4(3), 241–253, https://doi.org/10.1080/19475705.2012.707153.

Hazra, P., Krishna, A. P. (2022): Assessment of Proglacial Lakes in Sikkim Himalaya, India for Glacial Lake Outburst Flood (GLOF) Risk Analysis using HEC-RAS and Geospatial Techniques. Journal of the Geological Society of India 98(3), 344–352, https://doi.org/10.1007/s12594-022-1986-1.

Huggel, C., Haeberli, W., Kääb, A., Bieri, D., Richardson, S. (2004): An assessment procedure for glacial hazards in the Swiss Alps. Canadian Geotechnical Journal 41(6), 1068–1083, https://doi.org/10.1139/t04-053.

Huss, M., Bookhagen, B., Huggel, C., Jacobsen, D., Bradley, R. S., Clague, J. J., Vuille, M., Buytaert, W., Cayan, D. R., Greenwood, G., Mark, B. G., Milner, A. M., Weingartner, R., Winder, M. (2017): Toward mountains without permanent snow and ice. Earth's Future 5(5), 418–435, https://doi.org/10.1139/t04-053.

Islam, N., Patel, P. P. (2022): Inventory and GLOF hazard assessment of glacial lakes in the Sikkim Himalayas, India. Geocarto International 37(13), 3840–3876, https://doi.org/10.1080/10106049.2020.1869332.

Iturrizaga, L. (2014): Glacial and glacially conditioned lake types in the Cordillera Blanca, Peru: a spatiotemporal conceptual approach. Progress in Physical Geography 38(5), 602–636, https://doi.org/10.1177/0309133314546344.

Ives, J. D., Shrestha, R. B., Mool, P. K. (2010): Formation of Glacial Lakes in the Hindu Kush-Himalayas and GLOF Risk Assessment. International Centre for Integrated Mountain Development (ICIMOD), https://doi.org/10.53055/ICIMOD.521.

Kaushik, S., Rafiq, M., Dharpure, J. K., Howat, I., Moortgat, J., Joshi, P. K., Singh, T., Dietz, A. J. (2024): Increasing risk of glacial lake outburst flood in Sikkim, Eastern Himalaya under climate warming. Remote Sensing Applications: Society and Environment 36: 101286, https://doi.org/10.1016/j.rsase.2024.101286.

Kovanda, M. (2024): Interpretation of the glacial lake outburst floods database in relation to climatic conditions in different world regions. AUC Geographica 59(2), 282–293, https://doi.org/10.14712/23361980.2024.12.

Kumar, A., Mal, S., Schickhoff, U., Dimri, A. P. (2025): Basin-scale spatio-temporal development of glacial lakes in the Hindukush-Karakoram-Himalayas. Global and Planetary Change 245: 104656, https://doi.org/10.1016/j.gloplacha.2024.104656.

Kumar, P., Sharma, M. C. (2023): Frontal changes in medium-sized glaciers in Sikkim, India during 1988–2018: Insights for glacier-climate synthesis over the Himalaya. IScience 26(10): 107789, https://doi.org/10.1016/j.isci.2023.107789.

Kumar, P., Sharma, M. C., Saini, R., Singh, G. K. (2020): Climatic variability at Gangtok and Tadong weather observatories in Sikkim, India, during 1961–2017. Scientific reports 10(1): 15177, https://doi.org/10.1038/s41598-020-71163-y.

Maharjan, S. B., Mool, P. K., Lizong, W., Xiao, G., Shrestha, F., Shrestha, R. B., Khanal, N. R., Bajracharya, S. R., Joshi, S., Shai, S., Baral, P. (2018): The Status of Glacial Lakes in the Hindu Kush Himalaya – ICIMOD Research Report 2018/1. International Centre for Integrated Mountain Development (ICIMOD), https://doi.org/10.53055/ICIMOD.742.

Mohanty, L., Maiti, S. (2021): Probability of glacial lake outburst flooding in the Himalaya. Resources, Environment and Sustainability 5: 100031, https://doi.org/10.1016/j.resenv.2021.100031.

Mohanty, L., Maiti, S., Dixit, A. (2023): Spatio-temporal assessment of regional scale evolution and distribution of glacial lakes in Himalaya. Frontiers in Earth Science 10, https://doi.org/10.3389/feart.2022.1038777.

NWIA/SAC/ISRO (2010): High altitude lakes of India. Available online: http://www.sac.gov.in (accessed on 10 February 2025).

NWIA/SAC/ISRO (2012): High altitude lakes of India. Available online: http://www.sac.gov.in (accessed on 10 February 2025).

Racoviteanu, A. E., Arnaud, Y., Williams, M. W., Manley, W. F. (2015): Spatial patterns in glacier characteristics and area changes from 1962 to 2006 in the Kanchenjunga-Sikkim area, eastern Himalaya. Cryosphere 9(2), 505–523, https://doi.org/10.5194/tc-9-505-2015.

Sattar, A., Cook, K. L., Rai, S. K., Berthier, E., Allen, S., Rinzin, S. et al. (2025): The Sikkim flood of October 2023: Drivers, causes, and impacts of a multihazard cascade. Science 387: eads2659, https://doi.org/10.1126/science.ads2659.

Schmidt, S., Nüsser, M., Baghel, R., Dame, J. (2020): Cryosphere hazards in Ladakh: the 2014 Gya glacial lake outburst flood and its implications for risk assessment. Natural Hazards 104, 2071–2095, https://doi.org/10.1007/s11069-020-04262-8.

Sharma, D., Singh, I. E., Paul, K., Mukherjee, S. (2019): Assessment of hydromorphological conditions of upper and lower dams of river Teesta in Sikkim. Journal of Spatial Hydrology 15(2). Available online: https://scholarsarchive.byu.edu/josh/vol15/iss2/1 (accessed on 5 January 2025).

Shugar, D. H., Burr, A., Haritashya, U. K., Kargel, J. S., Watson, C. S., Kennedy, M. C., Bevington, A. R., Betts, R. A., Harrison, S., Strattman, K. (2020): Rapid worldwide growth of glacial lakes since 1990. Nature Climate Change 10(10), 939–945, https://doi.org/10.1038/s41558-020-0855-4.

Shukla, A., Garg, P. K., Srivastava, S. (2018): Evolution of glacial and high-altitude lakes in the Sikkim, Eastern Himalaya over the past four decades (1975–2017). Frontiers in Environmental Science 6: 81, https://doi.org/10.3389/fenvs.2018.00081.

Song, C., Fan, C., Ma, J., Zhan, P., Deng, X. (2025): A spatially constrained remote sensing-based inventory of glacial lakes worldwide. Scientific Data 12(1): 464, https://doi.org/10.1038/s41597-025-04809-z.

Verma, P., Ramsankaran, R. (2022): Semi-automated mapping of glacial lakes – a study in Sikkim Himalayas, India. Geocarto International 37(25), 8254–8272, https://doi.org/10.1080/10106049.2021.1996640.

Wang, X., Guo, X., Yang, C., Liu, Q., Wei, J., Zhang, Y., Liu, S., Zhang, Y., Jiang, Z., Tang, Z. (2020): Glacial lake inventory of high-mountain Asia in 1990 and 2018 derived from Landsat images. Earth System Science Data 12(3), 2169–2182, https://doi.org/10.5194/essd-12-2169-2020.

Worni, R., Huggel, C., Stoffel, M. (2013): Glacial lakes in the Indian Himalayas – From an area-wide glacial lake inventory to on-site and modeling based risk assessment of critical glacial lakes. Science of the Total Environment 468–469, S71–S84, https://doi.org/10.1016/j.scitotenv.2012.11.043.

Zhang, G., Bolch, T., Yao, T., Rounce, D. R., Chen, W., Veh, G., King, O., Allen, S. K., Wang, M., Wang, W. (2023): Underestimated mass loss from lake-terminating glaciers in the greater Himalaya. Nature Geoscience 16(4), 333–338, https://doi.org/10.1038/s41561-023-01150-1.

Zhang, G., Yao, T., Xie, H., Wang, W., Yang, W. (2015): An inventory of glacial lakes in the Third Pole region and their changes in response to global warming. Global and Planetary Change 131, 148–157, https://doi.org/10.1016/j.gloplacha.2015.05.013.

Land cover change and its effects on catchment hydrology: A quantitative analysis using SWAT in Horní Úpa

Nikol Zelíková[1,2], Markéta Potůčková[4], Marek Purm[1,3], Václav Šípek[1,*], Kristýna Falátková[1], Michael Hofbauer[1,2], Alex Šrollerů[4], Lucie Červená[4], Zuzana Lhotáková[5], Jana Albrechtová[5], Lucie Kupková[4]

[1] Institute of Hydrodynamics of the Czech Academy of Sciences, Czechia
[2] Czech University of Life Sciences Prague, Faculty of Environmental Sciences, Department of Water Resources and Environmental Modelling, Czechia
[3] Charles University, Faculty of Science, Department of Physical Geography and Geoecology, Czechia
[4] Charles University, Faculty of Science, Department of Applied Geoinformatics and Cartography, Czechia
[5] Charles University, Faculty of Science, Department of Experimental Plant Biology, Czechia
[*] Corresponding author: sipek@ih.cas.cz

ABSTRACT

The changes in land cover, particularly in vegetation, directly influence regional water systems through various processes and have potential to alter not only microclimates and local hydrological regimes, but also local ecosystems and downstream water resources. This article investigates the interplay between land cover change and hydrological processes over a 30-year period in the mid-latitude mountainous catchment. Key vegetation trends were identified by supervised classification of geometrically and radiometrically corrected Landsat satellite imagery. Specifically, the replacement of coniferous forests with transitional woodland-shrub and a gradual increase in mixed forests, influenced by disturbance events, such as windthrows and bark beetle outbreaks, were observed. The SWAT model was successfully calibrated and validated using long-term discharge data, allowing for the simulation of land cover scenarios. Results suggest that land cover changes exerted a limited influence on total water balance, indicating a degree of hydrological resilience of the catchment. However, they affected the partitioning of runoff components, such as direct flow, subsurface lateral flow, and groundwater recharge. The study demonstrates the value of integrating the satellite-based land cover analysis with process-based modelling to understand long-term land-hydrology interactions in complex terrain. The findings underscore the importance of improving spatial resolution, dynamic vegetation modelling, and soil-vegetation parameterization for future assessments under changing environmental conditions. This research contributes to a growing body of knowledge essential for sustainable water resource management in sensitive mountain regions.

KEYWORDS

land cover change; hydrological modelling; SWAT model; Landsat; Czechia

Received: 25 June 2025
Accepted: 22 September 2025
Published online: 24 October 2025

Zelíková, N., Potůčková, M., Purm, M., Šípek, V., Falátková, K., Hofbauer, M., Šrollerů, A., Červená, L., Lhotáková, Z., Albrechtová, J., Kupková, L. (2025): Land cover change and its effects on catchment hydrology: A quantitative analysis using SWAT in Horní Úpa. AUC Geographica 60(2), 186–203
https://doi.org/10.14712/23361980.2025.18

1. Introduction

In the last six decades, land cover change concerned about one-third of the global land area (Winkler et al. 2021). The changes in land cover, particularly in vegetation, directly influence the regional water systems through various processes, such as evapotranspiration, interception or infiltration (Wang et al. 2018; Chen et al. 2021). This is caused by the role of vegetation as an interface between the soil and the atmosphere (Dubbert and Werner 2018). These alterations in the hydrological cycle have also profound effects on the surface energy balance, carbon cycle and land surface-atmosphere interactions, influencing air temperature and precipitation patterns, humidity levels, and surface's albedo (Meier et al. 2021). They have potential to alter not only microclimates and local hydrological regimes, but also local ecosystems and downstream water resources.

A wide range of remote sensing and GIS-based techniques are used to analyze land cover change, enabling the assessment of spatial and temporal dynamics in vegetation cover (Feranec et al. 2014, 2016; Ramon et al. 2020; Kupková et al. 2021, 2023). One of the most common approaches is the classification of multi-temporal satellite imagery, such as Landsat or Sentinel-2, which allows for the detection of land cover transitions over decades (Kupková et al. 2018, Chaves et al. 2020). Change detection methods, including supervised classification, vegetation indices (e.g., NDVI, EVI, DI, forest z-score), and machine learning algorithms, are commonly applied to quantify deforestation, afforestation, or shifts in vegetation structure (Haeley et al. 2005; Huang et al. 2010; Talukdar et al. 2020; Gao et al. 2020).

As land cover affects the hydrological cycle in numerous ways, it is vital to analyze the land coverage by particular vegetation types. Forested areas use significant amount of water for transpiration (Jasechko et al. 2013; Sutanto et al. 2014) and a notable part of precipitation is not reaching the soil surface due to interception (Kofroňová et al. 2021). Infiltration from precipitation is in the forest also increased due to root growth and organic matter, enhancing the soil's ability to absorb and retain water (Alaoui et al. 2011; Archer et al. 2013) compared to urban or agricultural lands, affecting the flow regimes of rivers and streams (Yang and Zhang 2011). In contrast, meadows and urban areas have markedly different impacts on the hydrological cycle. Meadows, with their lower vegetation height and density, tend to have higher surface runoff (Blackburn et al. 2021) and less transpiration compared to forests (Pielke et al. 1998). Water table height in montane meadows determines their potential to be a large carbon sequestration sink in the context of changing hydroclimates and different land management decisions impacting meadow hydrology (Blackburn et al. 2021). Agricultural or urban areas, characterized by less permeable or impervious surfaces, dramatically reduce infiltration and increase surface runoff, leading to higher peak flows in urban streams and rivers (Sieker 2000; Blöschl 2022).

To predict the effects of land cover changes on the hydrological regime, hydrological models serve as valuable tools for managing water resources more effectively (Mensah et al. 2022). Numerous studies have utilized hydrological models to simulate the impacts of deforestation (Danáčová et al. 2020), afforestation (Sonnenborg et al. 2017), urbanization (Trinh and Chui 2013), and other land cover changes on various components of the hydrological cycle, such as stream flow (Cognard-Plancq et al. 2001), evapotranspiration (Launiainen et al. 2019), and groundwater recharge (Sonnenborg et al. 2017). However, the correct representation and parametrization of land cover in hydrological models is crucial for predicting hydrological responses to environmental changes. Inaccurate representation of land cover can lead to significant errors in model outputs (Yu et al. 2016).

In this study, we therefore combine a detailed analysis of landcover change over 30 years with a modelled catchment water balance. The selected study area is a forested mountain basin in the Krkonoše Mountains, Czechia. Our goal is to contribute to the understanding of the complex interactions between land use changes and the hydrological cycle. The specific aims of this study are: (1) quantification of land cover change in mid-latitude mountainous catchment over 30 years, (2) correct representation of the catchment water balance using a semi-distributed model, and (3) assessing the influence of time-varying land cover on the hydrological response, including two scenarios – afforestation and deforestation of the catchment.

2. Experimental area and methods

2.1 Experimental area

The Horní Úpa catchment is situated in the headwaters of the Krkonoše Mountains in the Czechia, Central Europe, covering an area of 82 km^2 (Fig. 1). The experimental area is a part of the KRNAP National Park. The altitude ranges from 600 to 1,600 meters above sea level. The region is part of the Lugicum unit of the Bohemian Massif and is primarily composed of metamorphic gneisses and schists with granite intrusions. Hydrogeologically, it is classified as a hard rock system (Lachassagne et al. 2011). The area is predominantly covered by shallow soils, namely crypto-podzols, podzols, rankers or cambisols. Forests and transitional forest zones dominate over 80% of the catchment, with coniferous species, particularly spruce, comprising 89% of the forest cover.

According to the Köppen climate classification, the catchment falls within the subarctic (Dfc) climate zone, with coldest month below 0 °C and 1–3 months averaging above 10 °C characterized by warm summers and relatively evenly distributed precipitation (Tolasz 2007). From 1990 to 2023, observations at the Pec pod Sněžkou meteorological station recorded an average annual air temperature of 5.4 °C and an average annual rainfall of 1,319 mm. The annual runoff height from the catchment is approximately 920 mm, representing ~70% of the total precipitation.

2.2 Land cover data

A 30-year time series of land cover in the Horní Úpa catchment was derived by supervised classification of geometrically and radiometrically corrected Landsat satellite imagery (L2 processing level, surface reflectance) acquired by the Thematic Mapper (TM), Enhanced Thematic Mapper + (ETM+) and Operational Land Imager (OLI) sensors. Due to the frequent occurrence of cloud cover in this area, cloud-free mosaics were first created by setting an optimized percentile of the valid pixel values of all available imagery within each growing season (day of year [DOY] 150–270). The quality band of the L2 product was used to select invalid pixels, i.e., pixels with cloud cover or obscured by cloud cover, pixels out of radiometer range, etc. (Zhu and Woodcock 2014). Years in which it was not possible to derive a cloud-free mosaic were omitted or replaced by a two- or three-year composite if the gap was greater than one year. For the given period 1991 to 2022, 23 datasets were created (Table 1).

The resulting cloud-free mosaic contained six spectral bands of a given sensor, covering the visible, near, and mid-infrared parts of the spectrum, and calculated spectral indices suitable for monitoring vegetation condition – normalized difference vegetation index, normalized difference infrared index, normalized burn ratio, greenness, brightness and wetness of the Tasselled Cap (TC) transform (Crist and Cicone 1984). For the TC calculation, images containing the reflectance at the top of the atmosphere and the coefficients from Crist and Cicone (1984) for Landsat 5 and Baig et al. (2014) for Landsat 8 were used. Elevations from the SRTM model were added as an additional layer. Processing was automated using the Google Earth Engine and the provided online access to the Landsat Collection 2 data (pre-processed by the US Geological Survey Earth Resources Observation and Science Center).

The landcover classes correspond to the catchment conditions and to the requirements of the SWAT hydrological model. Overall, we used 12 land cover classes (see Table 2). A supervised approach was chosen for classification (Zagajewski et al. 2021; Potůčková et al. 2021). The collection of training sets since 1997 was done by visual interpretation of available colour (RGB or CIR) orthophotos with a spatial resolution of 0.2 to 0.5 m in combination with Landsat

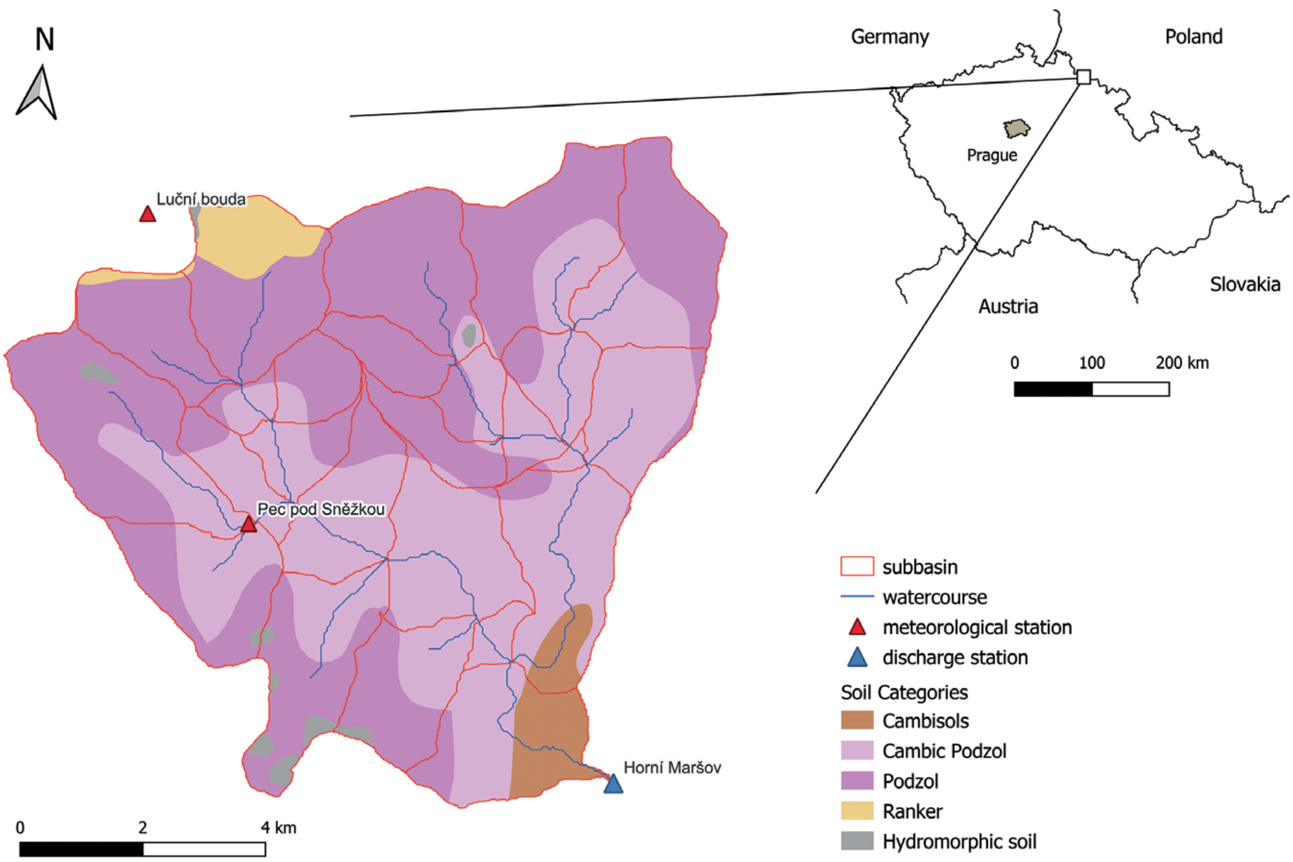

Fig. 1 Overview of the soil types distribution at the Horní Úpa catchment.

Tab. 1 Data sources for land cover analysis. The years where it was possible to produce annual cloud-free mosaics of satellite data are indicated by crosses on light blue shading. Dark blue shading marks the periods when a two- or three-year composite was used (from the DOY 150–270 imagery only). Percentile L5/L8 is equal to the threshold used for creating the cloud-free composite. The availability of aerial imagery used for training data collection is indicated by crosses in the AO column.

<table>
<tr><th rowspan="2">Year</th><th colspan="3">Landsat sensor</th><th rowspan="2">Percentile L5/L8</th><th rowspan="2">AO</th></tr>
<tr><th>L5/TM</th><th>L7/ETM+</th><th>L8/OLI</th></tr>
<tr><td>1991</td><td>x</td><td></td><td></td><td>40</td><td></td></tr>
<tr><td>1992</td><td rowspan="2">x</td><td></td><td></td><td rowspan="2">50</td><td></td></tr>
<tr><td>1993</td><td></td><td></td><td></td></tr>
<tr><td>1994</td><td>x</td><td></td><td></td><td>40</td><td></td></tr>
<tr><td>1995</td><td>x</td><td></td><td></td><td>40</td><td></td></tr>
<tr><td>1996</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>1997</td><td>x</td><td></td><td></td><td>40</td><td>x</td></tr>
<tr><td>1998</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>1999</td><td>x</td><td>x</td><td></td><td>60</td><td></td></tr>
<tr><td>2000</td><td>x</td><td>x</td><td></td><td>40</td><td>x</td></tr>
<tr><td>2001</td><td>x</td><td>x</td><td></td><td>40</td><td>x</td></tr>
<tr><td>2002</td><td>x</td><td>x</td><td></td><td>60</td><td></td></tr>
<tr><td>2003</td><td>x</td><td></td><td></td><td>40</td><td></td></tr>
<tr><td>2004</td><td>x</td><td></td><td></td><td>40</td><td>x</td></tr>
<tr><td>2005</td><td>x</td><td></td><td></td><td>40</td><td>x</td></tr>
<tr><td>2006</td><td>x</td><td>x</td><td></td><td>50</td><td></td></tr>
<tr><td>2007</td><td>x</td><td></td><td></td><td>40</td><td>x</td></tr>
<tr><td>2008</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2009</td><td>x</td><td></td><td></td><td>40</td><td></td></tr>
<tr><td>2010</td><td>x</td><td></td><td></td><td>40</td><td>x</td></tr>
<tr><td>2011</td><td>x</td><td></td><td></td><td>40</td><td>x</td></tr>
<tr><td>2012</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2013</td><td></td><td></td><td rowspan="3">x</td><td rowspan="3">50</td><td></td></tr>
<tr><td>2014</td><td></td><td></td><td>x</td></tr>
<tr><td>2015</td><td></td><td></td><td></td></tr>
<tr><td>2016</td><td></td><td></td><td rowspan="2">x</td><td rowspan="2">50</td><td>x</td></tr>
<tr><td>2017</td><td></td><td></td><td></td></tr>
<tr><td>2018</td><td></td><td></td><td>x</td><td>50</td><td>x</td></tr>
<tr><td>2019</td><td></td><td></td><td>x</td><td>50</td><td></td></tr>
<tr><td>2020</td><td></td><td></td><td>x</td><td>50</td><td>x</td></tr>
<tr><td>2021</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2022</td><td></td><td></td><td>x</td><td>50</td><td>x</td></tr>
</table>

data. The orthophotos provided by the Krkonoše National Park Administration and the Czech Office for Surveying, Mapping and Cadastre were available online as Web Map Services (KRNAP 2025; ČÚZK 2025). In addition, a normalised digital surface model (nDSM) derived from airborne laser scanning was provided by the Krkonoše National Park Administration for the years 2018 and 2022 (KRNAP 2025) and was used to discriminate the transitional woodlands and coniferous forest in training data. In 2022, a field survey was conducted to refine some categories (especially transitional woodlands). For the 1991, 1992, and 1994 imagery, training sets were collected by interpretation of Landsat imagery only. The buildings and roads class were generated from OpenStreetMap (as downloaded on January 28, 2023) and used for all time horizons. The area is located in the highly protected zone of the national park and new constructions are therefore minimal. Similarly, the water class, which consists of only a few pixels in the 30 m resolution Landsat data, was vectorized over the orthophoto and included in the final classification.

Tab. 2 Classified land cover classes for the SWAT hydrological model.

Land cover class	SWAT code
Artificial surfaces (urban, road network)	URBN
Meadows and pastures	PAST
Broad-leaved forest	FRSD
Coniferous forest	FRSE
Dead coniferous forest	BSVG
Mixed forest	FRST
Transitional woodlands and shrubs	SHRB
Forest clearings and sparsely vegetated areas	BSVG
Dwarf pine shrubs	SBRB
Peat bogs	MIGM
Rocks	BARR
Water bodies and water courses	WATR

The Random Trees algorithm, as implemented in ArcGIS Desktop v10 (Breiman 2001), was used for the classification. It is a non-parametric, robust classifier that can be applied to data of different scales. It provides information about the importance of each predictor. Moreover, the amount of training data required is smaller than for convolutional neural networks. Reflectivity values in six spectral bands, spectral indices, and height were used as features for classification, as mentioned above. Based on the tests performed, the classifier parameters were set to 300 trees and a maximum tree depth of 60.

To suppress random noise caused by the input data (radiometric values, training data), the following post-classification processing was carried out. First, we applied a majority filter to each single classification to remove pixels with a different class within or at the edge of homogeneous regions. Then, we defined a minimum classification mapping unit (1 × 2 pixels), and finally smoothed the time series (i.e., removed unrealistic changes in land cover within the time series).

After post-processing, the resulting 12 land cover layers were the basis for further analyses. Classification accuracy was assessed based on stratified random sampling of 1500 points at the 95% two-sided confidence level (Foody 2009). Classification classes were assigned to validation points based on orthophoto interpretation; 94 of these points were validated in the field in 2022. Due to the time cost associated with the number of validation points and lack of aerial orthoimages in the 1990s, validation was only performed in 5 time-horizons (2001, 2010, 2016, 2018, 2022). A standard error of the area determined for each class could be estimated, based on Olofsson et al. (2014).

2.3 The SWAT hydrological model

2.3.1 The SWAT model fundamentals

The SWAT model used in the study is a comprehensive basin-scale, continuous-time model (Arnold et al. 2012). As a semi-distributed model, SWAT provides a high level of spatial discretization by dividing the original catchment into multiple subbasins. Subbasins are further subdivided into series of hydrological response units – i.e. hydrotope (HRUs), which represent unique combinations of slope, land use, and soil type.

The main components of the model are the modules of direct runoff (SURQ), evapotranspiration, soil water infiltration, subsurface lateral runoff (LATQ), groundwater runoff (shallow and deep aquifer) (GWQ), water reservoirs, nutrients, bacteria and others. Most of the above components are calculated separately for each HRU in a daily step. In the evapotranspiration section, the water retained by interception is estimated first for a given land cover category. The Priestley-Taylor method (Priestley and Taylor 1972) was chosen for the estimation of potential evapotranspiration. Actual evapotranspiration is then determined separately for soil and plants using the approach of Ritchie (1972).

2.3.2 Required input data

SWAT requires the following spatially distributed inputs for model building: a digital elevation model (DEM), soil type, and land use layers. The DEM from the Shuttle Radar Topography Mission (SRTM) at a spatial resolution of 30 × 30 m was used. The soil type map was obtained from the Soil Subtypes Map of the Czech part of the Elbe Basin, created by Němečková (2008).

The SWAT model requires daily climate data, including minimum and maximum air temperatures (TMP), precipitation (PCP), average wind speed (WIN), solar radiation (SOL) and relative humidity (HUM). Meteorological observations covering the period from 1990 to 2019 were obtained from five meteorological stations (Fig. 1) from the Czech Hydrometeorological Institute. SWAT assigns meteorological data from the closest station to the central point of the subbasin. Each subbasin was then divided into several elevation bands spanning for 100 m. For each elevation band, the correction for the average observed elevation gradient was applied both for air temperatures and precipitations.

2.3.3 Modelling procedure

The SWAT model was calibrated at the beginning of the period of available data using the corresponding land cover. Calibration and validation were conducted using the SWAT-CUP 2012 software (Abbaspour 2013) in combination with the Sequential Uncertainty Fitting 2 (SUFI-2) optimization algorithm. Performance evaluation of the best-fit simulations included objective functions such as percent bias (PBIAS) and Nash-Sutcliffe efficiency (NSE). For further information on these objective functions, refer to Moriasi et al. (2015).

Further several distinct land cover maps were designed for the assessment of land cover influence on catchment rainfall-runoff relationship and water balance. The specific land cover maps corresponded to documented differences in aerial representation of particular land-covers. In order to eliminate the influence of different climatic conditions on the modelling results the model was run separately for every land-cover set-up using the entire period of meteorological data 1991–2021, so that the runoff characteristics can be assessed using exactly the same meteorological data and not only the ones present in the period of land-cover. Altogether four land cover set-ups were assessed (thoroughly described in section 3.2) representing four land cover set-ups and both catchment water balance as well as the rainfall-runoff. The four land cover set-ups are denoted as LU1991, LU2005, LU2011 and LU2022 always representing the land cover from particular year.

Finally, two extreme land cover scenarios were submitted to the hydrological model in order to check the sensitivity of the model results to pronounced changes in the land cover. In the first scenario, all areas originally attributed to transitional woodlands and shrubs and dwarf pine shrubs categories were altered to coniferous forest representing the gradual afforestation of the area induced by climate change. And in the second scenario, all coniferous forests were turned into transitional woodlands and shrubs, which may occur due to a bark beetle outbreak in the area.

3. Results

3.1 Calibration of the SWAT model

The hydrological model SWAT was calibrated in the period of 1993–1997. This period contains both wet (1995, 1997), dry (1996) and average years (1993, 1994) as recommended by Moriasi et al. (2015). The first two years of the available data were used as model warm-up. A total of 15 parameters were selected for the one-at-time sensitivity analysis. The selection of the parameters was based on a literature review on the application of SWAT models in streamflow and soil water calibration (Abbaspour 2015).

Within the calibration period, Nash-Suctliffe coefficient value of 0.47 was achieved for Horní Úpa watershed. In the validation period, the coefficient value was 0.50. The values of the Nash-Sutcliffe coefficient are satisfactory according to Moriasi et al. (2015), although there are at the lower acceptable margin. Largest discrepancies between modelled and observed discharges were found for winter and spring periods due to the uncertain estimation of snow accumulation and snow melt in particular dormant seasons. The Nash-Sutcliffe coefficient ranged from 0.22 in the worst-performing year (2012) to 0.76 in the best-performing simulation year (2007). The catchment water balance is correctly simulated as simulated runoff equaled to 98.9% of the observed one in the entire period of 31 simulated years. Only 20% (calibration) and 18% (validation) of the measured data were not enclosed by the 95PPU band.

3.2 Land cover characteristics

3.2.1 Classification accuracy

The assessment of land cover change was preceded by an analysis of the classification accuracy achieved in five time-horizons according to the methodology procedure outlined at the end of Section 2.2. The hydrological analysis focused on seven vegetation categories (i.e., grassland, coniferous, mixed, and broad-leaved forest, transitional woodlands, dwarf pine shrubs, clearings and areas with sparse vegetation) and rocks. As the classes of built-up and water areas were not of interest and were masked for the classifications, they were not included in the accuracy assessment. Table 3 shows that the overall classification accuracy was between 70 and 80%. Mixed and broad-leaved forest proved to be the most problematic categories, with both confusion between these two categories and confusion between mixed and coniferous forest and transitional woodlands. For each validation dataset, the area standard error was estimated for each class (Olofsson et al. 2014). The median of standard error calculated from the five validation observations was used as the best estimate of the achievable accuracy of area determination for a given category. Based on this, the significance of the difference in area of a given category between two consecutive time horizons was assessed using a 95% confidence interval. Significant changes are highlighted in Table 4. In general, the number of significant changes was low, with the most pronounced ones observed between 1992 and 1994 and between 2002 and 2003.

3.2.2 Land cover change

The result of the land cover classification from Landsat satellite data in the initial reference year 1991 and final year 2022 is shown in Fig. 2. Based on the classification, the area of each land cover class was calculated for each time horizon, see Table 4. According to the average values for the whole study period, coniferous forests (39%), transitional woodlands (29%) and meadows (10%) represent the highest proportion of the total classified area. The categories with a representation of less than 1% are peat bogs, water, and dead coniferous forest.

Changes in land cover categories within the time series can be used to express the dynamics of the development dynamics of the studied area of interest. The spatial distribution of pixels with unchanged

Tab. 3 Classification accuracy expressed in terms of F1-score and overall accuracy is based on 1,500 validation points obtained by visual interpretation of RGB and CIR aerial orthophotos with a spatial resolution of 0.2 m to 0.5 m.

	Landsat 8			Landsat 5	
	2022	2018	2016	2010	2001
	F1-score				
Meadows, pastures	0.91	0.83	0.89	0.86	0.75
Broad-leaved forest	0.53	0.56	0.73	0.60	0.56
Coniferous forest	0.89	0.88	0.89	0.80	0.92
Dead coniferous forest	0.75	0.56	0.71	–	–
Mixed forest	0.57	0.53	0.61	0.42	0.43
Transitional woodlands, shrubs	0.77	0.75	0.76	0.64	0.65
Clearings, sparsely vegetated areas	0.78	0.55	0.71	0.39	0.42
Dwarf pine shrubs	0.91	0.80	0.82	0.82	0.72
Peat bogs	0.98	0.93	0.85	0.68	0.83
Rocks	0.84	0.82	0.79	0.91	0.77
Overall Accuracy	**0.82**	**0.78**	**0.81**	**0.70**	**0.70**

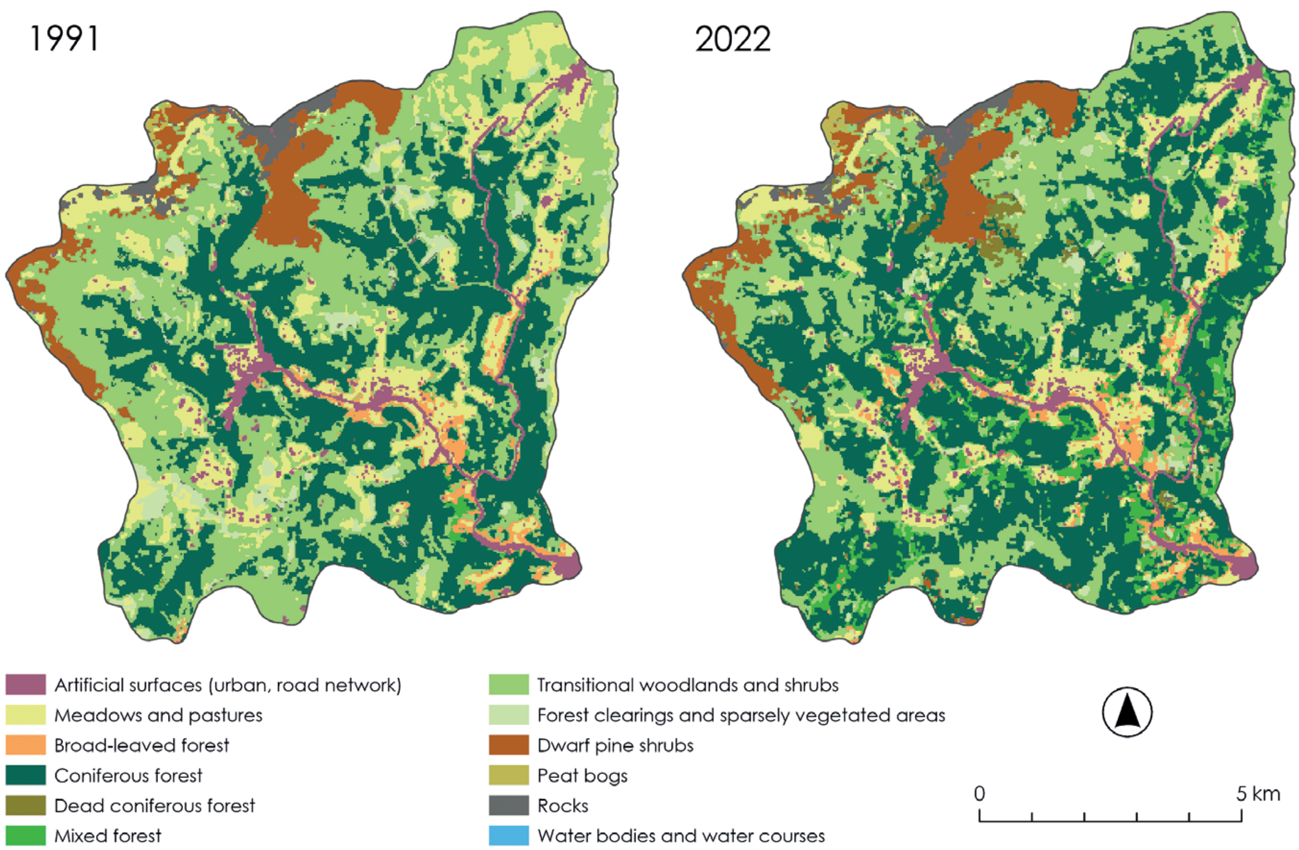

Fig. 2 Land cover classification from satellite images acquired by TM mapper and OLI sensors from Landsat 5 and Landsat 8 satellites, respectively, at the beginning and the end of the observed period (1991–2022).

categories throughout the time series, i.e., stable areas, and conversely, the number of changes in each pixel, is shown in Fig. 3. Stable areas represent 42.7% of the area of interest. The proportion of stable forest categories in the total area of the study catchment is as follows: deciduous forest 0.6%, mixed forest 0.2%, coniferous forest 20.4%, transitional woodlands 6.5%. Moreover, a detailed study of changes revealed that on 6.5% of the area, the change between matured deciduous, coniferous, and mixed forest categories occurred twice or more times. Similarly, pixels with repeated occurrences of dwarf pine shrubs, which do not belong to categories with high dynamics, represent 6.8% of the area. Such 'changes' indicate higher uncertainty in class distinction between the categories and may be attributed to classification inaccuracy rather than real change.

Breakpoints in the time series of the calculated areas of each category were sought. Fig. 4 depicts an example of the two most represented categories, coniferous trees and transitional woodlands. For coniferous trees and transitional woodlands, the most significant breaks in the time series occurred in 1995, 2001, and around 2005. The period between 1991 and 2005 is characterised by forest maturation, when the area of coniferous forest increased from its minimum (30 km^2) to its maximum (39 km^2), while the area of transitional forest decreased from its maximum (34 km^2) to its minimum (23 km^2). These trends changed after 2005. Since 2011 there has been no significant change. Both categories tend to decrease slightly. Based on this analysis four land covers maps were designed representing the land cover form the year 1991, 2005, 2011 and 2022 (denoted as LU1991, LU2005, LU2011 and LU2022).

3.3 Influence of land cover change on water budget and runoff

The following part is split into the assessment of evaluation of the gradual change of land cover on the hydrological regime of the Horní Úpa catchment and the water balance in hydrotopes covered with specific land cover.

3.3.1 Influence of land cover change on catchment runoff

The prominent source of information about land cover change influence on the catchment runoff characteristics are measured values of runoff in combination with precipitation sums. Hence first, the relationship between measured discharge and precipitation sums was analyzed by means of the double mass curve (Fig 5a). The curve demonstrates the stationary relationship between two observed variables from 1990 to 2014. After this breakpoint, less

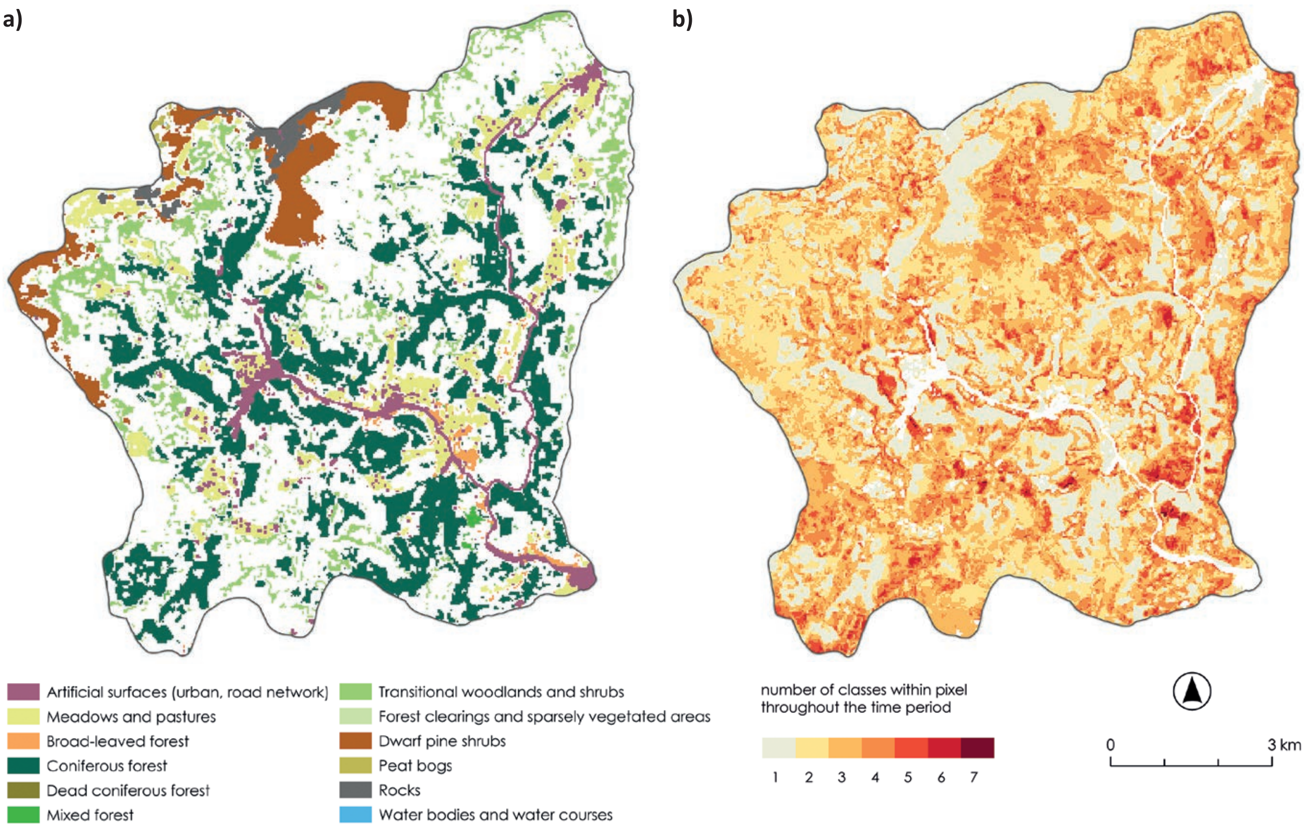

Fig. 3 Map of stable areas in Horní Úpa based on the classification of the satellite image time series between 1991 and 2022 (a). Pixels whose class has not changed over the entire time series are marked as stable. The number of land cover changes (variety) in a given pixel reflecting the dynamics of the area (b).

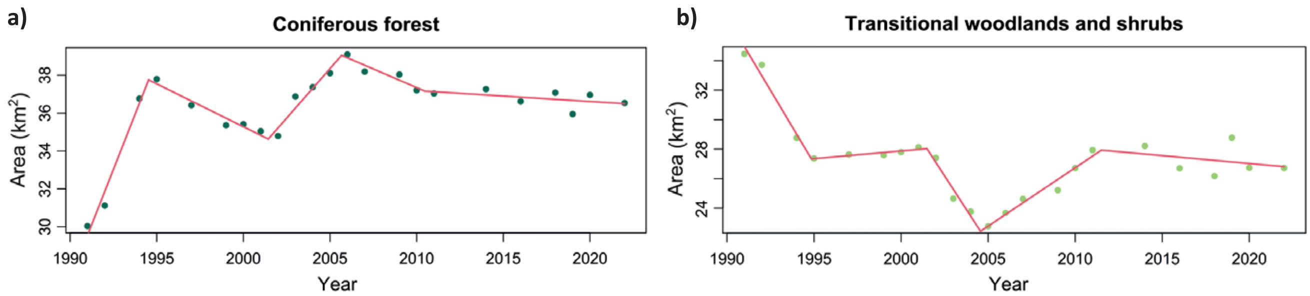

Fig. 4 Result of piecewise regression (red line) on the time series of the area calculated from Landsat image classification for the categories coniferous forest (a) and transitional woodlands (b). The main breakpoints are represented by the years 1995, 2001, and 2005 (or 2006 for the coniferous forest).

runoff was observed in comparison with precipitation sums. However, this was most probably caused by the occurrence of severe precipitation deficit during the drought period of 2014–2019 than by the changes in land cover. As mentioned in section 3.2, the most pronounced changes in land cover occurred from 1991 to 2005 and from 2005 to 2011. Since 2011 the spatial extent of particular land cover categories remained stable.

The documented change of land cover had only minor effect also on the modelled discharges as the long-term modelled average discharge differed only up to 3% among particular land cover set-ups. The only visible difference was that with increasing area covered with forest, more lateral subsurface and groundwater flow was produced at the expense of surface runoff. This was the case of set up representing the land cover from 2005 and 2011. However, the differences were up to 1.2% of the total runoff volume, hence statistically not significant (Fig. 5b). Nevertheless, the smaller extent of forest led to higher peak runoffs in a number of events (>25 $m^3.s^{-1}$).

3.3.2 Land cover influence on water balance

Most pronounced differences among investigated land covers were in terms of water partition in the

Tab. 4 Area of sub-categories based on land cover classification from Landsat data. The green and red colors highlight the maximum and minimum values for a given category. Light blue shading indicates years with a change in the area of a given category above the 95% confidence interval (based on the median of standard errors shown in Appendix A1). For abbreviations of the land cover classes Tab. 2.

	Area [km²]																								
	1991	1992	1994	1995	1997	1999	2000	2001	2002	2003	2004	2005	2006	2007	2009	2010	2011	2014	2016	2018	2019	2020	2022	Median	STD
URBN	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	0.00
PAST	12.71	12.90	10.77	10.75	11.09	11.95	11.26	10.86	10.93	8.84	8.58	8.35	8.43	9.06	9.24	8.10	7.84	8.42	8.36	7.71	7.52	7.86	8.25	8.84	1.69
FRSD	1.84	1.73	1.44	1.44	1.63	1.70	1.48	1.54	1.76	2.54	2.86	2.76	2.05	2.08	2.16	2.12	2.27	1.80	2.03	2.42	1.95	2.41	2.22	2.03	0.41
FRSE	30.04	31.12	36.77	37.79	36.41	35.36	35.41	35.04	34.79	36.87	37.36	38.11	39.09	38.19	38.03	37.20	37.03	37.26	36.62	37.08	35.95	36.96	36.53	36.87	2.10
BSVG																		0.24	0.63	0.50	0.77	0.80	1.10	0.70	0.29
FRST	0.37	0.46	2.19	2.31	2.34	2.87	3.39	3.81	4.26	7.53	8.06	8.66	6.56	6.02	5.29	4.76	4.43	5.97	6.82	5.41	4.47	5.23	5.96	4.76	2.23
SHRB	34.45	33.71	28.77	27.38	27.64	27.60	27.80	28.14	27.41	24.65	23.77	22.77	23.68	24.63	25.22	26.72	27.93	28.22	26.70	26.18	28.78	26.74	26.72	27.38	2.75
BSVG	3.38	2.89	2.03	2.39	3.03	2.51	2.28	2.41	2.44	2.88	2.69	2.50	2.19	2.21	2.37	3.58	2.95	0.35	1.10	2.73	2.64	1.99	1.68	2.44	0.69
SBRB	4.97	5.11	5.90	5.76	5.61	5.77	6.16	5.96	6.09	4.67	4.71	4.86	5.94	5.84	5.72	5.53	5.36	5.81	5.78	5.87	5.79	5.92	5.42	5.77	0.44
MIGM	0.13	0.14	0.15	0.18	0.22	0.23	0.22	0.22	0.24	0.17	0.18	0.18	0.23	0.23	0.24	0.20	0.22	0.24	0.25	0.27	0.26	0.24	0.26	0.22	0.04
BARR	1.58	1.41	1.45	1.47	1.50	1.49	1.46	1.48	1.56	1.31	1.26	1.28	1.30	1.20	1.20	1.24	1.43	1.16	1.18	1.30	1.33	1.31	1.32	1.32	0.13
WATR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

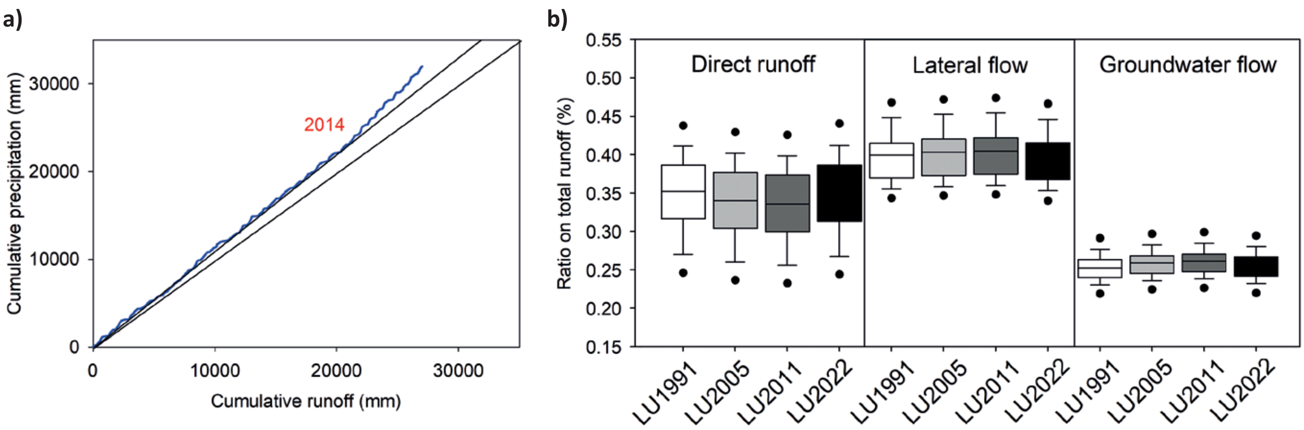

Fig. 5 Double mass curve and the ratio of direct (a), lateral subsurface and groundwater runoff in the period of 1991–2021 for all four land covers (b).

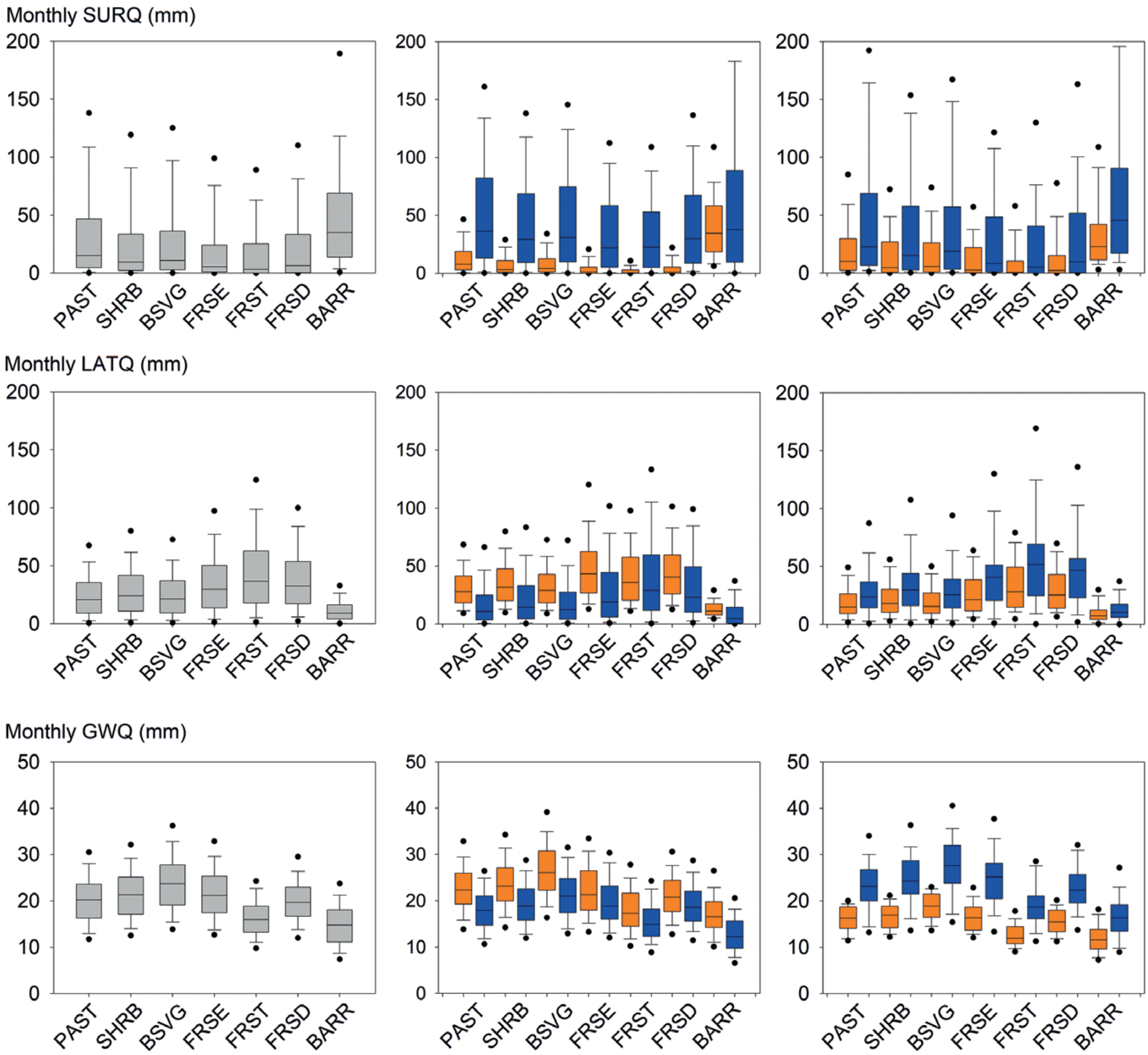

Fig. 6 Monthly sums of investigated water fluxes (SURQ, LATQ, GWQ in rows) in 1991–2021 for all hydrotopes divided according land cover categories (shortcuts explained in Table 2). The first column: annual sums, the second column: sums for summer (orange) and winter (blue), the third column: a 3-year period of wet (blue) (2000–2002) and dry (orange) (2014–2016) years.

soil profile into direct runoff, subsurface lateral flow and groundwater recharge with subsequent formation of groundwater contribution to streamflow. The differences in the rate of evapotranspiration were not statistically significant and merely not observable. This is given by the fact that under current climate conditions the mountainous regions are still energy limited in terms of evapotranspiration demand (Zelíková et al. 2025).

The fast direct runoff (SURQ) was highest (>600 mm.y^{-1}) in hydrotopes not covered by vegetation (BARR). Compared to all other land covers, it formed on average 2–3× more runoff. The lowest values of SRUQ were modelled from the forested areas (220–300 mm.y^{-1}). The average differences among forest types were at maximum 80 mm.year^{-1}. Among all three forest types, the highest SURQ was observed in deciduous forests and the lowest in mixed forests. All other land covers comprising pastures (PAST) and land with shrubs (SHRB) or sparse vegetation (BSVG) attained higher long-term averages of SURQ of 413, 326 and 349 mm.year^{-1}, respectively. The differences were more pronounced both in above-average wet and above-average dry years with dry years being most variable. SURQ was observed significantly more in winter months, only in the case of BARR there is no difference between summer and winter season (Fig. 6).

On the other hand, the lateral subsurface flow (LATQ) was formed at highest rate in forested hydrotopes forming average annual values of runoff from 450 to 550 mm.y^{-1}. The highest LATQ was on average formed in mixed forest (554 mm.year^{-1}), followed by deciduous forest (477 mm.year^{-1}) and finally the less

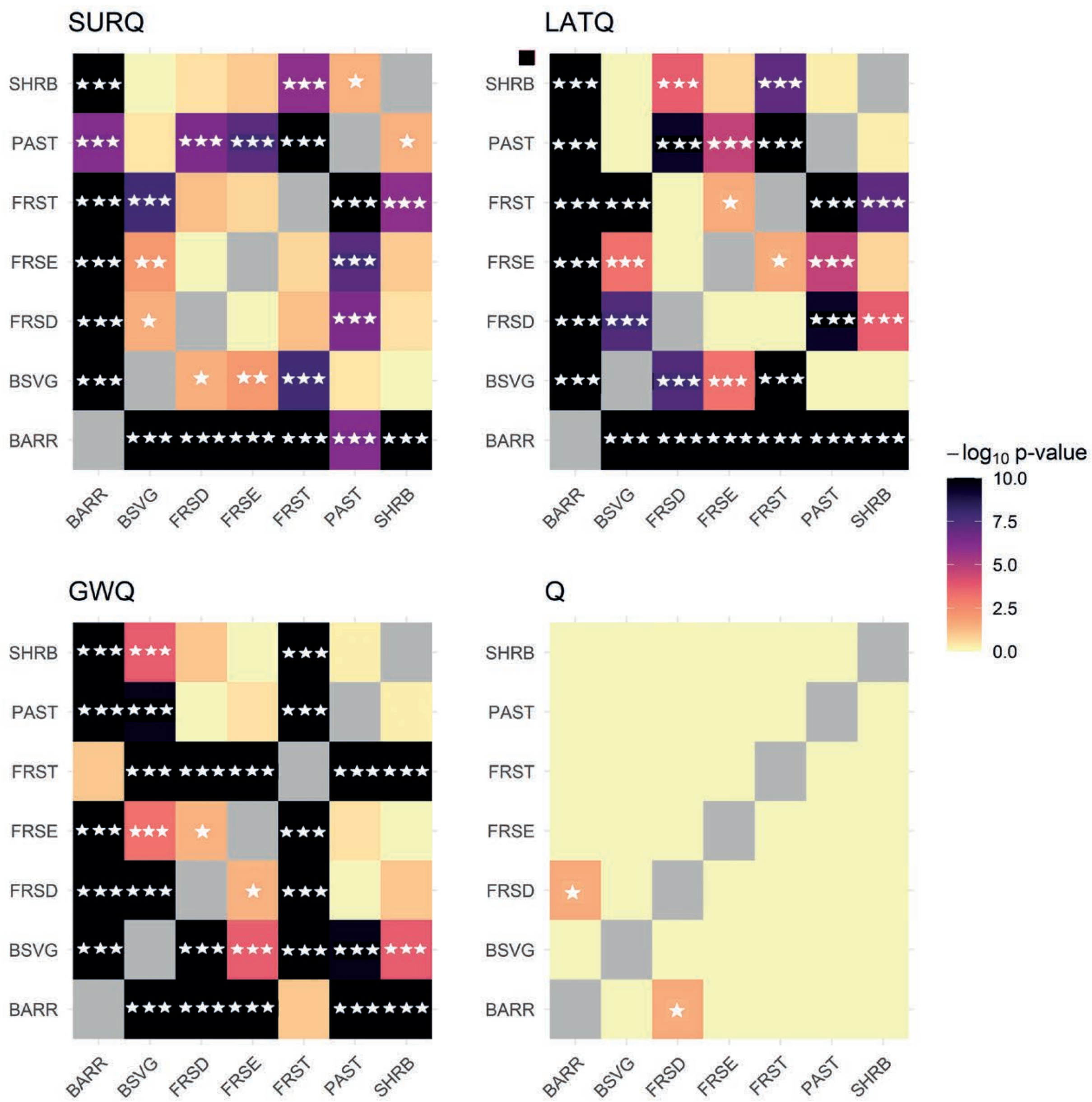

Fig. 7 Results of the Kruskal-Wallis test across all landcovers and inspected seasons. Asterisks indicate significantly different means (at the 5%, 2,5% and 1% significance level) and colours indicate the strength of a relationship.

prone was evergreen forest (438 mm.year^{-1}). Apart from forested hydrotopes, LATQ formation ranged between 300 and 400 mm.y^{-1} in pasture, shrubland and sparse vegetation covered hydrotopes. LATQ from pasture and shrubland covered hydrotopes statistically differ from the rest. Areas covered by sole bare land attained significantly less LATQ (<150 mm.y^{-1}) as the soil profile is the shallowest. There are no profound changes in the described pattern in wet and dry periods (Fig. 6), only the values of LATQ are higher or lower. In summer season, the differences among hydrotopes with particular landcover were not statistically significant. In winter, only the deciduous forest significantly differed from other forest types and less vegetated surfaces (Fig. 7).

The average annual values of groundwater contribution to runoff (GWQ) ranged from 179 mm.y^{-1} in the hydrotopes without vegetation cover to 289 mm.y^{-1} in the hydrotopes with sparse vegetation. Vegetated hydrotopes (PAST, SHRUB, FRSD, FRSE) attained average annual rates of groundwater flow from 241 to 260 mm.y^{-1}. Lower rate of groundwater flow was modelled for the mixed forest hydrotopes (197 mm.y^{-1}) as lateral subsurface flow was estimated to be the major flow pathway. All forested sites differed significantly from each other. Similarly, as for the LATQ, the same pattern of GWQ rates was observed among hydrotopes with particular land cover for summer periods and little less differences in winter period. The same applies for the dry and wet years, respectively.

The long-term average contributions of all three types of runoff for particular land covers are shown in Fig. 8. It demonstrates that in hydrotopes covered by bare soil, the direct runoff formed 63% of total runoff on average and this percentage was stable over the course of the year. Only 13% of runoff was modelled to be formed by lateral subsurface flow and less than 25% by the groundwater contribution. In the hydrotopes covered by sparse vegetation, the average contribution of all runoff types was merely equal (33% SURQ, 33% LATQ, 34% GWQ) with surface runoff dominating in the winter and equal contribution of direct and lateral flow in summer (~45%). The prevailing winter role of surface runoff was observed in all inspected hydrotopes documented by a distinct group of circles (December to April) in the right-hand side of the triangle plots (Fig. 8). This is given by the saturated soil in this period resulting in propensity of melt/rainwater to flow over soil surface or through the soil profile using shallow lateral pathways. In areas covered by shrubs and pastures, the lateral flow dominated summer runoff (~45%) with groundwater flow forming 30–40%. Lateral flow was on average highest in forest-covered hydrotopes (>45%), especially in mixed forests where it attained 56.5% of total runoff on average. The groundwater contribution in the forested sites ranged from 22.2% (FRST), over 25.8% (FRSD) to 30.5% (FRSE). Direct flow was least prominent in the forest-covered sites contributing up to 10% of runoff in summer. In summer months, mixed forests differed from evergreen and deciduous ones by higher contribution of lateral subsurface flow (groundwater flow at the expense of direct (up to 5%) and groundwater flow contributions.

3.3.3 Influence of two extreme land cover scenarios on catchment water balance

Deforestation represented by the change of all areas covered by conifer forest to transitional woodlands and shrubs led to slightly higher peak discharges and faster recession to baseflow. On average, twenty peak discharges were higher by 6.6%. The total runoff volume increased only by 2.7% compared to the reference scenario which was represented by the land cover from 1991. The surface runoff contribution to total runoff increased by 4.1% on average (1991–2021) at the expense of lateral subsurface runoff and groundwater runoff contributions, which decreased by 4.0% and 0.6%, respectively. Contrarily afforestation of the transitional areas led to slightly lower peak discharges and non-detectable changes in the recession limb of the hydrographs. The average decrease of the twenty peak discharges was 0.8%. Surface runoff decreased by 1.8% on average and it was compensated by the increase of groundwater contribution to runoff. The total runoff volume increased by 2.7% when the area was deforested and decreased by 0.8% when it was afforested. The overall influence of forest thinning and contrarily afforestation of all transitional woodlands and shrubs led only to minor changes in runoff volumes, timing and other runoff characteristics.

4. Discussion

4.1 Land cover classification

The Landsat data classification showed that at least 42.7% of the area of interest remained stable between 1991 and 2022. Although the overall accuracy for selected time horizons was only 70–80%, it is comparable to other studies that used Landsat data to quantify changes in different temperate forest types (Griffiths et al. 2014; Zagajewski et al. 2021) or forest disturbances (Senf et al. 2017) over decades. However, so far the Landsat archive is the only one that can provide consistent multispectral data continuously for such a long period.

Analysis of the time series of areas derived from the classification of each category revealed significant breakpoints in 1995, 2001, 2005, and 2011. Communication with experts from the KRNAP National Park Administration confirmed the observed changes in coniferous forest. Its increase from 1991, the beginning of the observation period, to 1995 can be attributed to the growth of forest plantations following extensive damage to Norway spruce stands in connection

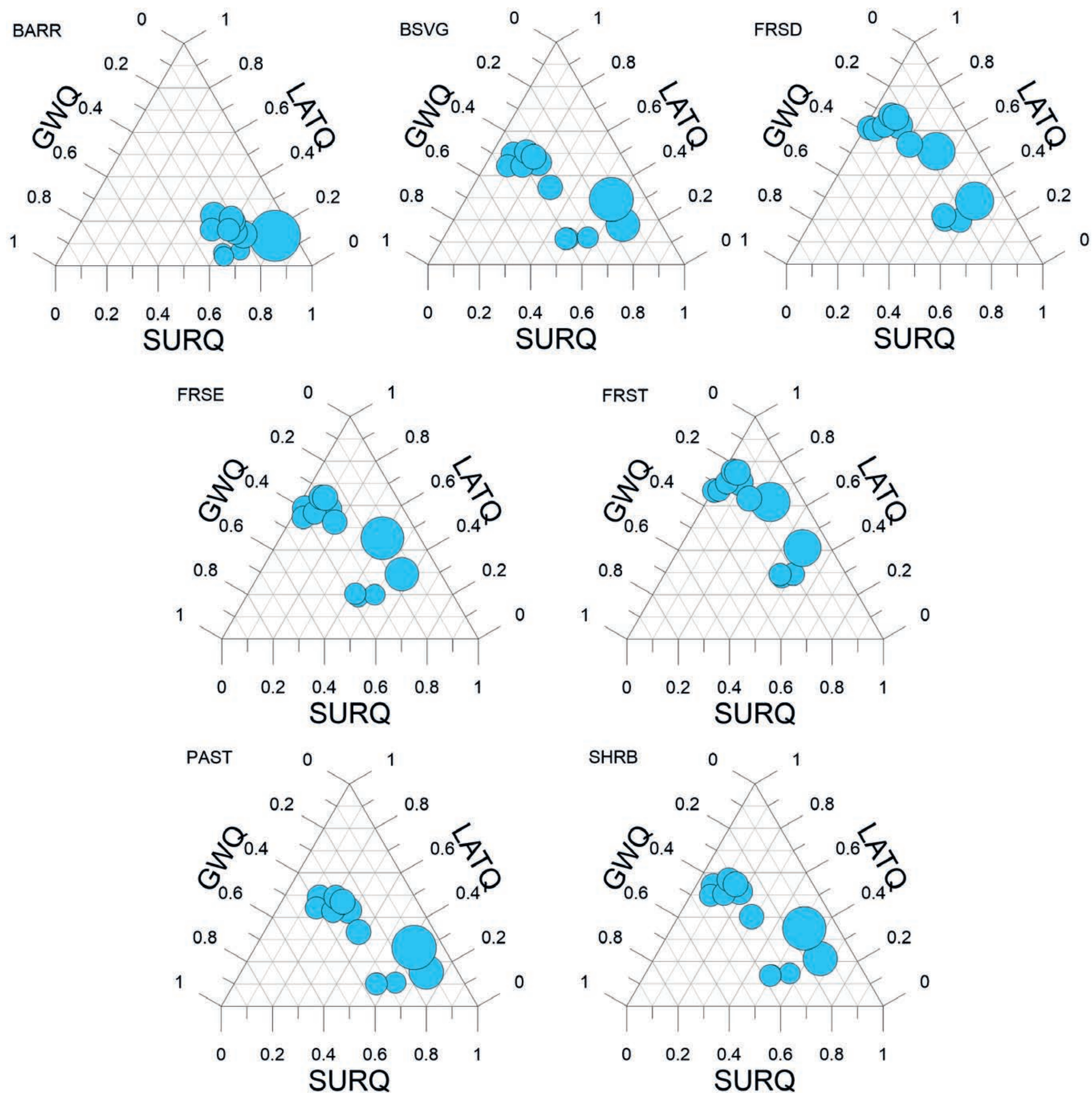

Fig. 8 Monthly average contributions to total runoff for hydrotopes covered by all seven inspected landcovers (indexes explained in Table 2). Obtained as an average from the SWAT model simulation in 1991–2021. GWQ, LATQ and SURQ represent groundwater, subsurface lateral and direct flow.

with huge immission loads in the 1980s (Vacek et al. 1999). The slight decline from 1997 onwards was due to damages caused by icing. After a period of recovery between 2001 and 2006, the cyclonic storm Kyrill in 2007 followed by repeated bark beetle outbreaks, are the main factors for the subsequent decline of coniferous forest in the studied catchment (ÚHÚL 2022).

In central European mountainous environments such as the Krkonoše Mts., the Krušné Mountains or the Tatra Mountains, land cover changes have been extensively studied, particularly in relation to the decline of spruce-dominated forests due to air pollution, bark beetle outbreaks and wind disturbances (Zagajewski et al. 2021; Kupková et al. 2018). The expansion of subalpine shrubland and secondary succession in formerly forested areas has been documented through remote sensing analysis, highlighting the influence of climate change and disturbance regimes on vegetation dynamics (Arekhi et al. 2018; Mašek et al. 2023). Additionally, historical aerial imagery has been used to reconstruct long-term land cover trajectories, revealing for example shrub encroachment due to artificial planting of the dwarf pine in the past as well as the cessation of former farming and, most probably, also by global change in the KRNAP National Park (Potůčková et al. 2021).

4.2 Input data and hydrological modelling

The input data and the performance of the hydrological model are subject to various uncertainties. One major source of bias is the underestimation of precipitation when using standard rain gauges. This occurs because rain gauge measurements often contain significant errors, particularly during the winter months (Dingman 2015). Even after corrections, the accuracy of winter precipitation estimates remains uncertain (error up to 40% precipitation volume), as measuring precipitation in mountainous regions involves several challenges and limitations (Sevruk 2005).

The hydrological SWAT was chosen as a modelling tool as it is commonly used for the assessment of land cover change influence on the hydrological regime (Siqueira et al. 2021; Valencia et al. 2024). The primary limitation in the model's performance stemmed from the snowmelt and snow accumulation/melt routine, which relied on the degree-day approach. However, it was documented that the degree-day approach proved to be the second-best choice (radiation balance approach was not chosen as the data were not available) for the modelling of snowmelt dynamics in Central Europe (Girons Lopez et al. 2020). The representation of particular processes representing the evaporation from the catchment also contains several crucial simplifications that do not take into account the non-stationary character of rainfall interception in the mountainous regions (Kofroňová et al. 2021), the differences in physiological response of different vegetation to drought-stress (Gebhardt et al. 2023) or forest floor evaporation (Floriancic et al. 2023). All these model simplifications can cause less pronounced effect of land cover change on hydrological régime of the area (Fatichi et al. 2016). Hence, the results of such studies represent the lower boundary of future possible development.

4.3 The influence of land cover change on the hydrological regime

The observed changes of land cover in Upper Úpa catchment did not cause any significant alternation of the hydrological regime. However, we have observed significant changes in runoff generation among inspected landcovers, namely in the representation of direct runoff and deeper and slower pathways represented by subsurface lateral and groundwater flows. It is necessary to stress that the observed changes in runoff generation processes are results of the modelling exercise and it was not observed in the real environment. A brief investigation of precipitation patterns preceding the modelling itself revealed no shift in daily precipitation totals in the ten-year periods of 1990–2000, 2001–2010, and 2011–2021, and thus climate-induced changes should not represent a significant contributing factor to the presented results.

In Central Europe, the non-detectable changes in hydrological regime despite changes in land cover were also reported by Bernsteinová et al. (2015) in Šumava Mts. and Wojkowski et al. (2022) in Upper Vistula basin. All these changes were only of local importance and they were probably dampened on the larger scale of gauging station, which was hypothesized by Blöschl et al. (2007). It needs to be stressed that these studies are limited to mountainous areas having high precipitation sums and low evapotranspiration. This possibly stands behind negligible differences in actual evapotranspiration among particular vegetation covers. We hypothesize that the differences in actual evapotranspiration will be larger in lower altitudes with more favorable meteorological conditions (Ji et al. 2024), lower amount of deposited water (influencing the importance of interception) and occurrence of soil water deficits will play more important role.

Experiments aimed at revealing the influence of extreme land cover changes on the local hydrology showed that changes from all coniferous forest to transitional woodlands and shrubs (including dwarf pine) and vice versa did not cause significant changes in the results of the hydrological modelling. The modelled changes affected about 30% of the catchment area. Thus, the real changes in the area of land cover classes, which reach a maximum of 12% of the study area (corresponding to the change of transitional woodlands between 1991 and 2005), as well as the accuracy gained, do not have a major impact on the changes in runoff.

5. Conclusion

This study aimed to quantify land cover change in a mid-latitude mountainous catchment over the past 30 years, to represent the catchment water balance using a semi-distributed hydrological model, and to assess the influence of time-varying land cover on the hydrological response, including afforestation and deforestation scenarios.

The land cover analysis captured the main patterns of vegetation change, including the transition from coniferous forests to transitional woodland-shrub and the gradual increase in mixed forests at the expense of grasslands. The influence of disturbance events, such as windthrows and bark beetle outbreaks, was evident, especially after 2007. Although finer-scale structural changes in vegetation could not be fully resolved, the results offered a sufficiently robust basis to explore long-term shifts in vegetation cover and their potential hydrological implications.

The hydrological simulations indicated that land cover changes had only a limited impact on the overall water balance, while influencing the partitioning between different runoff components. The results suggest that the Horní Úpa catchment exhibits a certain resilience to land cover variability under current climatic conditions. However, these results are not generally applicable, they apply only on the impact of land cover change of mountainous forest-dominated catchments. Additionally, the static representation of land cover and soils in the SWAT model, together with the fixed delineation of hydrological response units, may have constrained the model's sensitivity to more subtle changes in vegetation dynamics.

Despite these limitations, the study provides an important contribution to the understanding of interactions between land cover change and hydrological processes in complex mountainous environments. It demonstrates the potential of combining satellite-based land cover mapping with process-based hydrological modelling to explore landscape-hydrology linkages over long timescales. Future research should focus on integrating higher-resolution and temporally dynamic land cover data, improving field-based validation, refining soil and vegetation parameterization, and employing models capable of capturing dynamic vegetation processes. Additionally, we encourage analogous studies in catchments with different topography, climate, soil and land cover characteristics. These steps will help to further enhance the reliability of hydrological assessments under changing land use and climate conditions in sensitive mountain catchments such as the Krkonoše Mountains.

Acknowledgments

This work was supported by the Technological Agency of the Czech Republic (SS05010124), the institutional support of the Czech Academy of Sciences (RVO: 67985874).

References

Abbaspour, K. C. (2013): SWAT-CUP 2012: SWAT calibration and uncertainty programs – A User manual. Swiss Federal Institute of Aquatic Science and Technology, Eawag. Available online: https://swat.tamu.edu/media/114860/usermanual_swatcup.pdf (accessed on 11 October 2025).

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., Kløve, B. (2015): A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology 524, 733–752, https://doi.org/10.1016/j.jhydrol.2015.03.027.

Alaoui, A., Lipiec, J., Gerke, H. H. (2011): A review of the changes in the soil pore system due to soil peformation: A hydrodynamic perspective. Soil and Tillage Research 115–116, 1–15, https://doi.org/10.1016/j.still.2011.06.002.

Archer, N. A. L., Bonell, M., Coles, N., MacDonald, A. M., Auton, C. A., Stevenson, R. (2013): Soil characteristics and landcover relationships on soil hydraulic conductivity at a hillslope scale: A view towards local flood management. Journal of Hydrology 497, 208–222, https://doi.org/10.1016/j.jhydrol.2013.05.043.

Arekhi, M., Yesil, A., Ozkan, U. Y. et al. (2018): Detecting treeline dynamics in response to climate warming using forest stand maps and Landsat data in a temperate forest. For Ecosyst 5: 23, https://doi.org/10.1186/s40663-018-0141-3.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santi, C., Harmel, R. D., Van Griensven, A., Van Liew, M. W., Kannan, N., Jha, M. K. (2012): SWAT: model use, calibration, and validation. Transactions of the ASABE 55(4), 1491–1508, http://dx.doi.org/10.13031/2013.42256.

Baig, M. H. A., Zhang, L., Shuai, T., Tong, Q. (2014): Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. Remote Sensing Letters 5(5), 423–431, https://doi.org/10.1080/2150704X.2014.915434.

Bernsteinová, J., Bässler, C., Zimmermann, L., Langhammer, J., Beudert, B. (2015): Changes in runoff in two neighbouring catchments in the Bohemian Forest related to climate and land cover changes. Journal of Hydrology and Hydromechanics 63(4), 342–352, https://doi.org/10.1515/johh-2015-0037.

Blackburn, D. A., Oliphant, A. J., Davis, J. D. (2021): Carbon and Water Exchanges in a Mountain Meadow Ecosystem, Sierra Nevada, California. Wetlands 41: 39, https://doi.org/10.1007/s13157-021-01437-2.

Blöschl, G., Ardoin-Bardin, S., Bonell, M. et al. (2007): At what scales do climate variability and land cover change impact on flooding and low flows? Hydrological Process 21(9), 1241–1247, https://doi.org/10.1002/hyp.6669.

Blöschl, G. (2022): Three hypotheses on changing river flood hazards. Hydrology and Earth System Sciences 26(19), 5015–5033, https://doi.org/10.5194/hess-26-5015-2022.

Breiman, L. (2001): Random Forests. Machine Learning 45, 5–32, https://doi.org/10.1023/A:1010933404324.

Chaves, M. E. D., Picoli, M. C. A., Sanches, I. D. (2020): Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. Remote Sensing 12(18): 3062, https://doi.org/10.3390/rs12183062.

Chen, Z., Wang, W., Woods, R. A., Shao, Q. (2021): Hydrological effects of change in vegetation components across global catchments. Journal of Hydrology 595: 125775, https://doi.org/10.1016/j.jhydrol.2020.125775.

Cognard-Plancq, A. L., Marc, V., Didon-Lescot, J. F., Normand, M. (2001): The role of forest cover on streamflow down sub-Mediterranean mountain watersheds: a modelling approach. Journal of Hydrology 254(1–4), 229–243, https://doi.org/10.1016/S0022-1694(01)00494-2.

Crist, E. P., Cicone, R. C. (1984): A Physically-Based Transformation of Thematic Mapper Data – The TM Tasseled Cap. IEEE Transactions on Geoscience and Remote Sensing 22(3), 256–263, https://doi.org/10.1109/TGRS.1984.350619.

ČÚZK – Czech Office for Surveying, Mapping and Cadastre (2025): Geoprohlížeč [Online]. Available online: https://ags.cuzk.gov.cz/geoprohlizec/ (accessed on 11 October 2025).

Danáčová, M., Földes, G., Labat, M. M., Kohnová, S., Hlavčová, K. (2020): Estimating the effect of deforestation on runoff in small mountainous basins in Slovakia. Water 12(11): 3113, https://doi.org/10.3390/w12113113.
Dingman, S. L. (2015): Physical hydrology. 3rd ed., Long Grove: Waveland Press.
Dubbert, M., Werner, C. (2018): Water fluxes mediated by vegetation: emerging isotopic insights at the soil and atmosphere interfaces. New Phytologist 221(4), 1754–1763, https://doi.org/10.1111/nph.15547.
Fatichi, S., Vivoni, E. R., Ogden, F. L. et al. (2016): An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. Journal of Hydrology 537, 45–60, https://doi.org/10.1016/j.jhydrol.2016.03.026.
Feranec, J., Solin, L., Kopecká, M., Oťahel, J., Kupková, L., Štych, P., Bičík, I., Kolář, J., Čerba, O., Soukup, T., Brodský, L. (2014): Analysis and expert assessment of the semantic similarity between land cover classes. Progress in Physical Geography: Earth and Environment 38(3), 301–327, https://doi.org/10.1177/0309133314532001.
Feranec, J., Soukup, T., Hazeu, G., Jaffrain, G. (2016): European Landscape Dynamics: CORINE Land Cover Data. 1st ed., CRC Press, https://doi.org/10.1201/9781315372860.
Floriancic, M. G., Allen, S. T., Meier, R., Truniger, L., Kirchner, J. W., Molnar, P. (2023): Potential for significant precipitation cycling by forest-floor litter and deadwood, Ecohydrology 16(2), 1–16, https://doi.org/10.1002/eco.2493.
Foody, G. M. (2009): Sample size determination for image classification accuracy assessment and comparison. International Journal of Remote Sensing 30(20), 5273–5291, https://doi.org/10.1080/01431160903130937.
Gao, L., Wang, X., Johnson, B. A., Tian, Q., Wang, Y., Verrelst, J., Mu, X., Gu, X. (2020): Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review. ISPRS Journal of Photogrammetry and Remote Sensing 159, 364–377, https://doi.org/10.1016/j.isprsjprs.2019.11.018.
Gebhardt, T., Hesse, B. D., Hikino, K., Kolovrat, K., Hafner, B. D., Grams, T. E. E., Häberle, K. H. (2023): Repeated summer drought changes the radial xylem sap flow profile in mature Norway spruce but not in European beech. Agricultural and Forest Meteorology 329: 109285, https://doi.org/10.1016/j.agrformet.2022.109285.
Girons Lopez, M., Vis, M. J. P., Jenicek, M., Griessinger, N., Seibert, J. (2020): Assessing the degree of detail of temperature-based snow routinesfor runoff modelling in mountainous areas in central Europe. Hydrology and Earth System Sciences 24(9), 4441–4461, https://doi.org/10.5194/hess-24-4441-2020.
Griffiths, P., Kuemmerle, T., Baumann, M. et al. (2014): Forest disturbances, forest recovery, and changes in forest types across the Carpathian ecoregion from 1985 to 2010 based on Landsat image composites. Remote Sensing of Environment 151, 72–88, https://doi.org/10.1016/j.rse.2013.04.022.
Healey, S. P., Cohen, W. B., Zhiqiang, Y., Krankina, O. N. (2005): Comparison of tasseled cap-based Landsat data structures for use in forest disturbance detection. Remote Sensing of Environment 97(3), 301–310, https://doi.org/10.1016/j.rse.2005.05.009.
Huang, C., Goward, S. N., Masek, J. G., Thomas, N., Zhu, Z., Vogelmann, J. E. (2010): An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. Remote Sensing of Environment 114(1), 183–198, https://doi.org/10.1016/j.rse.2009.08.017.
Jasechko, S., Sharp, Z. D., Gibson, J. J., Birks, J., Yi, Y., Fawcett, P. J. (2013): Terrestrial water fluxes dominated by transpiration. Nature 496, 347–350, https://doi.org/10.1038/nature11983.
Ji, Z., Wang, Y., Wang, L. (2024): Threshold identification of evapotranspiration under different land-use types in the Loess Plateau, China. Journal of Hydrology: Regional Studies 53: 101780, https://doi.org/10.1016/j.ejrh.2024.101780.
Kofroňová, J., Šípek, V., Hnilica, J., Vlček, L., Tesař, M. (2021): Canopy interception estimates in a Norway spruce forest and their importance for the hydrological modelling. Hydrological Sciences Journal 66(7), 1233–1247, https://doi.org/10.1080/02626667.2021.1922691.
KRNAP – Krkonošský národní park (Krkonoše National Park Administration) (2025): Veřejný mapový portal. Available online: https://ags.krnap.cz/mapy/prohlizecka/ (accessed on 11 October 2025).
Kupková, L., Potůčková, M., Lhotáková, Z., Albrechtová, J. (2018): Forest cover and disturbance changes, and their driving forces: A case study in the Ore Mountains, Czechia, heavily affected by anthropogenic acidic pollution in the second half of the 20th century. Environmental Research Letters 13: 095008, https://doi.org/10.1088/1748-9326/aadd2c.
Kupková, L., Bičík, I., Jeleček, L. (2021): At the Crossroads of European Landscape Changes: Major Processes of Landscape Change in Czechia since the Middle of the 19th Century and Their Driving Forces. Land 10(1): 34, https://doi.org/10.3390/land10010034.
Kupková, L., Červená, L., Potůčková, M., Lysák, J., Roubalová, M., Hrázský, Z., Březina, S., Epstein, H. E., Müllerová, J. (2023): Towards reliable monitoring of grass species in nature conservation: Evaluation of the potential of UAV and PlanetScope multi-temporal data in the Central European tundra. Remote Sensing of Environment 294: 113645, https://doi.org/10.1016/j.rse.2023.113645.
Lachassagne, P., Wyns, R., Dewandel, B. (2011): The fracture permeability of Hard Rock Aquifers is due neither to tectonics, nor to unloading, but to weathering processes. Terra Nova 23(3), 145–161, https://doi.org/10.1111/j.1365-3121.2011.00998.x.
Launiainen, S., Guan, M., Salmivaara, A., Kieloaho, A. J. (2019): Modeling boreal forest evapotranspiration and water balance at stand and catchment scales: a spatial approach. Hydrology and Earth System Sciences 23(8), 3457–3480, https://doi.org/10.5194/hess-23-3457-2019.
Mašek, J., Tumajer, J., Lange, J. et al. (2023): Shifting climatic responses of tree rings and NDVI along environmental gradients. Science of The Total Environment 908: 168275, https://doi.org/10.1016/j.scitotenv.2023.168275.
Meier, R., Schwaab, J., Seneviratne, S. I. et al. (2021): Empirical estimate of forestation-induced precipitation

changes in Europe. Nature Geoscience 14, 473–478, https://doi.org/10.1038/s41561-021-00773-6.
Mensah, J. K., Ofosu, E. A., Yidana, S. M., Akpoti, K., Kabo-bah, A. T. (2022): Integrated modeling of hydrological processes and groundwater recharge based on land use land cover, and climate changes: A systematic review. Environmental Advances 8: 100224, https://doi.org/10.1016/j.envadv.2022.100224.
Moriasi, D. N., Gitau, M. W., Pai, N., Daggupati, P. (2015): Hydrologic and water quality models: performance measures and evaluation criteria. Transactions of the ASABE 58(6), 1763–1785, https://doi.org/10.13031/trans.58.10715.
Němečková, S. (2008): The Map of soil subtypes of the Czech part of the Elbe river basin for the hydrological model SWIM. IH CAS, CD-ROM.
Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., Wulder, M. A. (2014): Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment 148, 42–57, https://doi.org/10.1016/j.rse.2014.02.015.
Pielke, R. A., Avissar, R., Raupach, M., Dolman, A. J., Zeng, X., Denning, A. S. (1998): Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate. Global Change Biology 4(5), 461–475, https://doi.org/10.1046/j.1365-2486.1998.t01-1-00176.x.
Potůčková, M., Kupková, L., Červená, L., Lysák, J., Krause, D., Hrázký, Z., Březina, S., Müllerová, J. (2021): Towards resolving conservation issues through historical aerial imagery: vegetation cover changes in the Central European tundra. Biodiversity and Conservation 30, 3433–3455, https://doi.org/10.1007/s10531-021-02255-y.
Priestley, C. H. B., Taylor, R. J. (1972): On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. Monthly Weather Review 100(2), 81–92, https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2.
Ramon, F. B., Millington, J. D. A., Moran, E. F., Batistella, M., Liu, J. (2020): Three decades of land-use and land-cover change in mountain regions of the Brazilian Atlantic Forest. Landscape and Urban Planning 204: 103948, https://doi.org/10.1016/j.landurbplan.2020.103948.
Ritchie, J. T. (1972): A model for predicting evaporation from a row crop with incomplete cover. Water Resources Research 8(5), 1204–1213, https://doi.org/10.1029/WR008i005p01204.
Senf, C., Pflugmacher, D., Hostert, P., Seidl, R. (2017): Using Landsat time series for characterizing forest disturbance dynamics in the coupled human and natural systems of Central Europe. ISPRS Journal of Photogrammetry and Remote Sensing 130, 453–463, https://doi.org/10.1016/j.isprsjprs.2017.07.004.
Sevruk, B. (2005): Rainfall measurement: gauges. In Anderson, M. G. (Ed.), Encyclopedia of Hydrological Sciences. 3rd ed., New York: Wiley, 230–257, https://doi.org/10.1002/0470848944.hsa038.
Sieker, F. (2000): Investigations of the Effects of On-Site Stormwater Management Measures in Urban and Agricultural Areas on Floods. In Marsalek, J., Watt, W. E., Zeman, E., Sieker, F. (Eds.), Flood Issues in Contemporary Water Management. NATO Science Series 71. Dordrecht: Springer, https://doi.org/10.1007/978-94-011-4140-6_31.
Siqueira, P. P., Oliveira, P. T. S., Bressiani, D., Neto, A. A. M., Rodrigues, D. B. B. (2021): Effects of climate and land cover changes on water availability in a Brazilian Cerrado basin. Journal of Hydrology: Regional Studies 37: 100931, https://doi.org/10.1016/j.ejrh.2021.100931.
Sonnenborg, T. O., Christiansen, J. R., Pang, B., Bruge, A., Stisen, S., Gundersen, P. (2017): Analyzing the hydrological impact of afforestation and tree species in two catchments with contrasting soil properties using the spatially distributed model MIKE SHE SWET. Agricultural and Forest Meteorology 239, 118–133, https://doi.org/10.1016/j.agrformet.2017.03.001.
Sutanto, S. J., van den Hurk, B., Dirmeyer, P. A., Seneviratne, S. I., Röckmann, T., Trenberth, K. E., Blyth, E. M., Wenninger, J., Hoffmann, G. (2014): HESS Opinions "A perspective on isotope versus non-isotope approaches to determine the contribution of transpiration to total evaporation". Hydrology and Earth System Sciences 18(8), 2815–2827, https://doi.org/10.5194/hess-18-2815-2014.
Talukdar, S., Singha, P., Mahato, S., Shahfahad, P. S., Liou, Y. A., Rahman, A. (2020): Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations – A Review. Remote Sensing 12(7): 1135, https://doi.org/10.3390/rs12071135.
Tolasz, R. (2007): Atlas podnebí Česka: Climate atlas of Czechia. Prague: Czech Hydrometeorological Institute.
Trinh, D. H., Chui, T. F. M. (2013): Assessing the hydrologic restoration of an urbanized area via an integrated distributed hydrological model. Hydrology and Earth System Sciences 17(12), 4789–4801, https://doi.org/10.5194/hess-17-4789-2013.
ÚHÚL (2022): Oblastní plán rozvoje lesů: Přírodní lesní oblast 22 – Krkonoše. Analýza stavu a vývoje. ÚHÚL, Brandýs nad Labem – Stará Boleslav.
Vacek, S., Bastl, M., Lepš, J. (1999): Vegetation changes in forest of the Krkonoše Mts over a period of air pollution stress (1980–1995). Plant Ecology 143, 1–11, https://doi.org/10.1023/A:1009833313509.
Valencia, S., Villegas, J. C., Hoyos, N., Duque-Villegas, M., Salazar, J. F. (2024): Streamflow response to land use/land cover change in the tropical Andes using multiple SWAT model variants. Journal of Hydrology: Regional Studies 54: 101888, https://doi.org/10.1016/j.ejrh.2024.101888.
Wang, H., Tetzlaff, D., Soulsby, C. (2018): Modelling the effects of land cover and climate change on soil water partitioning in a boreal headwater catchment. Journal of Hydrology 558, 520–531, https://doi.org/10.1016/j.jhydrol.2018.02.002.
Winkler, K., Fuchs, R., Rounsevell, M., Herold, M. (2021): Global land use changes are four times greater than previously estimated. Nature Communications 12: 2501, https://doi.org/10.1038/s41467-021-22702-2.
Wojkowski, J., Walega, A., Radecki-Pawlik, A., Mlynski, D., Lepeška, T. (2022): The influence of land cover changes on landscape hydric potential and river flows: Upper Vistula, Western Carpathians. Catena 210: 105787, https://doi.org/10.1016/j.catena.2021.105878.
Yang, J. L., Zhang, G. L. (2011): Water infiltration in urban soils and its effects on the quantity and quality of runoff.

Journal of Soils and Sediments 11, 751–761, https://doi.org/10.1007/s11368-011-0356-1.
Yu, X., Lamačová, A., Duffy, C., Krám, P., Hruška, J. (2016): Hydrological model uncertainty due to spatial evapotranspiration estimation methods. Computers and Geosciences 90(Part B), 90–101, https://doi.org/10.1016/j.cageo.2015.05.006.
Zagajewski, B., Kluczek, M., Raczko, E., Njegovec, A., Dabija, A., Kycko, M. (2021): Comparison of Random Forest, Support Vector Machines, and Neural Networks for Post-Disaster Forest Species Mapping of the Krkonoše/Karkonosze Transboundary Biosphere Reserve. Remote Sensing 13(13): 2581, https://doi.org/10.3390/rs13132581.
Zelíková, N., Toušková, J., Kocum, J., Vlček, L., Tesař, M., Bouda, M., Šípek, V. (2025): Divergent water balance trajectories under two dominant tree species in montane forest catchment shifting from energy- to water-limitation. Hydrology and Earth System Sciences (Preprint, accepted), https://doi.org/10.5194/hess-2024-244.
Zhu, Z., Woodcock, C. E. (2014): Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. Remote Sensing of Environment 152, 217–234, https://doi.org/10.1016/j.rse.2014.06.012.

An introduction to the demography of genocide: The Timor-Leste case study

Adam Dušek*

Charles University, Faculty of Science, Department of Demography and Geodemography, Czechia
* Corresponding author: adam.dusek@natur.cuni.cz

ABSTRACT
This paper introduces the demography of genocide, an interdisciplinary branch of demographic research that examines the impact of genocides on populations. This is achieved by defining the field and discussing its limitations, such as a limited institutional and methodological framework and a lack of data sources. The analytical section presents a case study on the demographic effects of the Indonesian occupation of Timor-Leste, starting with the historical context. It then analyzes population dynamics, mortality, fertility, and population structures based on demographic indicators, primarily derived from the World Population Prospects 2024 database. The findings show that the genocide profoundly affected Timor-Leste's population development, primarily through a significant deterioration in mortality, which was highly uneven across age and gender. Excess mortality was particularly pronounced among children and men. Large-scale migration waves occurred, indirectly influencing demographic processes. The impact on fertility is complex and cannot be generalized, but the decline in female mortality likely contributed to fertility increases. The simultaneous effect of these factors significantly altered the structure of the population, leading to sharp fluctuations in the sex ratio, an acceleration of population aging, and disruptions in age distribution. Total population losses due to the occupation were estimated at 177,000 people using a counterfactual population projection.

KEYWORDS
demography; genocide; Timor-Leste; Indonesia; excess mortality

Received: 31 March 2025
Accepted: 2 October 2025
Published online: 24 October 2025

Dušek, A. (2025): An introduction to the demography of genocide: The Timor-Leste case study. AUC Geographica 60(2), 204–220
https://doi.org/10.14712/23361980.2025.19

1. Introduction

Despite its significant impact on population dynamics, genocide remains an underexplored topic in demography compared to other areas of research. Genocide, a term of relatively recent origin with deeply negative connotations, describes a phenomenon that has occurred throughout human history, from ancient to modern times. One of the most studied examples is the Holocaust, where Nazi Germany's policies resulted in approximately six million Jewish deaths, with the total death toll reaching up to 17 million when including other ethnic groups (Hillberg 2003; Niewyk and Nicosia 2000). However, many genocides, despite their significance, have received limited attention from experts, particularly in demographic research.

Genocides not only inflict immense suffering but also deeply disrupt population processes such as mortality, fertility, and migration. These events can have lasting consequences, modifying population structures and dynamics beyond the immediate crisis period. For example, migration caused by genocide can indirectly impact populations originally unexposed to such events. Thus, deeper demographic analysis of genocides is necessary to understand their effects fully.

This paper briefly introduces the demography of genocides as a branch of demographic research. The main part of this paper presents a case study on Timor-Leste during the late 20th century, one of the lesser-studied genocides, to illustrate the importance of such research.

Case studies are a fundamental tool in genocide studies, as they illustrate the diverse forms and contexts in which genocides occur. Comprehensive documentation and reliable data are crucial for demographic research on genocidal cases, enabling the analysis of mortality, fertility, migration, and changes in gender and age structures. An essential aspect of the case study is situating the studied events within their historical context, providing clarity and deeper understanding of the subject matter.

Although genocide is not a frequent topic of demographic research, the growing global security challenges highlight its importance. For example, Europe's largest conflict since WWII is ongoing, alongside other global armed conflicts prone to genocidal acts. A demographic analysis of genocides helps to understand their causes, development, and effects on population change. Such insights can be applied, for instance, to support legal processes or to forecast the recovery of affected populations following an event, including during ongoing crises. By advancing this area of research, demography can contribute to better understanding and responding to these tragic events. The UN Convention defines genocide as "any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial, or religious group" (UN 1948, p. 1). These acts include killing members of the group, causing them serious bodily or mental harm, imposing living conditions intended to destroy the group, preventing births, and forcibly transferring children out of the group. To ensure research validity, only events meeting the UN convention's definition of genocide should be analyzed.

1.1 Genocidal research in population studies

Lemkin's (1944) work defined genocide and sparked academic interest in the field (Bloxham and Moses 2010). Genocide studies, a multidisciplinary field involving history, sociology, and law, aims to prevent genocides by understanding their causes, mechanisms, and consequences (Schaller 2011; Jones 2017). Initially focused on the Holocaust, research expanded post-Cold War to colonial and gendered violence, including genocidal rape (Von Joeden-Forgey 2010). Despite its growth, the field faces challenges with definitions, methodologies, and limited use of demographic methods (Weiss-Wendt 2008; Jones 2017). Even so, Ball et al. (2000) made a significant contribution by emphasizing a data-driven approach to investigating mass atrocities through the systematic collection of primary data. For instance, the work outlined how to improve the creation of victim databases by incorporating statistical methods that reduce the margin of error in victim estimates, especially when the data are incomplete.

While early demographic research on war, exemplified by the works of Hauser (1942) and Vincent (1947) analyzed its impacts, genocide remained largely overlooked. Interest grew in the 1980s with studies on the Holocaust and the Khmer Rouge regime, including victim estimates (Yehuda and Magos 1983; Ea 1981; Heuveline 1998). Seltzer (1998) presents an uncommon perspective, showing that the role of statistics and demography goes beyond their beneficial use in legislative and evidentiary processes, as they have also been misused to facilitate the planning and execution of large-scale violence. This perspective can be extended to include another crucial view: the deliberate manipulation of data to conceal the consequences of violence. The 1937 Soviet census is a prime example. Its count of approximately 162 million people fell short of the leadership's expected 180 million (Johnson 1992). The data also revealed a population decline in specific regions since the 1926 census, which could, among other things, document the actual impact of ethnic cleansing, including the infamous famines in Ukraine and Kazakhstan (Poliakov et al. 1992). The census was subsequently invalidated, and its authors were persecuted.

In 2002, the IUSSP launched a panel on the demography of armed conflicts, an emerging field that facilitates interdisciplinary research on the impacts not only of wars but also of genocides (Brunborg and Tabeau 2005). The first comprehensive study specifically addressing the demographic effects of genocide was The Demography of Genocide by Kugler (2016), which examined disruptions to gender-age structures, migration patterns, and recovery rates.

The demography of genocides focuses on immediate and long-term impacts on mortality, fertility, and migration for directly and indirectly affected populations. While primarily employing demographic methods, this field also integrates interdisciplinary approaches from history, sociology, and genocide studies to contribute to genocide prevention.

The accompanying violence leads to infrastructure destruction, and limited access to high-quality data remains a significant research limitation (Brunborg and Tabeau 2005; Kugler 2016). Nevertheless, demographic research aids genocide prevention and legal processes, such as victim estimation in the Bosnian genocide (Brunborg et al. 2003).

1.2 The Timor-Leste case: Historical context

The Indonesian invasion of Timor-Leste in December 1975 led to catastrophic conditions, causing tens of thousands of deaths. The crisis stems from colonial conflict between Portugal and the Netherlands, which divided Timor in 1904, leaving the eastern half under Portuguese rule (Joseph and Hamaguchi 2014). After Portugal's 1974 Carnation Revolution, decolonization efforts began. Two major factions, FRETILIN and UDT, initially formed a coalition for independence (Dunn 2012). But tensions escalated as Indonesia opposed East Timorese independence, fearing separatist movements and FRETILIN's leftist ideology (Kiernan 2008).

In August 1975, UDT, under Indonesian influence, attempted a coup, triggering civil war with FRETILIN. By December, Indonesia invaded with the tacit consent of the West (Dunn 2012). Despite armed resistance, Indonesian forces quickly occupied Timor-Leste, forcing many FRETILIN fighters and civilians into guerrilla warfare. The occupation was marked by severe repression, including massacres, torture, and sexual violence (Kiernan 2008). Forced relocations, agricultural disruption, and disease led to widespread famine. An information embargo prevented external scrutiny until 1979, when the International Red Cross exposed the crisis and prompted condemnation (Robinson 2023).

By the late 1970s, FRETILIN's resistance weakened due to starvation and military losses. Indonesian forces employed brutal tactics such as village massacres, human shields, and forced displacement to sever guerrilla support (Budiardjo and Liong 1984). A fragile ceasefire in 1983 failed, leading to renewed violence and cultural suppression, including banning the Portuguese language and promoting Indonesian migration (Kiernan 2008). By the 1990s, Indonesian settlers comprised a fifth of the population and controlled key positions (Dunn 2012).

The 1991 Santa Cruz massacre, in which 300 peaceful protesters were killed, triggered global outrage after footage surfaced, pressuring Western nations to act (Joseph and Hamaguchi 2014). The Nobel Peace Prize awarded to East Timorese leaders in 1996 and increased UN involvement further challenged Indonesia's claims (Robinson 2023). Following the May 1998 riots, Indonesia agreed to a 1999 UN-supervised referendum. Despite violent intimidation, nearly 80% voted for independence, leading to full independence on May 20, 2002 (Dunn 2012; Robinson 2023).

Post-independence, Timor-Leste faced instability, poverty, and weak accountability for wartime atrocities. Indonesia refused to extradite perpetrators, and only a quarter of the indicted individuals were prosecuted (Dunn 2012). Despite criticism, Timor-Leste has reportedly prioritized economic development over justice (Thu 2019).

2. Materials and methods

The World Population Prospects 2024 (WPP 2024) serves as the primary demographic data source for this case study. It is one of the most comprehensive sources of global population data and is particularly valuable for Timor-Leste. Due to the country's fragmented history under Portuguese, Japanese, and Indonesian rule, there is a lack of unified population data, making WPP 2024 essentially the only comprehensive source.

While Timor-Leste has its own statistical office, the data it provides are limited and difficult to access, and they only sporadically cover the period before the office's founding in 2000. The registration of vital events in the country remains significantly incomplete. For example, a 2016 Demographic and Health Surveys Program (DHS) survey found that 40% of children under five years of age were not registered (DHS 2024).

Consequently, the data in WPP 2024 are based on a combination of censuses, sample surveys, and numerical adjustments. It is important to emphasize that these figures are estimates and may be subject to errors during both initial data collection and subsequent processing. Therefore, all data presented in this study should be interpreted with appropriate caution.

To outline the historical context, literature cited in the case study itself was used including publications by historians, genocidologists, and demographers. The following subsections detail the specific procedures and indicators used to analyze demographic development within the case study.

2.1 Methods of population dynamics analysis

The population dynamics analysis utilized population growth, net migration, and vital event data, focusing on total population, births, deaths, net migration, natural increase, and growth rate. A population projection, constructed using Spectrum 6 software (Avenir Health 2024), estimated population losses due to

Timor-Leste's decolonization, civil war, and Indonesian occupation. Spectrum's DemProj tool employs the cohort-component method which conceptualizes population development as a composite process consisting of fertility, mortality, and migration (Preston et al. 2001). The projection modeled a hypothetical scenario removing excess mortality by linearly interpolating life expectancy at birth between 1973 and 2000. Survival probabilities were derived using UN model life tables. Population losses were calculated as the difference between empirical (WPP 2024) and projected population counts. While acknowledging limitations, such as ignoring interdependence of mortality, fertility and migration or relying on estimated input data, this projection offers valuable insights into estimating population losses.

2.2 Methods of mortality analysis

The study utilized WPP (2024) indicators, including infant and child mortality rates and life expectancy at various ages, to assess mortality trends. A key metric was the directly standardized mortality rate for men and women, using the WHO (2001) New Standard Population for comparison. Age- and sex-specific mortality rates ($ASMR_x$) were calculated from WPP (2024) data, allowing standardization and analysis of mortality changes.

$$ {}_tASMR_x = \frac{{}_tD_x}{{}_tP_x} $$

${}_tD_x$ = the number of deaths between ages x and x+1 for year t
${}_tP_x$ = the population between ages x and x+1 for year t

Next step leads to a directly standardized mortality rate (SDR) per 1000:

$$ SDR = \frac{\Sigma({}_tASMR_x \cdot P_x^{st})}{P^{st}} \cdot 1000 $$

P_x^{st} = the population exposed to the risk of death at age x last birthday in the standard population
P^{st} = the standard population exposed to the risk of death

The decomposition of life expectancy changes, using Shkolnikov and Andreev's (2010) Excel-based program, quantified the contributions of different age groups to the changes in life expectancy between 1973 and 1978, covering both the pre-crisis period and the peak of mortality. The method applies a stepwise replacement of mortality rates to track changes (Andreev et al. 2002).

2.3 Methods of fertility analysis

Indicators from WPP (2024) were used, including the crude birth rate to examine fertility trends and the impact of sex-age structure changes. Other indicators included the total fertility rate (TFR) and net reproduction rate, and the average age of mothers at childbirth. DHS (2003; 2024) data were used to gain a more detailed insight into fertility trends, focusing on completed fertility and the ideal number of children. The WPP contains age-specific fertility rates ($ASFR_x$), however it excludes births to mothers under 15 and over 49. Therefore, the indicator was calculated as follows:

$$ ASFR_x = \frac{B_x}{P_x^f} \cdot 1000 $$

B_x = the number of live births to females at age x,
P_x^f = the total number of females at age x.

The corresponding age limits were chosen as denominators for ages 15 and 49. However, the numerator for $ASFR_{15}$ was adjusted to include all children born to mothers aged 15 and younger. Similarly, the numerator for $ASFR_{49}$ was adjusted to include children born to mothers aged 49 and older.

2.4 Methods of the population structure analysis

The analysis of population structure utilized indicators based on WPP data, including the sex ratio, calculated for individual ages and broad age groups. Population pyramids were used to illustrate changes in the sex-age distribution between 1950 and 2023, reflecting the key changes during and after the Indonesian invasion. To assess demographic aging, we focused on the median age and dependency ratios, including young-age and old-age dependency ratios (YADR and OADR), calculated as follows:

$$ YADR = \frac{P_{0-14}}{P_{15-64}} \cdot 100 $$

$$ OADR = \frac{P_{65+}}{P_{15-64}} \cdot 100 $$

P_{0-14} = the mid-year population aged 0–14,
P_{15-64} = the mid-year population aged 15–64,
P_{65+} = the mid-year population aged 65 and above.

3. Results

In this chapter, we present a demographic analysis of Timor-Leste's population from 1950 to 2023, focusing on the impact of the Indonesian occupation. The historical context is provided in Chapter 1.2. The following demographic analysis examines population dynamics, mortality, fertility, and gender-age structure shifts.

3.1 Population dynamics in Timor-Leste

Timor-Leste's population experienced major fluctuations throughout the 20th century, heavily impacted by conflicts, political instability, and external intervention. In 1920, the population was estimated at 400,000, rising to over 470,000 by 1930 (Hull 2003). However, the Japanese occupation during World War II caused significant disruption, reducing the population to 403,000 by 1946. Recovery began post-war, with the population increasing to 442,000 by 1950 (Fig. 1). However, the specific reasons behind these drastic changes remain unclear.

Population growth continued in the following decades, reaching 592,000 by 1974 (WPP 2024). However, political turmoil in 1975, including a civil war and the Indonesian invasion, marked a turning point. Deaths spiked from around 14,000 annually to over 24,000, leading to the first natural population decline, which occurred again in 1978 (Fig. 2). Despite an almost zero natural increase between 1975 and 1979, Indonesia's migration policy resulted in a

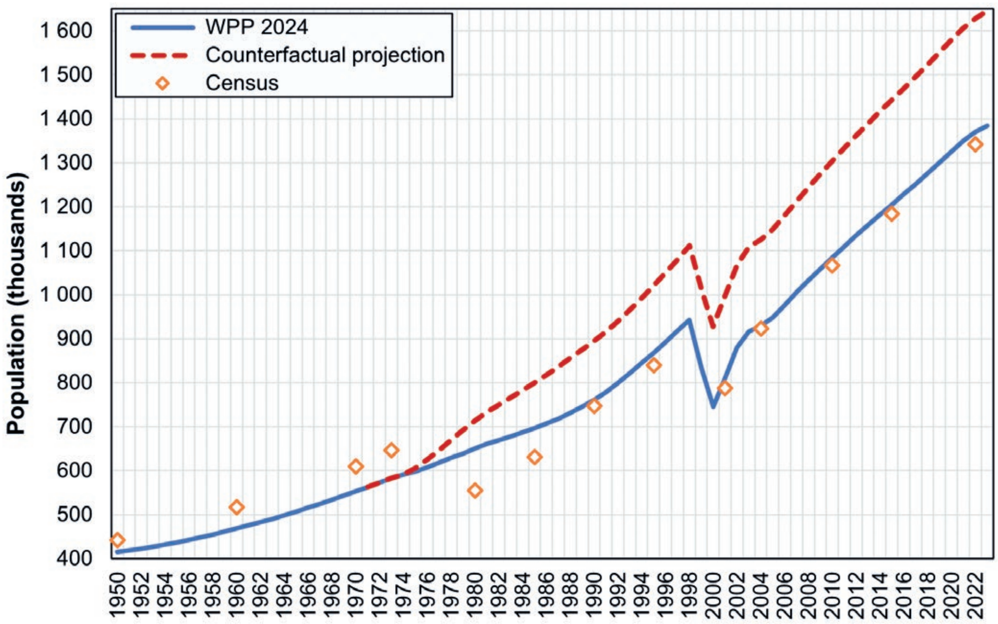

Fig. 1 Population of Timor-Leste, 1950–2023.
Note: The data correspond to July 1 of each given year, while census data refer to various dates within the respective years. The counterfactual projection, developed by the author, aimed to model a hypothetical scenario assuming the absence of excess mortality between 1974 and 1999 (see Chapter 2.1).
Source: WPP 2024, Hull 2003, BPS 1991, Dasvarma 2011, INETL 2023a, INETL 2023b and author's calculation.

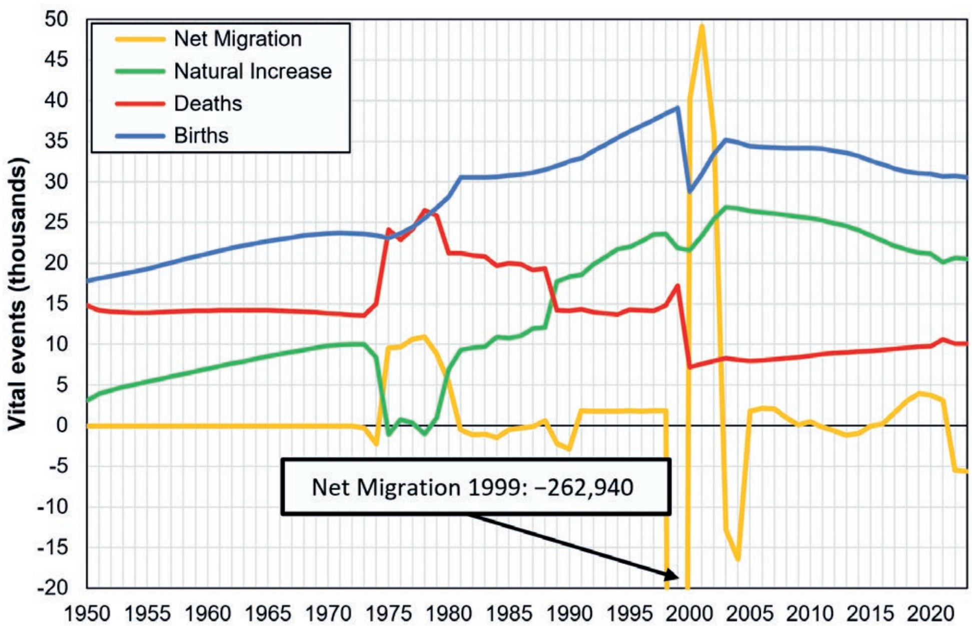

Fig. 2 Vital events in Timor-Leste, 1950–2023.
Source: WPP 2024.

population growth of 50,000 during this period (WPP 2024).

The 1980 census conducted under Indonesian rule reported 555,000 inhabitants, reflecting a loss of 91,000 compared to the 1973 census and 35,500 fewer than WPP's 1974 estimates. Cribb (2001) suggests that these figures may be inaccurate, as many East Timorese likely avoided enumeration due to the ongoing conflict, seeking refuge in remote areas.

As the conflict eased in the 1980s, natural population growth resumed, although in relative terms, it remained below pre-crisis levels. A significant growth occurred in 1989, primarily due to a decrease in mortality. This trend continued throughout the 1990s, driven by declining mortality rates and rising birth rates.

This trend fluctuated in 1999, with a significant increase in the number of deaths, likely connected to the important political changes, such as the withdrawal of Indonesian troops and the independence referendum (Kiernan 2008). However, far more significant was the effect of an unprecedented wave of emigration, which led to a negative migration balance of nearly 263,000. By the end of 1999, Timor-Leste's population had decreased by almost 90,000 people, leaving a total population of 745,000 (WPP 2024).

Following independence, the natural increase surged, driven by nearly 50% reduction in mortality. Between 2000 and 2002, there was also a very positive migration balance, totaling 125,000. Although negative migration balances were recorded again in subsequent years, the population grew rapidly. By 2023, the population of Timor-Leste exceeded 1,384,000 inhabitants.

To estimate population losses, we created a counterfactual population projection for Timor-Leste from 1971 to 2023. This projection assumes that the excess mortality observed between 1974 and 1999 – attributed to Portuguese decolonization, the civil war, and, most significantly, the Indonesian invasion and occupation – did not occur. The objective of this projection is to model a hypothetical scenario where these excess deaths were averted (Chapter 2.1).

The projection's results clearly show that these events severely disrupted Timor-Leste's population growth. Assuming the pre-crisis mortality trend had continued, the population would have exceeded actual levels from 1974 onward. By 1975, this difference would have reached nearly 7,000 individuals, exceeding 64,000 by the early 1980s. By 1999, when Indonesia withdrew, the population would have been 176,000 higher than the actual figure. This difference represents an estimate of the cumulative direct and indirect population losses resulting from the occupation (Fig. 3).

3.2 Mortality in Timor-Leste

Mortality rates in Timor-Leste had historically been high, with standardized death rates in 1950 reaching 40‰ for men and 35‰ for women (based on the WHO New Standard Population, Fig. 4). Mortality gradually declined in line with global trends, dropping by approximately one percentage point for both genders by 1973. This trend abruptly reversed in 1974 due to the collapse of Portuguese administration and violence during Timor-Leste's struggle for independence.

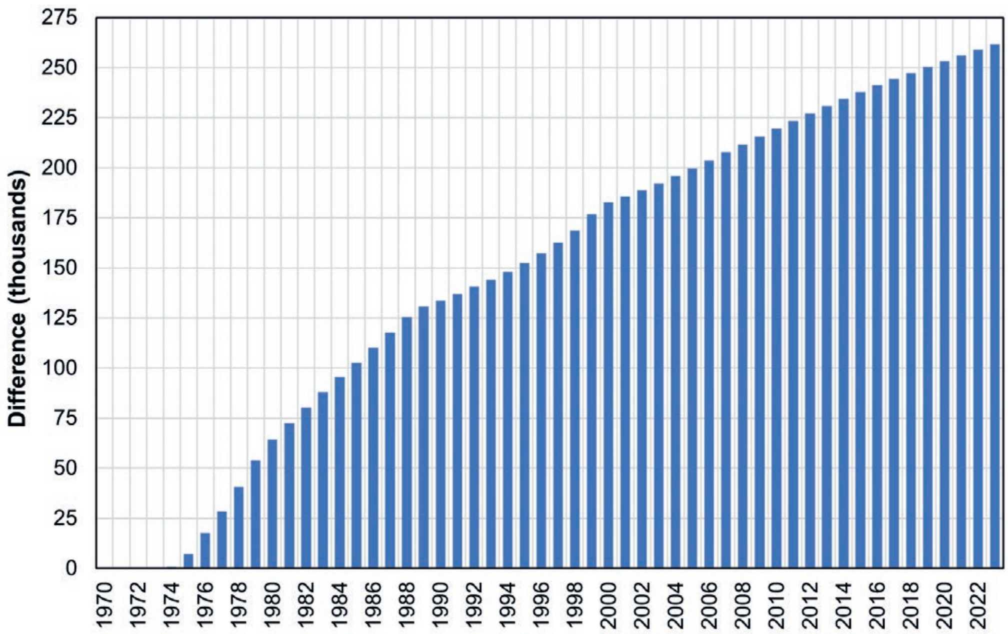

Fig. 3 Differences between projected and observed population.
Source: WPP 2024 and author's calculation.

Mortality rose dramatically in 1975 as civil war erupted and Indonesia invaded. The standardized mortality rate for males spiked to over 59‰, while for females it rose to nearly 35‰, exceeding 1951 levels. This excess male mortality caused by the conflict persisted until 1999. Gender disparities were most significant between 1980 and 1988. While female mortality returned to pre-crisis levels by 1984, male mortality remained elevated due to ongoing violence and persecution by the Indonesian army. By 1988, male mortality remained higher than pre-crisis levels by almost 42%, highlighting the conflict's selective nature of excess mortality, particularly among younger and adult men (Fig. 5).

Age-specific death rates provide further insights. From 1972 to 1978, mortality among boys aged 5–14 more than doubled, though it subsequently declined to 1972 levels by 1988. However, a similar decline occurred only in the 0–4 age group. Particularly, young and adult men aged 15–54 experienced a sustained increase in mortality. The death rate among men aged 15–39 tripled between 1972 and 1978 and remained

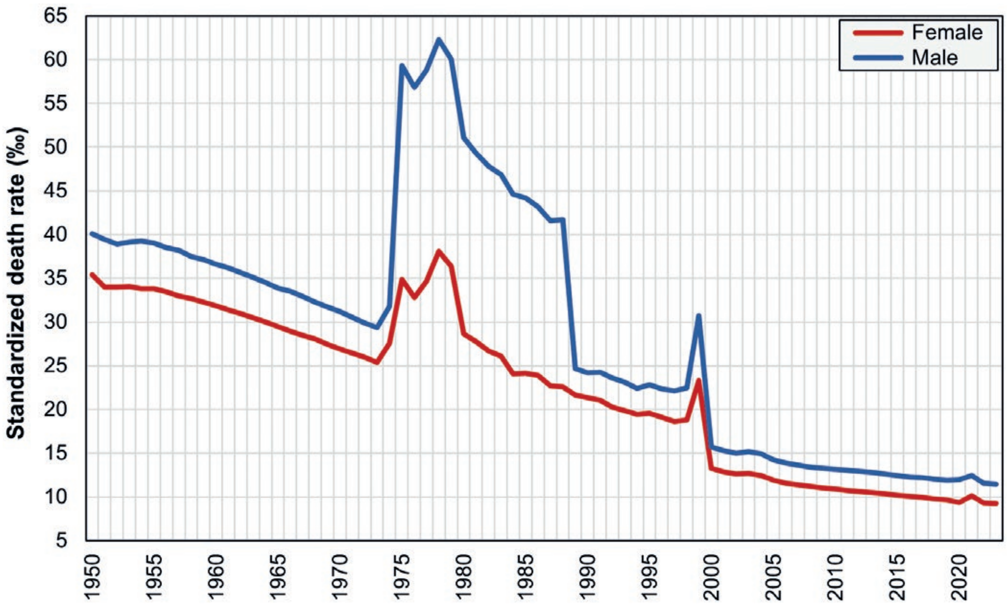

Fig. 4 Standardized death rate in Timor-Leste, 1950–2023.
Source: WPP 2024 and author's calculation.

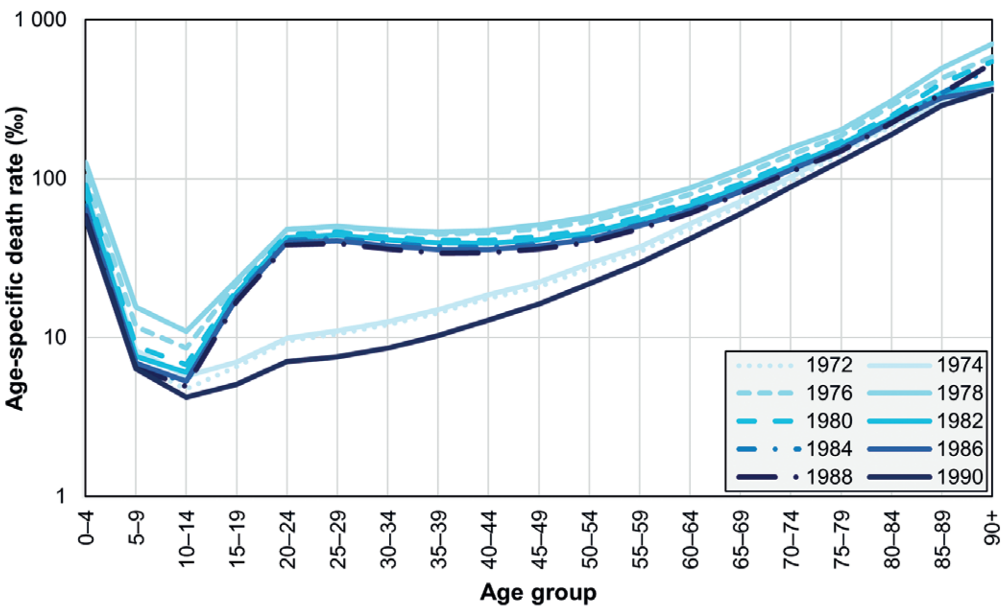

Fig. 5 Age-specific death rate in Timor-Leste, Males, 1972–1990.
Note: The Y-axis is on a logarithmic scale.
Source: WPP 2024 and author's calculation.

elevated until 1988, after which it began to decline significantly. While mortality during 1978 affected male age groups more evenly, subsequent years witnessed a concentration of excess mortality among men aged 15–39. By 1990, mortality declined across all age groups, falling below 1972 levels (Fig. 5).

There was a sharp increase of female mortality in 1976 and 1978, surpassing early 1950s levels. However, female mortality was less severe and less age-selective than male. For instance, mortality among men aged 20–24 increased fivefold between 1972 and 1978, whereas the highest rise among women occurred in the 5–9 age group, rising from 7‰ to 16‰ (Fig. 6). By 1984, the death rate among women had returned to pre-crisis levels, except for the 40–49 and 90+ age groups. Due to the mortality of very young girls, we believe women were more likely to be affected by the indirect consequences of war, such as malnutrition, rather than by direct violence.

Infant (between birth and exact age 1) and child (between birth and exact age 5) mortality rates also reflect the conflict's impact. From extraordinarily high 1950s levels – over 270‰ for infants and 400 ‰ for children – these rates declined significantly by 1973 (Fig. 7). However, the Indonesian invasion reversed this trend. By 1975, infant mortality rose to 244‰

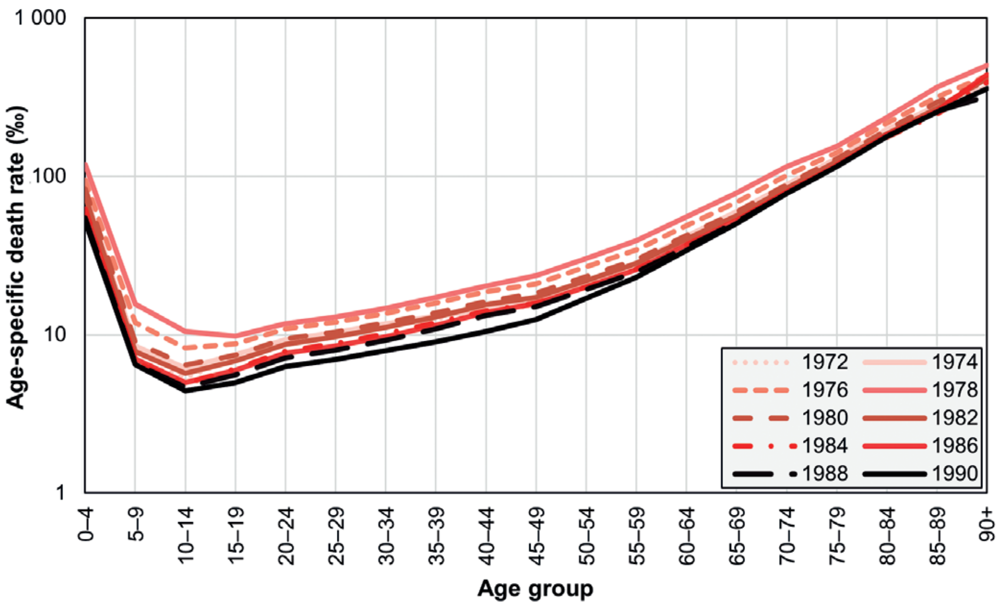

Fig. 6 Age-specific death rate in Timor-Leste, Females, 1972–1990.
Note: The Y-axis is on a logarithmic scale.
Source: WPP 2024 and author's calculation.

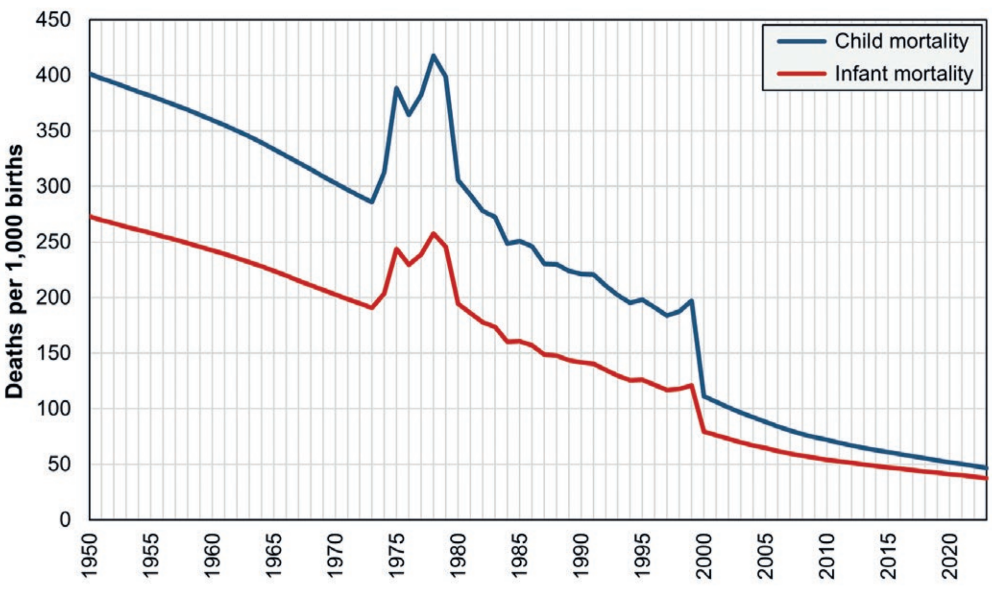

Fig. 7 Child and infant mortality in Timor-Leste, 1950–2024.
Source: WPP 2024.

and child mortality to 388‰, peaking in 1978 at 258‰ and 418‰, respectively (WPP 2024). Conditions began improving in the early 1980s, with both infant and child mortality falling below pre-crisis levels by 1982. During the conflict, the share of infant mortality within child mortality decreased to 62% (compared to a pre-crisis average of 67%), indicating a worsening situation for children aged 1–4. Mortality spiked briefly in 1999 amid the chaos of Indonesia's withdrawal but quickly declined. Since 2000, infant mortality has steadily declined, now accounting for over 80% of child mortality. However, child mortality remains high in global terms at 47‰ (WPP 2024).

Life expectancy at birth was initially low, at 27.8 years for men and 30.6 years for women in 1950, increasing to 37.6 and 40.5 years by 1973. These gains were reversed during the occupation (Fig. 8). In 1978, life expectancy plummeted to 19.4 years for men and 28.3 years for women (WPP 2024). High child mortality and excess male mortality, especially among young and adult men, contributed to this decline. Between 1980 and 1988, male life expectancy at birth exceeded that at age 15, unprecedented in Timor-Leste, driven by high mortality among young men and declining child mortality. Male life expectancy returned to pre-crisis levels by 1989, while females

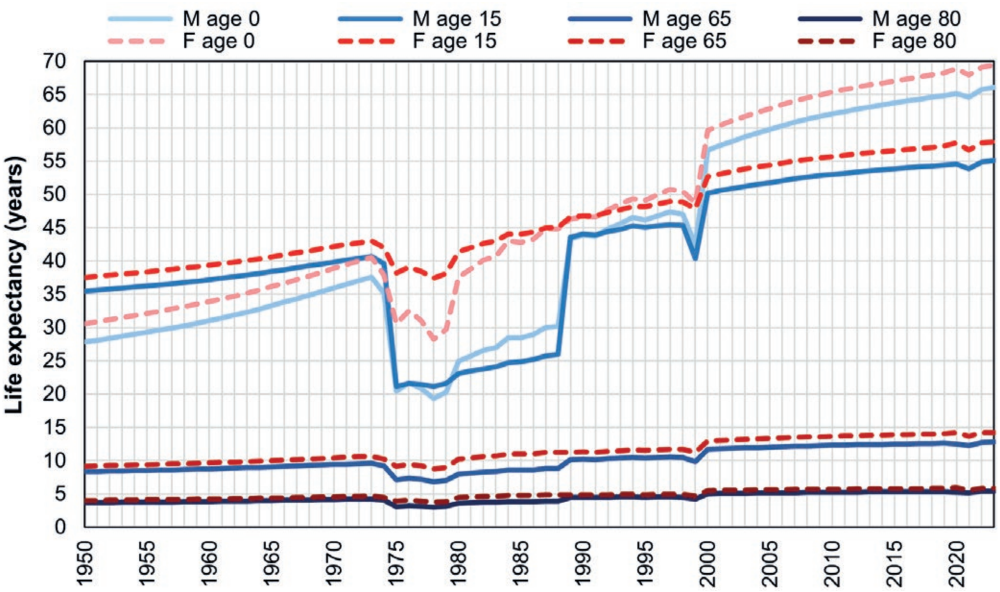

Fig. 8 Life expectancy at age 0, 15, 65 and 80 in Timor-Leste, females and males, 1950–2024.
Source: WPP 2024.

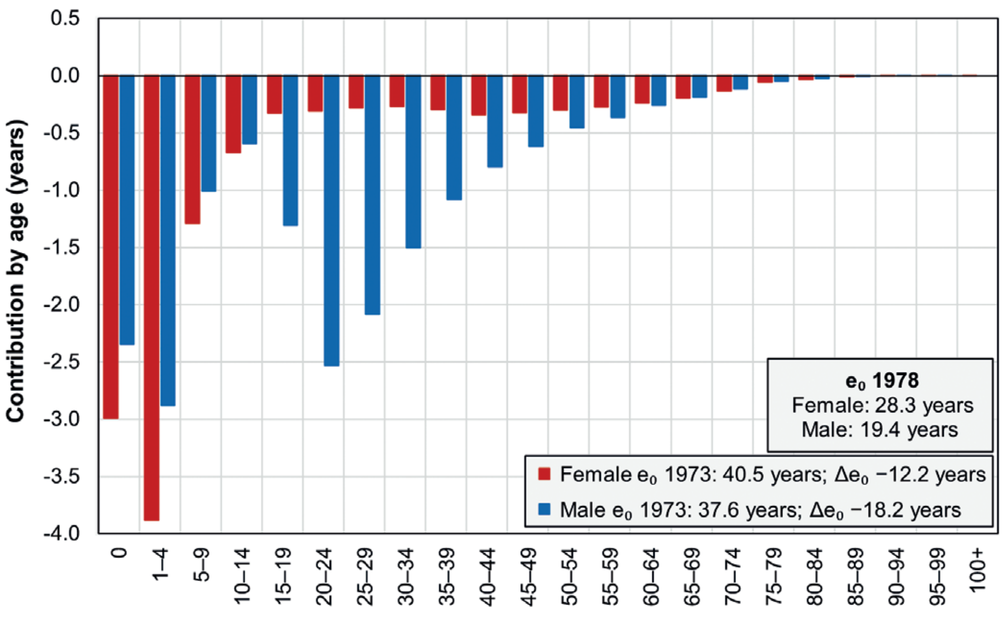

Fig. 9 Age-decomposition of life expectancy change in Timor-Leste, 1973–1978.
Source: WPP 2024 and author's calculation.

reached this level by 1984. Life expectancy generally increased thereafter, except in 1999, when mortality temporarily rose amid the chaos of Indonesia's withdrawal. Another decline occurred in 2021 due to the COVID-19 pandemic.

The age-decomposition of change in life expectancy reveals the selective impact of the conflict. Between 1973 and 1978, spanning the pre-crisis period and the peak of mortality, life expectancy at birth dropped dramatically. Decomposition analysis shows that all age groups, except men aged 100+, contributed to this decline. The most significant contributions for both sexes were observed in children, particularly those aged 1–4. Excess child mortality was more pronounced among females, with the 0–4 age group contributing −6.9 years, accounting for more than half of the total decline in life expectancy, while for males, the contribution was −5.2 years. However, a key sex difference lies in the contributions of the 15–44 age groups, which were more than five times higher for males. These groups accounted for more than half of the total decline in life expectancy at birth, which was 18.2 years. This confirms that excess mortality among young and adult men during the occupation was exceptionally high.

3.3 Fertility in Timor-Leste

Historically, the country has been characterized by high fertility. In 1950, the TFR was 6.6 children per woman, gradually declining to 5.1 by 1977 (Fig. 10). The lowest fertility level coincided with the Indonesian invasion, indicating its potential negative impact on fertility. However, the most significant interannual declines occurred earlier, between 1972 and 1974, with decreases of 0.1 children per woman.

Fertility rebounded after 1978, likely attributable to a decline in mortality in 1980 (Chapter 3.2). The most substantial increase occurred between 1980 and 1981, with the TFR rising by 0.31 children per woman (WPP 2024). This surge was likely driven by compensatory fertility, aiming to compensate for previous child losses or delayed births. Following this sharp increase, fertility continued to rise, albeit at a slower pace. Notably, the mean age of childbearing remained stable at 31 years from 1950 to 2001, suggesting consistent reproductive behaviors throughout this period (WPP 2024). Since 2002, this indicator has been gradually declining with slight fluctuations, reaching a value of 29.4 years in 2023.

Fertility rose slightly in the late 1980s, with the most significant increase between 1988 and 1990, likely linked to a decline in male mortality. The TFR grew by 0.05 children per woman annually during this period, a significantly higher rate than in previous years. However, the TFR has experienced a sharp decline since 2000, decreasing by 3.23 children per woman. By 2023, it had reached 2.71, remaining above the global average of 2.32 (WPP 2024).

Age-specific fertility rates mirrored these patterns. The fertility rebound mentioned earlier, peaking in 1999, was primarily attributable to an elevated fertility rate among women aged 30–34 (Fig. 11). It is important to emphasize that fertility data from the mentioned age-specific rates do not align with the TFR values in WPP 2024 for this period. Our estimate for 1999 based on age-specific fertility rates is 6.9 children per woman, 0.9 higher than the WPP estimate

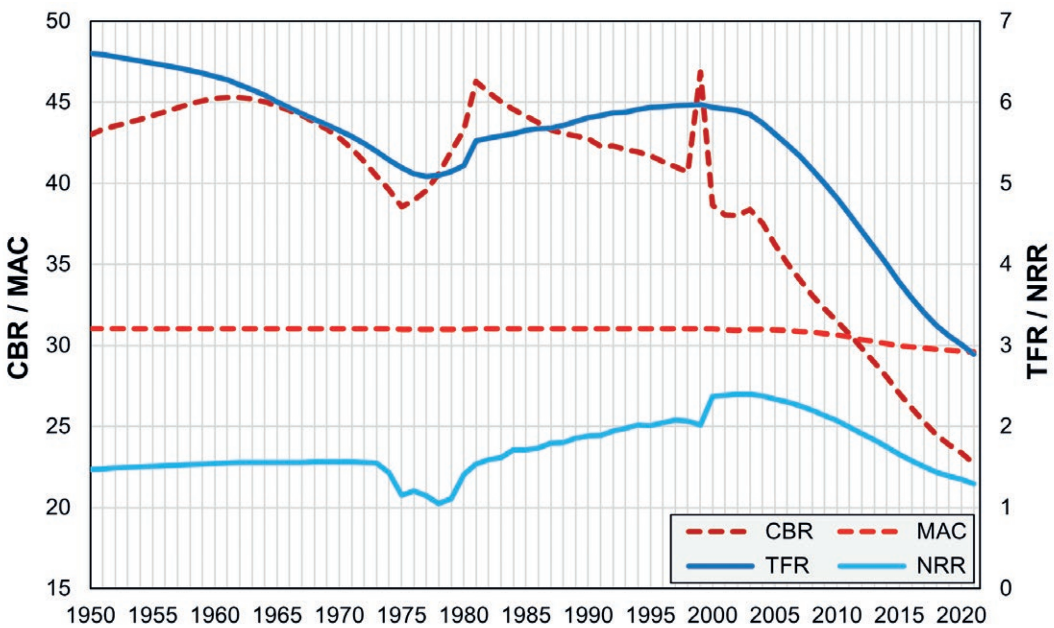

Fig. 10 Trends in fertility indicators in Timor-Leste, 1950–2023.
Note: CBR – Crude birth rate, MAC – Mean age at childbearing, TFR – Total fertility rate, NRR – Net reproduction rate.
Source: WPP 2024.

(Chapter 2.3). This discrepancy likely stems from significant population outflow that year, along with the smoothing of fertility trends in the WPP data. The differences are most pronounced between 1999 and 2002 but become negligible afterward. Subsequently, fertility has declined across all age groups. While the 25–29 age group consistently exhibited the highest fertility rate, its share has steadily increased (Fig. 11). Notably, the second-highest fertility group shifted from 30–34 to 20–24 years, reflecting a decrease in the mean age of childbearing. This leads to a shift toward a fertility regime characterized by decreasing fertility rates and a reduced role of older age groups in reproductive activity.

DHS data (2003; 2024) show a decline in completed fertility, from 5.8 to 5.0 children per woman, between 2009–2010 and 2016. However, the ideal number of children has decreased even more, from 5.7 (2003) to 3.7 (2016), suggesting an unmet need for family planning. Public distrust to such programs may stem from Indonesia's forced contraception policy during the occupation (Saikia et al. 2009).

3.4 Population structure of Timor-Leste

Timor-Leste's population has been historically very young. The 65+ age group remained very low, from 3.1% in 1950 to 2.2% in 1991. A significant increase only began in the early 2000s, peaking at 5.6% in 2019 (Fig. 12).

The 0–14 age group made up nearly 44% of the population in 1950 and decreased to 39% by 1965. A slight increase to a 1973 peak coincided with high, particularly adult, mortality, despite declining fertility. At the beginning of the 1970s, there was a gradual increase in the share. However, this was disrupted in

Tab. 1 Fertility and family planning indicators, Timor-Leste, 2003–2016.

DHS survey	Completed fertility of women aged 40–49	Mean ideal number of children for all women
2003	–	5.7
2009–2010	5.8	5.0
2016	5.0	3.7

Note: Data not available for entries marked with –.
Source: DHS 2003 and DHS 2024.

1974, after which the share began to decline, falling to 37.4% in 1980. This was likely influenced by the Indonesian occupation, as mortality in this age group increased significantly, while declining fertility rates led to fewer births.

Since the 1980s, the child population share rapidly increased due to rising fertility, reaching 43% in 1996. This trend was disrupted primarily by the increasing elderly population (Fig. 12). Since 2009, the child population share's decline has accelerated, while the working-age population increased (WPP 2024). It should be noted that the shares of the child and working-age population were essentially inversely related during the observed period, due to the very low share of the elderly.

Given the findings on the presence of excess male mortality during the occupation period, it is certainly relevant to analyze the sex ratio. It showed a unique trend during the Indonesian occupation. The sex ratio gradually declined until 1974, reaching a value of 101.3 men per 100 women. Following the onset of the occupation, a sharp decline occurred, stabilizing only in 1980 at 93.6. This phenomenon largely reflected the mortality situation, characterized by high male

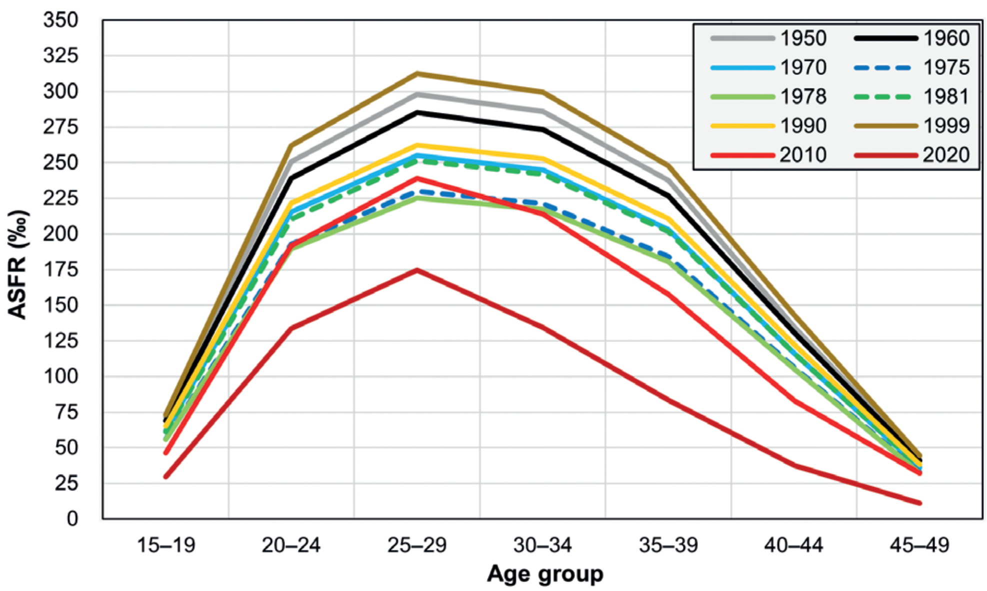

Fig. 11 Age-specific fertility rate in Timor-Leste, 1950–2020.
Source: WPP 2024 and author's calculation.

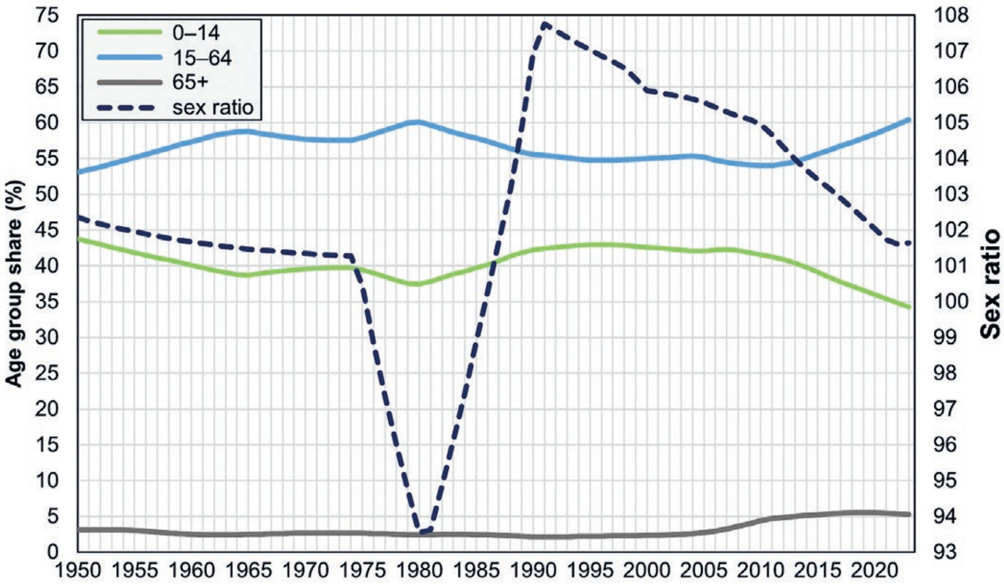

Fig. 12 Broad age groups share and sex ratio in Timor-Leste, 1950–2023.
Source: WPP 2024 and author's calculation.

excess mortality, which began to gradually decrease towards the end of the 1970s.

In the 1980s, according to WPP data (2024), the sex ratio saw a sharp increase, significantly exceeding pre-crisis levels. In 1991, at the peak, the sex ratio was 107.8 men per 100 women. However, it is highly unlikely that this increase can be solely attributed to a decline in male excess mortality. According to WPP data (2024), there was no change in the sex ratio at birth. Assuming the sex ratio estimates in the WPP database (2024) are reasonably accurate, it is likely that this phenomenon was driven by migration. Given that net migration was negative during the 1980s, the sharp increase in the sex ratio would have been primarily caused by male immigration and female emigration. However, data on the structure of migration are not available.

There was a continuous decline in the sex ratio, with the most significant drop occurring at the turn of the millennium, during the withdrawal of Indonesian troops and the gaining of independence, and again after 2010. In 2023, the sex ratio was 101.6 men per 100 women.

The median age was 17.1 years in 1950, rising to 19.4 in 1974 (WPP 2024). After the onset of the occupation, the median age increased to 20 years by 1979, likely due to significant excess mortality among children and young people (Tab. 2). This was followed by a period of gradual decline, which halted in 1999, before a renewed increase, reaching 21 years by 2023.

Tab. 2 Age structure indicators, Timor-Leste, 1950–2020.

Age group		1950	1960	1970	1980	1990	2000	2010	2020
Age group share of the total population (%)	0–14	43.8	40.1	39.6	37.4	42.2	42.6	41.6	36.0
	15–64	53.1	57.4	57.7	60.1	55.6	55.0	54.0	58.5
	65+	3.1	2.5	2.7	2.5	2.2	2.4	4.4	5.5
Age group share of the male population (%)	0–14	43.8	40.1	39.8	39.4	42.8	42.6	42.0	36.3
	15–64	53.5	57.8	58.0	58.4	55.2	55.2	53.8	58.6
	65+	2.7	2.1	2.2	2.2	2.0	2.2	4.2	5.1
Age group share of the female population (%)	0–14	43.8	40.1	39.4	35.6	41.6	42.6	41.1	35.7
	15–64	52.6	56.9	57.4	61.7	56.1	54.9	54.2	58.3
	65+	3.6	3.0	3.2	2.7	2.3	2.5	4.7	6.0
Indicators of population aging	Median age	17.1	18.2	19.5	20.0	18.0	17.5	17.8	20.1
	YADR	82.5	69.9	68.6	62.3	75.9	77.4	77.0	61.6
	OADR	5.9	4.4	4.7	4.1	3.9	4.3	8.2	9.5

Note: YADR – the number of children (aged 0–14) per 100 working-age people (15–64); OADR – the number of elderly (65+) per 100 working-age people (15–64).
Source: WPP 2024 and author's calculation.

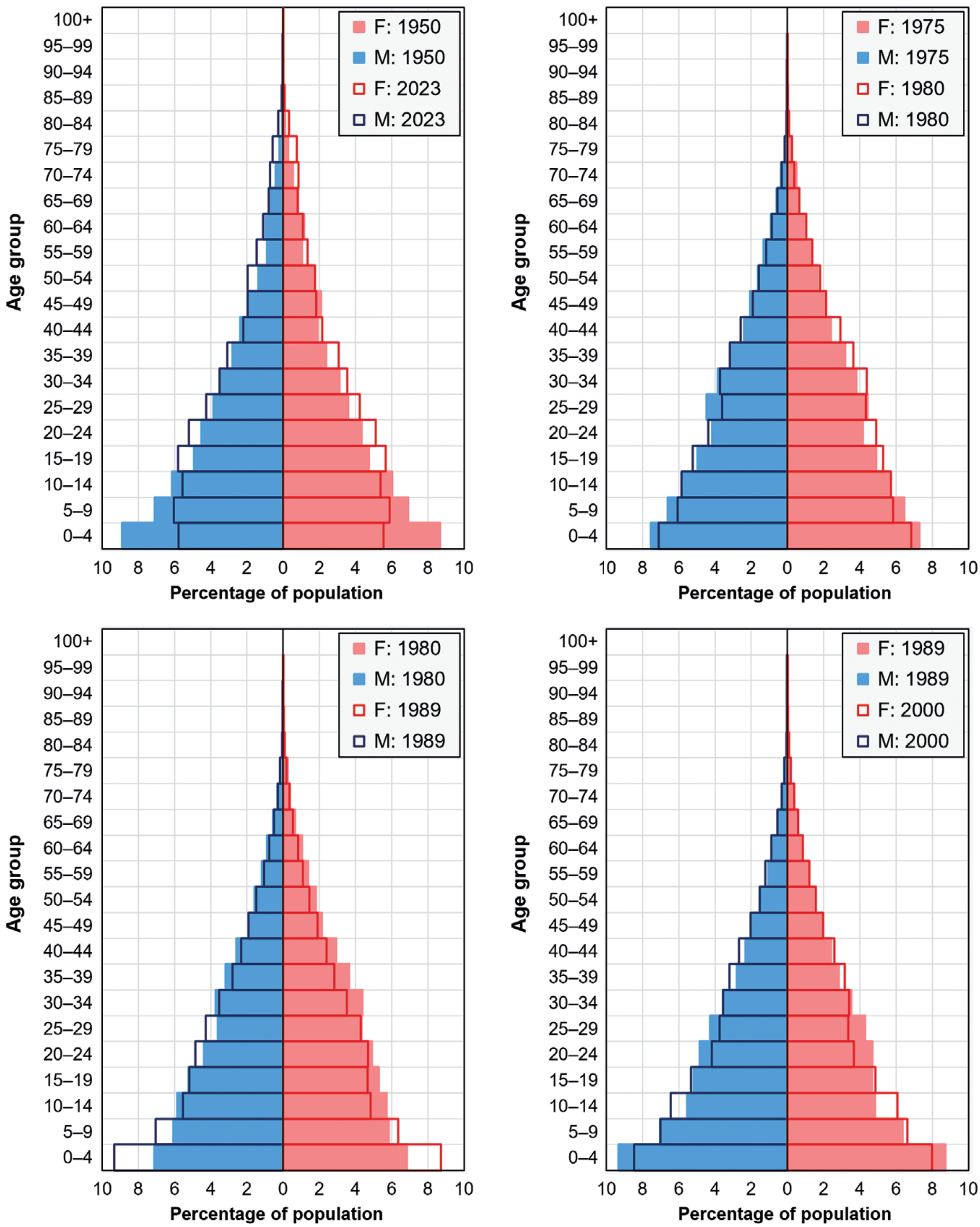

Fig. 13 Age-sex structure – Timor-Leste, 1950–2023, 1975–1980, 1980–1989 and 1989–2000.
Note: F – Females; M – Males. Source: WPP 2024.

Although demographic aging is not an immediate concern, the old-age dependency ratio is rising, while the young-age dependency ratio is declining, both driven by falling fertility and mortality rates. Since 2010, the working-age population has grown, signaling the onset of the demographic window. However, to harness such economic potential, well-designed socio-economic policies are essential (Kotschy et al. 2020).

Between 1950 and 2023, Timor-Leste's population structure became more stationary, with a declining proportion of children and a growing share of older adults, particularly men aged 70 and above (Fig. 13).

A comparison of age and gender between 1975 and 1980, the period of highest mortality, shows that declining fertility and rising mortality led to a decrease in the youngest age groups (Fig. 13). The share of the 0–9 age group declined for both sexes, while men also saw a significant drop in the proportion of those aged 25–34. Despite these trends, the population grew by 37,000 women, compared to just 14,000 men during this period (WPP 2024).

Between 1980 and 1989, the 0–9 age group grew due to higher fertility and declining mortality (Fig. 13). In contrast, the 10–14 age group saw a significant decline, likely reflecting lower fertility in the late 1970s. The male population in the 15–24 age group grew nearly four times more than the female population in the same age group (WPP 2024). At the same time, the share of those aged 30–44 declined considerably, particularly among women. The gradual reduction in male excess mortality during this period is also reflected in overall population growth, with the male population increasing by 67,200, while the female population grew by only 27,700.

Between 1989 and 2000, during the period between the significant decline in mortality and the end of the occupation, mortality rates fell, and the child population grew, particularly among women, despite a declining share of individuals in the youngest age group (Fig. 13). During this period, the male population grew by 1,900, while the female population declined by 1,600. This was likely due to a major wave of emigration in 1999, when net migration reached a negative balance of 263,000. Both men and women saw a significant decline in the 20–29 age group.

4. Discussion

The analysis of various aspects of Timor-Leste's population development has shown that the Indonesian occupation had a clear impact on both individual demographic processes and overall population trends.

The counterfactual demographic projection presented in this paper indicates that due to decolonization, civil war, and, most notably, the subsequent Indonesian occupation, Timor-Leste's population in 1989 was 130,000 lower than it would have been without disruptions to mortality trends. This finding broadly aligns with Kiernan (2008), who estimates 145,000 excess deaths. The UN estimated 103,000 deaths between 1974 and 1999, while some sources suggest the toll exceeded 200,000 (Silva and Ball 2008). By 1999, the counterfactual projection suggests the population of Timor-Leste was 176,000 lower than its estimated value, representing about 30% of the pre-crisis population. In comparison, the relative population loss during the Cambodian genocide was lower, at approximately 26% of its pre-crisis population, with an absolute death toll of around 1.9 million (Heuveline 2015). Both events occurred in a similar period and region, although it is important to note that the Cambodian genocide lasted only a few years.

Given that the highest values of net migration, whether positive or negative, were recorded at the beginning of the invasion and at its end, it is evident that the Indonesian occupation had a significant impact on migration patterns. Kugler (2016) observes similar effects of genocide on migration. However, an interesting contrast emerges: in the cases he examined, genocides typically led to large-scale emigration at the outset, followed by an immigration wave after the crisis ended. In Timor-Leste, however, the pattern was reversed. A possible explanation is that the country's geographical isolation, combined with the occupation, effectively prevented inhabitants from leaving. It is important to note that such migration also has a substantial indirect influence on other demographic processes, particularly through changes in the age and sex structure. This applies not only to the emigrant population but also to the population of the destination areas.

The events mentioned had a profound impact on the country's mortality levels due to their very nature. The mortality analysis reveals several key findings. The invasion and subsequent occupation from 1975 to 1999 significantly influenced mortality rates. Notably, excess mortality among children, as well as young and adult men (particularly those aged 15–39) was a defining characteristic of this period. The conflict's negative impact on male mortality was more intense, selective, and prolonged compared to its effect on female mortality. Similar mortality patterns were also observed during the Cambodian genocide (Heuveline 1998). However, female mortality was also severely affected, with excess deaths most concentrated among girls aged 5–14. Following the invasion, infant mortality increased by more than 30% compared to pre-crisis levels, while child mortality rose by over 40%. The scale and selectivity of the crisis can be effectively illustrated through life expectancy at birth. In 1973, female life expectancy was 40.5 years, but by 1978, it had dropped to 28.2 years. The decline was even more pronounced for men, falling from 37.5 to just 19.3 years.

The demographic analysis indicates a significant decline in fertility in Timor-Leste. Prior to this decline, the TFR was exceptionally high, exceeding 6.5 children per woman (WPP 2024). The results of the analysis reveal a gradual decline in TFR, dropping from approximately 6.6 children per woman to 5.3 before the invasion. Fertility continued to decrease until 1978, coinciding with the period of the highest mortality rates. The subsequent increase in fertility between 1979 and 1999 may have been a response to population losses caused by the occupation. However, it is important to note that TFR estimates in WPP (2024) do not align with the figures calculated for this study based on WPP (2024) data. Furthermore, the targeted and violent Indonesianization of the country, along with the marginalization of the Timorese population, likely triggered resistance and reinforced strong pronatalist intentions among the local population (Saikia et al. 2009).

The Indonesian occupation deeply altered Timor-Leste's population structure through excess male mortality, declining fertility, and forced migration. The sex ratio dropped sharply, especially by 1980, due to high male mortality. However, compared to other cases of genocides and conflicts, its subsequent recovery was unusually rapid. For example, in the case of Germany after World War II, Urlanis (1971) suggests that the sex ratio would not return to pre-war levels until the 21st century, which, according to WPP (2024), has not yet occurred. Another example is the Cambodian genocide, which also took place in the late 1970s. Neupert and Prum (2005) state that in 1980, the sex ratio was 87 males per 100 females. However, it has only reached 96 so far (WPP 2024).

The youngest age groups saw a reduction due to both increased child mortality and declining fertility, while the working-age population, especially men aged 25–34, also declined. Migration likely further disrupted the balance, with male immigration and female emigration affecting the sex ratio. However, this remains a hypothesis, as data on the age and gender structure of migrants is unavailable. As mortality declined in the 1980s, fertility rebounded, briefly increasing the child population. By the time of independence, Timor-Leste's population had shifted toward a more stationary model, but the demographic scars of the occupation remain evident today, such as the lower numbers of persons born between 1975 and 1980.

Given that one of the greatest challenges in demographic research on genocide is the severe limitation of high-quality data sources, it is essential for future research to focus on expanding methodological approaches.

A more detailed analysis of the impact of genocide-induced migration on population dynamics would also be highly valuable. Additionally, further research should address the long-term effects of genocide on affected populations, such as changes in morbidity patterns or shifts in causes of death. Assessing the relationship between genocide and mortality in relation to socioeconomic status could also be beneficial.

5. Conclusions

This paper introduces the demography of genocide as a specialized branch of demographic research. To illustrate its importance, the main section presents a case study on the demographic impacts of the Indonesian occupation of Timor-Leste.

The demography of genocide can be described as a branch of research that focuses on the immediate and long-term impacts on mortality, fertility, and migration among both directly and indirectly affected populations. While it primarily employs demographic methods, it also incorporates interdisciplinary approaches from history, sociology, and genocide studies to aid in genocide prevention.

However, it is important to note that the demography of genocide still lacks institutional recognition: it is not formally taught at universities, nor is it the focus of dedicated research institutions. Currently, it remains largely a subfield within the demography of armed conflict and violence. An equally significant challenge is the lack of methodological frameworks, whether for assessing impacts or improving the efficiency of high-quality data collection. The scarcity of such data, due to the violent and destructive nature of genocides and related conflicts, represents a major obstacle to research in this field.

The case study clearly demonstrates the profound demographic impact of the genocide perpetrated by Indonesia against the population of Timor-Leste between 1975 and 1999.

There was a significant deterioration in mortality across the entire population. This overall rise in mortality was disproportionately severe, leading to particularly high excess mortality among children and young men.

While the effect on fertility was less significant and largely aligned with pre-existing long-term trends, the genocide nevertheless fundamentally altered the country's population dynamics. The relatively rapid rebound of fertility could indicate that a population's response to genocidal violence is not merely passive but may also be an active process of resilience. This demographic shock, combined with notable migratory shifts at the beginning and end of the occupation, significantly reshaped the age and sex structure of Timor-Leste's population. The pattern of migration, where population movement was likely constrained by the country's geographical isolation, suggests that migratory responses to mass violence may not be uniform and should be understood within their specific geopolitical and geographical contexts.

Based on the projection results presented in this paper, the total losses by the end of the Indonesian occupation in 1999 amounted to a staggering toll, representing nearly a third of the country's pre-crisis population. This methodology could be valuable for evaluating the demographic consequences in contexts where data are scarce, a situation that is common in genocide research.

Overall, the demographic scars of the Indonesian occupation remain visible today, highlighting the lasting consequences of acts of mass violence and demonstrating how such events can fundamentally transform a nation's population structure for generations.

Acknowledgments

I would like to thank my doctoral supervisor, Assoc. Prof. Klára Hulíková Tesárková, for her valuable guidance throughout the preparation of this paper. I would also like to thank the anonymous reviewers for their constructive comments, which significantly improved the final manuscript.

References

Andreev, E. M., Shkolnikov, V. M., Begun, A. (2002): Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. Demographic Research 4(7), 499–522, https://doi.org/10.4054/DemRes.2002.7.14.

Avenir Health (2024): Spectrum Manual, Avenir Health. Available online: https://avenirhealth.org/Download/Spectrum/Manuals/SpectrumManualE.pdf (accessed on 12 January 2025).

Ball, P., Spirer, L., Spirer, H. F. (2000): Making the Case: Investigating Large Scale Human Rights Violations Using Information Systems and Data Analysis. Washington, DC: American Association for the Advancement of Science.

Bloxham, D., Dirk Moses, A. (2010): Editor's Introduction: Changing Themes in the Study of Genocide. In Bloxham, D., Moses, A. D., The Oxford Handbook of Genocide Studies. Oxford: Oxford University Press, 1–18, https://doi.org/10.1093/oxfordhb/9780199232116.001.0001.

BPS (1991): Penduduk Indonesia Hasil Sensus Penduduk 1990 Seri L1, Badan Pusat Statistik. Available online: https://www.bps.go.id/id/publication/1991/01/28/bb2dc5b6c70e83f3a00aa495/penduduk-indonesia-hasil-sensus-penduduk-1990-seri-l1.html (accessed on 12 January 2025).

Brunborg, H., Lyngstad, T. H., Urdal, H. (2003): Accounting for Genocide: How Many Were Killed in Srebrenica? European Journal of Population 19(3), 229–248, https://doi.org/10.1023/A:1024949307841.

Brunborg, H., Tabeau, E. (2005): Demography of Conflict and Violence: An Emerging Field. European Journal of Population 21(2–3), 131–144, https://doi.org/10.1007/s10680-005-6850-7.

Budiardjo, C., Liong, L. S. (1984): The War Against East Timor. London: Zed Books.

Cribb, R. (2001): How many deaths? Problems in the statistics of massacre in Indonesia (1965–1966) and East Timor (1975–1980). In Wessel, I., Wimhöfer, G., Violence in Indonesia. Hamburg: Abera Verlag, 82–98.

Dasvarma, G. L. (2011): Demography of Timor-Leste. A Brief Overview and Some Implications for Development and Security. In Harris, V., Goldsmith, A. Security, Development and Nation-Building in Timor-Leste: A cross-sectoral assessment. London: Routledge, 195–209, https://doi.org/10.4324/9780203828199.

DHS (2003): Timor-Leste 2003 DHS. The Demographic and Health Surveys Program, Ministry of Health and National Statistics Office, The Australian National University, ACIL Australia Pty Ltd., University of Newcastle. Available online: https://www.aidsdatahub.org/sites/default/files/resource/timor-leste-dhs-2003.pdf (accessed on 16 December 2024).

DHS (2024): The DHS Program STATcompiler. Timor-Leste: 2009–10 and 2016 DHS Survey Data, The Demographic and Health Surveys Program. Available online: https://www.statcompiler.com (accessed on 17 December 2024).

Dunn, J. (2012): Genocide in East Timor. In Totten, S., Parsons, W. S., Centuries of Genocide Essays and Eyewitness Accounts. London: Routledge, 279–315, https://doi.org/10.4324/9780203867815-9.

Ea, M. T. (1981): Kampuchea: A Country Adrift. Population and Development Review 7(2), 209–228, https://doi.org/10.2307/1972621.

Hauser, P. M. (1942): Population and Vital Phenomena. American Journal of Sociology 48(3), 309–322, http://dx.doi.org/10.1086/219178.

Heuveline, P. (1998): 'Between one and three million': towards the demographic reconstruction of a decade of Cambodian history (1970–79). Population Studies 52(1), 49–65, https://doi.org/10.1080/0032472031000150176.

Heuveline, P. (2015): The boundaries of genocide: Quantifying the uncertainty of the death toll during the Pol Pot regime in Cambodia (1975–79). Population Studies 69(2), 201–218, https://doi.org/10.1080/00324728.2015.1045546.

Hilberg, R. (2003): The Destruction of the European Jews. 3rd ed., New Haven: Yale University Press.

Hull, T. H. (2003): From province to nation: the demographic revolution of a people. In Fox, J. J., Soares, D. B., Out of the Ashes: Destruction and Reconstruction of East Timor. Canberra: ANU E Press, 28–40, http://doi.org/10.22459/OA.11.2003.02.

INETL (2023a): 2015 Census Wall Chart, Instituto Nacional De Estatística Timor-Leste. Available online: https://inetl-ip.gov.tl/2023/03/13/2015-census-wall-chart/ (accessed on 12 January 2025).

INETL (2023b): 2022 Census Wall Chart, Instituto Nacional De Estatística Timor-Leste. Available online: https://inetl-ip.gov.tl/2023/10/04/2022-census-wall-chart (accessed on 12 January 2025).

Johnson, R. E. (1992): Introduction. Russian Studies in History 31(1), 3–9, https://doi.org/10.2753/RSH1061-198331013.

Jones, A. (2017): Genocide: A Comprehensive Introduction. 3rd ed., London: Routledge.

Joseph, A., Hamaguchi, T. (2014): Timor-Leste: The History and Development of Asia's Newest Nation. Lanham: Lexington Books.

Kiernan, B. (2008): Genocide and Resistance in Southeast Asia: Documentation, Denial, and Justice in Cambodia and East Timor. London: Routledg.

Kotschy, R., Suarez Urtaza, P., Sunde, U. (2020): The demographic dividend is more than an education dividend. PNAS 117(42), 25982–25984, https://doi.org/10.1073/pnas.2012286117.

Kugler, T. (2016): The Demography of Genocide. In Anderton, C. H., Brauer, J., Economic Aspects of Genocide,

Mass Killing, and Their Prevention. Oxford: Oxford University Press, 102–124, https://doi.org/10.1093/acprof:oso/9780199378296.003.0004.
Lemkin, R. (1944): Axis Rule in Occupied Europe. Laws of Occupation. Analysis of Government. Proposals for Redress. Washington, DC: Carnegie Endowment for International Peace.
Neupert, T., Prum, V. (2005): Cambodia: Reconstructing the Demographic Stab of the Past and Forecasting the Demographic Scar of the Future. European Journal of Population 21(2–3), 217–246, https://doi.org/10.1007/s10680-005-6854-3.
Niewyk, D., Nicosia, F. (2000): The Columbia Guide to the Holocaust. New York: Columbia University Press, https://doi.org/10.7312/niew11200.
Poliakov, I. A., Zhiromskaia, V. B., Kiselev, I. N. (1992): Part I: The 1937 Census. Russian Studies in History 31(1), 10–39, https://doi.org/10.2753/RSH1061-1983310110.
Preston, S. H., Heuveline, P., Guillot, M. (2001): Demography: Measuring and Modeling Population Processes. Malden: Blackwell Publishing.
Robinson, G. (2023): Half a Century of Genocide and Extermination Indonesia, 1965–1966; East Timor, 1975–1999; and West Papua, 1963–2020. In Kiernan, B., Lower, W., Naimark, N., Straus, S., The Cambridge World History of Genocide, Volume III. Genocide in the Contemporary Era, 1914–2020. Cambridge: Cambridge University Press, 450–475, https://doi.org/10.1017/9781108767118.020.
Saikia, U. S., Dasvarma, G. L., Wells-Brown, T. (2009): The world's highest fertility in Asia's newest nation: an investigation into reproductive behaviour of women in Timor-Leste, International Union for the Scientific Study of Population. Available online: https://ipc2009.popconf.org/papers/90513 (accessed on 28 January 2025).
Schaller, D. J. (2011): From Lemkin to Clooney: The Development and State of Genocide Studies. Genocide Studies and Prevention 6(3), 245–256, https://doi.org/10.3138/gsp.6.3.245.
Seltzer, W. (1998): Population Statistics, the Holocaust, and the Nuremberg Trials. Population and Development Review 24(3), 511–552, https://doi.org/10.2307/2808153.
Shkolnikov, V. M., Andreev, E. M. (2010): Age-decomposition of a difference between two populations for any life-table quantity in Excel, Max Planck Institute for Demographic Research. Available online: https://dx.doi.org/10.4054/MPIDR-TR-2010-002 (accessed on 20 December 2024).
Silva, R., Ball, P. (2008): The Demography of Conflict-Related Mortality in Timor-Leste (1974–1999): Reflections on Empirical Quantitative Measurement of Civilian Killings, Disappearances, and Famine-Related Deaths. In Aasher, J., Banks, D., Scheuren, F. J., Statistical Methods for Human Rights. New York: Springer, 117–139, https://doi.org/10.1007/978-0-387-72837-7_6.
Thu, P. M. (2019): Internal displacement in Timor-Leste. In McWilliam, A., Leach, M., Routledge Handbook of Contemporary Timor-Leste. London: Routledge, 174–184, http://doi.org/10.4324/9781315623177-13.
UN (1948): Convention on the Prevention and Punishment of the Crime of Genocide, United Nations. Available online: https://www.un.org/en/genocideprevention/documents/atrocity-crimes/Doc.1_Convention%20on%20the%20Prevention%20and%20Punishment%20of%20the%20Crime%20of%20Genocide.pdf (accessed on 18 September 2025).
Urlanis, B. C. (1971): Wars and Population. Moscow: Progress Publishers.
Vincent, P. (1947): Guerre et population. Population (French Edition) 2(1), 9–30, http://dx.doi.org/10.2307/1523845.
Von Joeden-Forgey, E. (2010): Gender and Genocide. In Bloxham, D., Dirk Moses, A., The Oxford Handbook of Genocide Studies. Oxford: Oxford University Press, 61–80, https://doi.org/10.1093/oxfordhb/9780199232116.013.0004.
Weiss-Wendt, A. (2008): Problems in Comparative Genocide Scholarship. In Stone, D., The Historiography of Genocide. New York: Palgrave Macmillan, 42–70, http://dx.doi.org/10.1057/9780230297784_3.
WHO (2001): Age Standardization of Rates: A New WHO Standard, World Health Organization. Available online: https://cdn.who.int/media/docs/default-source/gho-documents/global-health-estimates/gpe_discussion_paper_series_paper31_2001_age_standardization_rates.pdf (accessed on 16 December 2024).
WPP (2024): World Population Prospects 2024. United Nations, Department of Economic and Social Affairs, Population Division. Available online: https://population.un.org/wpp/ (accessed on 15 December 2024).
Yehuda, D., Magos, G. (1983): The Demographic Development of Hungarian Jewry. Jewish Social Studies 45(3–4), 189–216.

Multiple births in an era of assisted reproduction and delayed motherhood: Evidence from Czechia and an international comparison

Anna Šťastná*, Boris Burcin

Charles University, Faculty of Science, Department of Demography and Geodemography, Czechia
* Corresponding author: stastna@natur.cuni.cz

ABSTRACT

This article examines recent international trends regarding multiple births, with a specific focus on Czechia. It stresses the role of assisted reproductive technology (ART) along with the fertility postponement process. Employing internationally harmonised datasets and detailed Czech data on deliveries and live births, we firstly provide a comparative overview of long-term trends concerning twinning/multiple birth rates prior to and following the widespread use of ART. Based on anonymised individual-level data for Czechia, we explore the role of ART and the maternal age in determining the proportion of multiple births and the structure of newborns. Particular attention is devoted to differences according to the maternal age and the mode of conception (with and without the use of ART). The results revealed the substantial international heterogeneity of twinning rates as driven primarily by differences in national ART practices rather than the overall volume of ART treatment. Czechia represents the most common pattern, i.e. a rapid increase in the twinning rate from less than 10 twin deliveries per 1,000 deliveries in the early 1990s to more than 21 twin deliveries per 1,000 in 2010, followed by a rapid decline. Regulatory changes in Czechia that have promoted single embryo transfers since 2012 have led to a marked decline in the incidence of multiple births. However, women aged 35 and older accounted for a significant share of multiple live births conceived through ART; 18% of all multiples born to women aged 35–39, and 36% of those born to women aged 40 and over, were conceived via ART in 2020.

KEYWORDS

fertility; multiple births; twinning rate; ART; Czechia

Received: 4 August 2025
Accepted: 2 October 2025
Published online: 18 November 2025

Šťastná, A., Burcin, B. (2025): Multiple births in an era of assisted reproduction and delayed motherhood: Evidence from Czechia and an international comparison. AUC Geographica 60(2), 221–232
https://doi.org/10.14712/23361980.2025.20

1. Introduction

The twinning rate, defined as the number of twin deliveries per 1,000 total deliveries (Human Multiple Births Database 2025), has almost doubled in many developed countries over the past forty years due primarily to the increasing age of mothers at birth and the increased use of medically assisted reproduction (MAR), i.e. hormonal treatment and assisted reproductive technology (Pison et al. 2015). According to Pison et al. (2015), the role of MAR is significant; in many developed countries, MAR explained a large proportion of the total change in the twinning rate that occurred between 1970 and 2005. The authors estimate that the contribution of MAR to the increase in twinning rates ranged widely – from 22% in Poland to 87% in Japan - with an estimated share of 56% in Czechia (Pison et al. 2015, pp. 640–641).

The increasing twinning and multiple birth rate represents an important public health concern in terms of the association of multiple pregnancies with maternal and neonatal risks, e.g. the higher risk of a preterm birth and a lower birth weight, the elevated risk of a Caesarion section delivery, more complications at birth than for single births, and the more frequent occurrence of prematurity and the associated occurrence of many long-term health problems, higher stillbirth and infant mortality rates, etc. (Fait et al. 2022; Marleen et al. 2024; Whittaker et al. 2023; Santana et al. 2016; Pison et al. 2015; Blondel et al. 2006). Therefore, concerns surrounding the impacts of MAR on multiple pregnancies has prompted regulatory changes and modifications to clinical practice, including reductions in the number of embryos transferred and the promotion of an elective single-embryo transfer strategy (ESHRE Guideline Group on the Number of Embryos to Transfer et al. 2024). These measures led to the stabilisation or decline of twinning rates in many developed countries in the 2000s, while in others, twinning rates have continued to rise over the past decade; considerable differences persist in terms of elective single-embryo transfer practice and adherence to expert recommendations varies from country to country (Pison et al. 2015; ESHRE Guideline Group on the Number of Embryos to Transfer et al. 2024). This divergence may have been influenced by the increased reliance on MAR in relation to fertility postponement, non-ART ovarian stimulation (concerning which the trends remain unclear, Pison et al. 2015), and the expansion of cross-border reproductive care (McKelvey et al. 2009).

This article examines multiple birth trends both internationally and focusing specifically on Czechia. We explore the role of assisted reproductive technology (ART), defined as all interventions that include the in vitro (outside the body) handling of human oocytes, sperm or embryos for the purpose of reproduction (Zegers-Hochschild et al. 2017), in shaping these trends, and highlight the combined impact of reproductive ageing and the expansion in the use of ART since the 1990s. The central research question examines patterns in multiple births in connection which ART and policy measures – especially the promotion of single embryo transfer (SET).

Employing both internationally harmonised datasets and Czech data on multiple births and ART, we firstly present a comparative overview of long-term developments concerning twinning birth rates prior to and following the widespread use of ART. Secondly, we focus on Czechia as an example of a country that has seen a dynamic increase in ART-related fertility (Kocourková et al. 2023). We investigate the timing and impacts of specific policy intervention measures aimed at promoting single embryo transfer. We analyse, at the anonymised individual data level, the extent to which the observed trends – namely the increasing use of ART accompanied by a reduction in the share of multiple embryo transfers and the continued postponement of childbearing to older maternal ages – are reflected in the proportion of multiple births and the structure of newborns from multiple deliveries according to the maternal age and mode of conception (with and without the use of ART).

2. Theoretical background

Numerous studies have identified two key factors that influence the intensity of multiple pregnancies: the age of childbearing and the development of medically-assisted reproduction (MAR). Twinning rates related to spontaneous conception without the use of MAR are associated with the maternal age (Bønnelykke 1990) and certain other factors, e.g. Bulmer (1970) highlighted the birth order and region/country of birth, which partly reflects genetic differences (Hoekstra et al. 2008). Regarding the 1960s, i.e. prior to the widespread use of MAR, Pison et al. (2015) identified a pattern of steadily increasing twinning rates up to the maternal age at birth of 35–39 years, followed by a rapid decline at age 45 and over. Even though the pattern was similar in all the developed countries studied, higher rates were identified for England and Wales than in France, with significantly lower rates in Japan, due possibly to genetic differences; Japan exhibits differing monozygotic twin and dizygotic twin patterns depending on the age of the mother. Prior to 1970, the monozygotic twin birth rate was consistent across maternal ages, birth orders and geographic origin, historically ranging from 3.5 to 4.5 per 1,000 deliveries (Bortolus et al. 1999; Bulmer 1970). In contrast, dizygotic twinning rates vary significantly based on maternal (age, parity) and region factors; the lowest rates have been determined for East Asia and Latin America (2–4 per 1,000) and the highest rates in sub-Saharan Africa (15 per 1,000), with intermediate rates in Europe and the US (7 per 1,000) (Hoekstra et al. 2008; Smits and Monden 2011).

Given the significant increase in the maternal age in developed countries since the 1970s (Frejka and Sobotka 2008), the question of the impact of ART, which acts to increase dizygotic twinning rates due to the transfer of two or more embryos, on the increase in twinning rates in developed countries is of significant importance. The proportion of single embryo transfer cycles at the end of the 1990s for those European countries that reported to ESHRE (The European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology [ESHRE] et al. 2023) comprised a mere 10% of all fresh in vitro fertilisation (IVF)/intracytoplasmic sperm injection (ICSI) cycles, while two and three embryo transfer cycles each accounted for approximately 40% of treatment cycles. Subsequently, the proportion of transfers involving three or more embryos declined steadily. In contrast, the proportion of two-embryo transfers increased and remained stable over the medium term at around 55% of all IVF/ICSI cycles; however, from 2014 onwards the proportion declined continuously to reach around 40% in 2019 (EIM for ESHRE et al. 2023). At the same time, the share of single embryo transfers increased – from slightly above 10% in the late 1990s to more than 55% in 2019 (EIM for ESHRE et al. 2023). As the number of embryos transferred per cycle progressively declined, the global proportion of twin and triplet births decreased accordingly. In 2019, the twin and triplet delivery rates following fresh IVF and ICSI cycles stood at 12.0% and 0.3%, respectively, compared to 26% and 3.5%, respectively in 1997. A similar downward trend was also observed for frozen embryo transfers (FET) (EIM for ESHRE et al. 2023).

While the overall number of embryos transferred per IVF/ICSI cycle has steadily decreased across Europe, significant differences between countries persist. In the late 1990s, countries such as Czechia, Greece and Spain exhibited the highest triple and quadruple embryo transfer rates – up to 76% – compared to 5–10% in Sweden and Finland (Kocourková, Fait 2009). In 2019, most countries reported single embryo transfer rates in excess of 50%, with some (e.g. Iceland and Finland) exceeding 95%; nevertheless, a small number of countries continue to report two and more embryo transfer rates of over 60% of all fresh IVF/ICSI cycles (e.g. Germany, Greece) (EIM for ESHRE et al. 2023).

Compared to the relatively well-documented field of assisted reproductive technology (ART), which is often covered by both international and national registries, significantly less information is available on fertility and children born following other forms of hormonal treatment, even though their contributions to the total number of births is significant. For instance, a study by Blondel et al. (2012) revealed that in France in 2010, 2.3 percent of children were conceived via ovulation induction (i.e. non-ART methods), identical to the percentage of children conceived through in-vitro fertilisation.

In recent decades, trends concerning multiple births have reflected the complex interplay between increasing maternal ages, broader societal changes and medical intervention, particularly assisted reproductive technology (ART). Building upon this context, this paper examines trends in the development of twinning/multiple birth rates and focuses in detail on Czechia via the analysis of the roles of legislative changes/professional guidelines and the maternal age in terms of shaping these trends.

2.1 Availability and public coverage of ART in Czechia

An age limit of 50 years was introduced in Czechia for access to assisted reproductive technology (ART) in 2012. However, legislation introduced in 1997 already provided for at least the partial reimbursement of ART costs for women under the age of 39 for up to three treatment cycles. A significant shift occurred in April 2012, when legislation explicitly encouraged the single embryo transfers via the extension of reimbursement to a maximum of four treatment cycles on the condition that only a single embryo is transferred during the first two cycles. Since 2022, ART treatment costs have been covered by the public health insurance system up to a woman's 40th birthday. However, the reimbursement of ART treatment costs by the statutory health insurance system excludes certain modern laboratory techniques (Kocourková et al. 2023).

ART in Czechia is limited to women who are able to provide written consent from a male partner, thus excluding single women and same-sex couples. Nevertheless, the number of ART cycles per woman of reproductive age in Czechia is one of the highest in Europe (EIM for ESHRE et al. 2023), largely due to cross-border reproductive care for foreign women who travel to Czechia to undergo ART treatment (Volejníková et al. 2025).

3. Data and methods

3.1 Data sources

We employed data obtained from various sources. For international comparison purposes, we relied on data on multiple deliveries obtained from the Human Multiple Births Database (HMBD) and on data on ART obtained from the European Society of Human Reproduction and Embryology (ESHRE), which is freely available. The HMBD (https://www.twinbirths.org/) provides demographic data on twin and higher-order multiple births. It contains annual statistics on deliveries by multiplicity – it distinguishes between singleton and multiple deliveries – and, where available, provides details of the number of children born in multiple deliveries (twin deliveries, triplet deliveries, etc.). The database further includes indicators of the twinning and multiple birth rates, defined as the number of twin or multiple deliveries per 1,000 total deliveries each year (Torres et al. 2023). Twin deliveries account for the vast majority of multiple deliveries – e.g. 97.9% in France in 2010–2015 (Torres et al. 2023), 99% in Czechia (Němečková 2016). Therefore, this article considers both indicators – multiple deliveries in general and, in certain sections, twin deliveries specifically for international comparison purposes.

We employed data from May 2025, at which time the Human Multiple Births Database included information on 25 national populations across Europe, the Americas, Oceania and Asia. We focused exclusively on data from European countries for the period 1950–2022 for the purposes of this analysis. Most countries report data on total births, including both live births and stillbirths. For more information on the data collection approach and the database methodology, see Torres et al. (2023). While the inclusion of stillbirths in reported deliveries varies, most countries report deliveries including stillbirths; whenever both classifications were available, we employed data that included stillbirths in the analysis.

Concerning the use of ART, we employed data that is collected and published annually by the European Society of Human Reproduction and Embryology (ESHRE, https://www.eshre.eu/) in "ART in Europe: results generated from European registries by ESHRE" journal articles, accompanied by a series of detailed tables. Data was provided by 40 European countries in 2019 covering ART treatment that commenced between 1 January and 31 December 2019. The data on deliveries cover the outcomes of treatment performed in 2019 (EIM for ESHRE et al. 2023). This approach differs from the publication of data by the HMBD and the Czech Statistical Office (CZSO) for Czechia (see below), which considers births that occur within a given calendar year. The number of countries displayed in each graph varied since their inclusion depended on the completeness of the data obtained from the two sources, i.e. the HMBD and ESHRE. Neither source covers all European countries, and in the case of ESHRE, not all countries report data from all the clinics in the respective country. Some countries do not report certain data at all, or they report data only from a subset of clinics. More detailed information on the scope of the data presented is provided in the notes below each graph.

Data obtained from the Czech Statistical Office (CZSO) were employed for the detailed analysis of the situation in Czechia. The analysis focused on births according to multiplicity, regardless of whether the children were born alive or stillborn; births according to multiplicity and the maternal age were also considered. The main source comprised the CZSO Demographic Yearbook for the period 1955–2023, specifically the "Births and Children Born from Twins and Triplets by Sex and Mother's Age" table (https://csu.gov.cz/casova_rada_demografie).

Moreover, we employed data on fertility obtained from the Czech Statistical Office and data on ART from the National Register of Assisted Reproduction for the detailed study of children born in twin deliveries according to the maternal age and the role of ART treatment. We considered individual level data stored in the National Register of Reproduction Health, managed by the Institute of Health Information and Statistics of the Czech Republic (IHIS CZ). The reporting of data on ART cycles is compulsory for all reproductive centres/clinics according to legislation introduced in 2006 (IHIS CZ 2022). We employed individual-level data for the period 2013–2020 obtained from the National Register of Assisted Reproduction that was subsequently linked via the women's birth identification numbers aimed at identifying those mothers who gave birth following an embryo transfer in Czechia. We excluded all the ART cycles of foreign women who received cross-border reproductive care in Czechia that were included in the National Register of Assisted Reproduction, since such live births do not form a part of the vital statistics data on live births in Czechia. ART live births were considered to be live births following an embryo transfer that occurred within a maximum of 40 weeks prior to the date of birth.

3.2 Methods

This article combines a variety of data sources aimed at illustrating both the overall situation and trends in multiple births, both internationally and in detail for Czechia; thus, it employs a range of indicators. The individual indicators and the formulas are presented below. The two main analytical approaches applied in the article comprise: a) the analysis of deliveries, and b) the analysis of live births. We therefore distinguished whether the input data refers to deliveries or to live-born children in the formulas. Since demography conventionally works with live births, in accordance with established international practice, we denoted them as N^v. In the case of deliveries, we provide the term in full in the formulas so as to avoid confusion.

Two indicators were used for the analysis of multiple-birth patterns – the twinning rate and the multiple birth rate. We followed the HMBD definition and defined the twinning rate as:

$$\text{Twinning rate } = \frac{\text{Deliveries}^{\text{Twins}}}{\text{Deliveries}} \times 1000$$

And, with respect to the twining rate, more specifically by the age group of women at delivery defined as:

$$\text{Twinning rate}_x = \frac{\text{Deliveries}_x^{\text{Twins}}}{\text{Deliveries}_x} \times 1000$$

where x comprises the −24 years, 25–29 years, 30–34 years, 35–39 years and 40+ years age groups. Three-year moving averages were used to smooth any significant year-to-year fluctuations – particularly for those age groups with relatively low numbers of births.

Similarly, we defined the multiple birth rate as the number of multiple deliveries per 1,000 total deliveries:

$$\text{Multiple birth rate} = \frac{\text{Deliveries}^{\text{Multiple}}}{\text{Deliveries}} \times 1000$$

With regard to the analysis of live births, we used the multiple birth rate in relation to the use of ART methods indicator, defined as the number of ART multiple live births per 1,000 ART live births (4) and the number of non-ART multiple live births per 1,000 non-ART live births (5).

$$\text{Multiple birth rate}^{\text{ART}} = \frac{N^{\text{v, ART, multiple}}}{N^{\text{v, ART}}}$$

$$\text{Multiple birth rate}^{\text{nonART}} = \frac{N^{\text{v, nonART, multiple}}}{N^{\text{v, nonART}}}$$

Furthermore, we calculated the shares of liveborn children from multiple pregnancies disaggregated by the use of ART for all children from multiple pregnancies:

$$\text{Share of multiple births}^{\text{ART}} = \frac{N^{\text{v, ART, multiple}}}{N^{\text{v, multiple}}}$$

$$\text{Share of multiple births}^{\text{nonART}} = \frac{N^{\text{v, nonART, multiple}}}{N^{\text{v, multiple}}}$$

We used the indicator published by the European Society of Human Reproduction and Embryology (ESHRE) – the number of ART cycles per 1 million females of reproductive age in each country for the international comparison of the prevalence of ART across countries. This indicator is published and presented in our analysis only for those countries that report data to ESHRE from all clinics in the respective country since the accuracy of the analysis depended on the calculation of the indicator only for countries that fully register the provision of ART treatment.

4. Results

4.1 Twinning rates in European countries and links to the use of ART

A significant increase in the proportion of twin births has been observed in European countries (Fig. 1) since the early 1990s, prior to which the share of twin births was generally in decline. This trend was associated with decreasing higher-order fertility, which led to a gradual decline in the mean age of mothers at childbirth (Human Fertility Database 2025). Moreover, European populations exhibited a relatively homogeneous pattern in terms of the prevalence of twin births before the 1990s, especially when compared to the divergence that emerged after 1990 (Fig. 1).

From the 1990s onwards, significant advances were made in assisted reproductive technology (ART) with respect to the volume, techniques and quality of care. In addition, differences in national practices – in terms of both the use of ART and the number of embryos transferred per cycle comprised

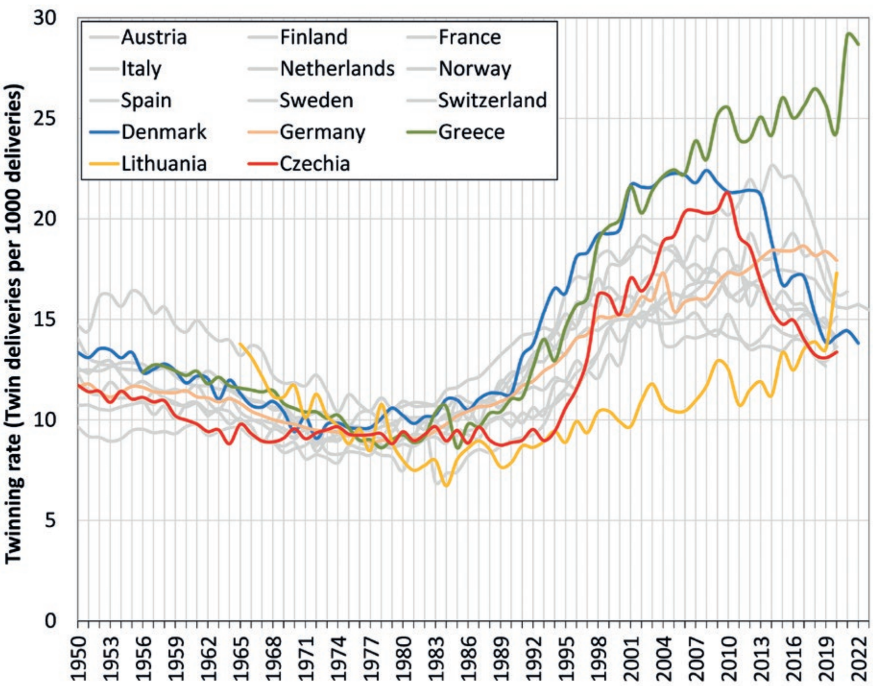

Fig. 1 Twinning rates in selected European countries 1950–2022.
Source: The Human Multiple Births Database (2025).

the principal drivers of the growing heterogeneity of twinning rates across European countries (Fig. 1). Figure 1 shows that Lithuania exhibits a persistently low twinning rate, which exceeded 15 twin deliveries per 1,000 deliveries only as recently as in 2020. Conversely, Greece experienced a sharp and sustained increase in the twinning rate during the 1990s. Unlike several other countries that have adopted more restrictive

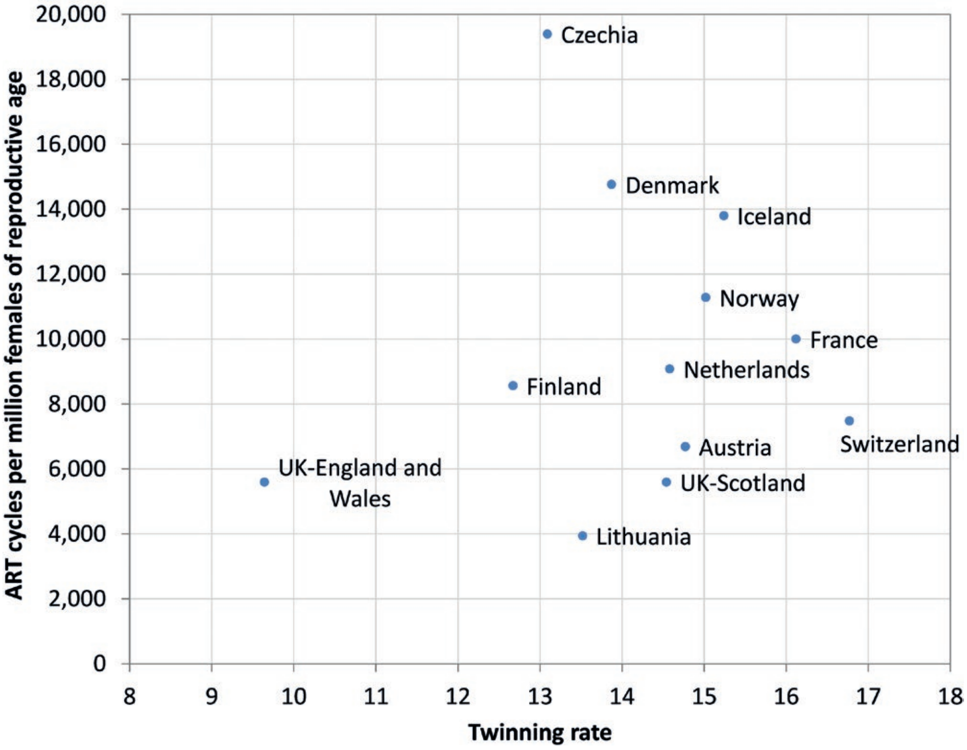

Fig. 2 Twinning rates and number of ART cycles per 1 million females of reproductive age, selected European countries, 2019.
Source: The Human Multiple Births Database (2025) and EIM for ESHRE et al. (2023).
Note: only those countries that reported events at all clinics to the national register of ART in 2019.

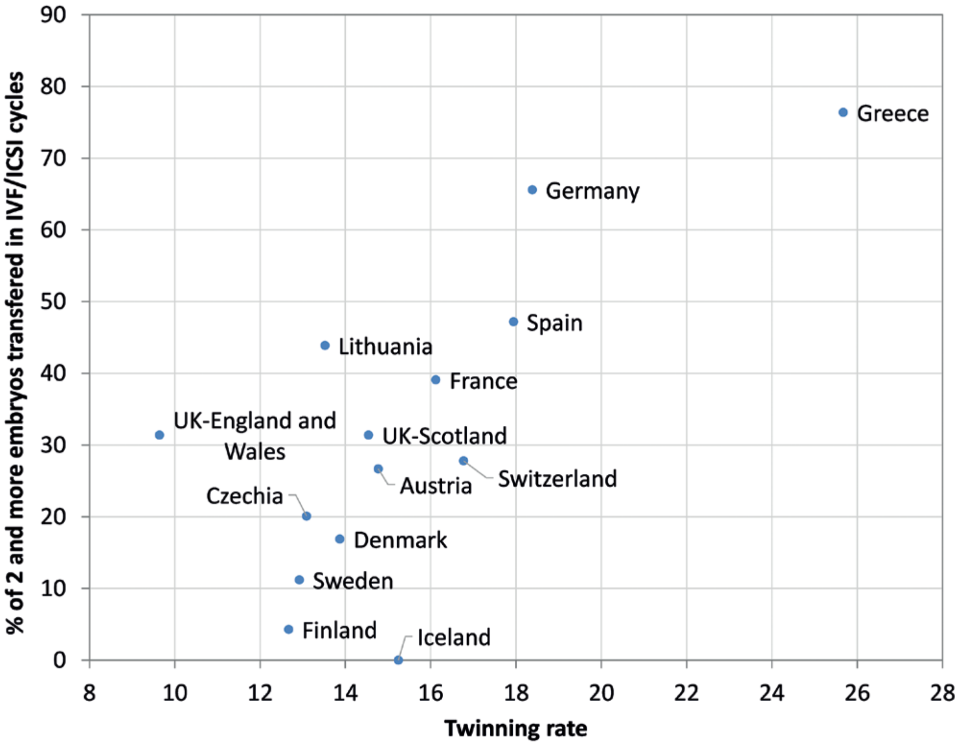

Fig. 3 Twinning rates and the proportion of 2+ embryos transferred in one IVF/ICSI cycle, selected European countries, 2019.
Source: The Human Multiple Births Database (2025) and EIM for ESHRE et al. (2023).

embryo transfer policies since the 1990s, the upward trend in Greece has not slowed significantly and the twinning rate is currently approaching 30 twin deliveries per 1,000 deliveries.

Czechia and Denmark (Fig. 1) shared a common pattern that was observed in many other countries, i.e. a rapid and pronounced initial increase in the twinning rate, which began to decelerate in the early 2000s. The slowdown coincided with the intensification of the professional discourse on, and the gradual implementation of, single embryo transfer policies specifically aimed at reducing the incidence of multiple births and the associated elevated health risks for both mothers and infants. In many countries, this shift – driven by both the medical community and, in part, by legislation – led to the stabilisation of, and in some cases, such as in Czechia and Denmark, to a relatively rapid decline in twinning rates.

The twinning rates currently observed in high-income countries are largely influenced by the expanding use of ART. However, as illustrated in Figures 2 and 3, the overall extent of ART use, as measured by the number of ART cycles per one million women of reproductive age (Fig. 2), is not the primary differentiating factor. Countries with vastly different levels of ART provision may exhibit very similar twinning rates. For example, Czechia, where the high volume of ART services provided is partly attributable to a significant share of foreign women seeking cross-border reproductive care (Volejníková et al. 2025), exhibits the highest level of ART provision relative to the population of women of reproductive age, but only average twinning rates. Indeed, the variations in twinning rates between countries appear to depend to a significant extent on clinical practice, specifically the proportion of cycles involving the transfer of more than one embryo (Fig. 3). For example, more than half of IVF/ICSI cycles provided in Germany and Greece, which report some of the highest twinning rates of the countries analysed, involve the transfer of two or more embryos. Conversely, twinning rates are considerably lower, i.e. typically around 13 to 15 twin deliveries per 1,000 deliveries, in countries in which multiple embryo transfers account for fewer than one-third of IVF/ICSI cycles.

4.2 Czechia – long term trends in the twinning rate focusing on the roles of age and ART

The most pronounced impact of the growing use of ART on the twinning rate was observed in Czechia in the 1990s and the first decade of the new millennium, at which time the twinning/multiple birth rate increased rapidly, with a notable rise in both the number and proportion of multiple births. At the beginning of the 1990s, the twinning rate stood at around 9 twin deliveries per 1,000 deliveries (Fig. 4). By the end of the 1990s, it had doubled, reaching its peak in 2010 at 21 twin deliveries per 1,000 deliveries. Although the sharp rise in the mean age of mothers at birth may also have contributed to the growing incidence of multiple births during the second half of the 1990s, the decisive factor at that time concerned the increasing use of ART and the high proportion of ART cycles

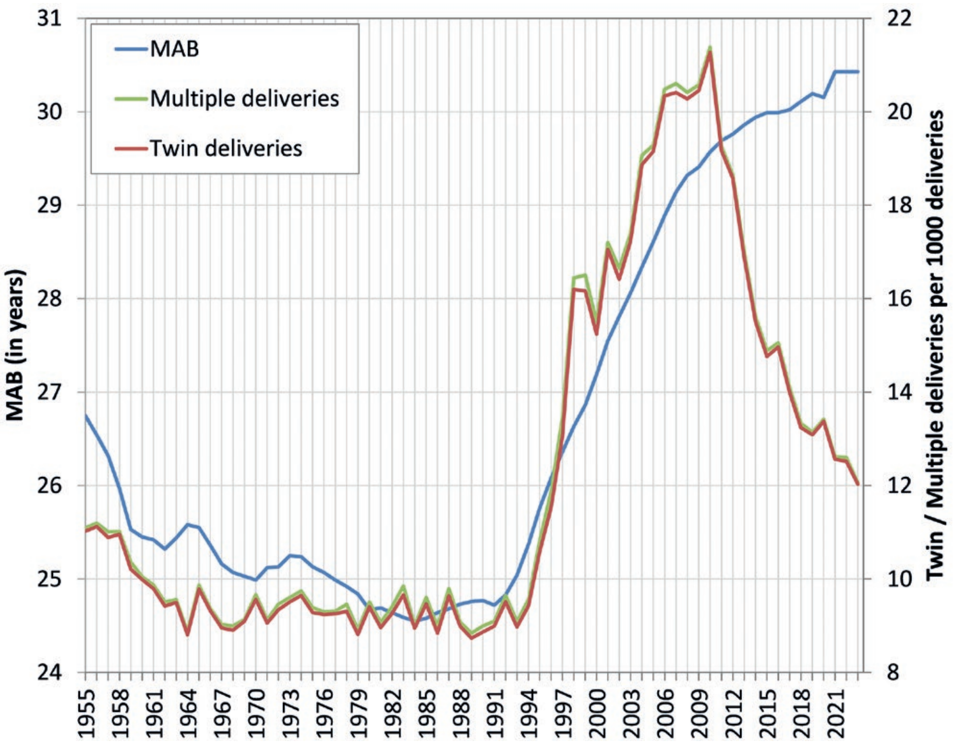

Fig. 4 Mean age at birth (MAB), twinning rate and multiple birth rate, Czechia, 1955–2023.
Source: CZSO (2025), authors' calculations.

involving the transfer of multiple embryos. According to long-term ESHRE statistics, only 10% of IVF/ICSI cycles involved the transfer of a single embryo in Czechia at the end of the 1990s, while the remaining 90% of cycles involved the transfer of multiple embryos – most commonly three or more embryos (76.2% in 1997 and 70.9% in 1999) (The European IVF-monitoring programme [EIM], for the European Society of Human Reproduction and Embryology [ESHRE] 2000; Nygren et al. 2001; Nygren et al. 2002).

The proportion of single embryo transfers subsequently increased gradually; however, the most significant change occurred following a change in public health insurance reimbursement legislation in April 2012, whereby the number of reimbursed cycles was increased from three to four provided that only one embryo was transferred during each of the first two cycles (Kocourková et al. 2015). While in 2009 single embryo transfers accounted for one-fifth of all fresh IVF/ICSI transfers, by 2011 the proportion had risen to 29%, and in 2012 it jumped sharply year-on-year to 47.4% (Ferraretti et al. 2013; The European IVF-Monitoring Consortium [EIM] for the European Society of Human Reproduction and Embryology [ESHRE] for the European Society of Human Reproduction and Embryology [ESHRE] et al. 2016b; The European IVF-Monitoring Consortium [EIM] for the European Society of Human Reproduction and Embryology [ESHRE] et al. 2016a). By 2019, the share of single embryo transfers of IVF/ICSI cycles had reached 80% (EIM for ESHRE et al. 2023). In contrast, transfers of three or more embryos gradually declined to 2.9% in 2012 and just 0.2% of IVF/ICSI cycles in 2019 in Czechia (EIM for ESHRE et al. 2016a; EIM for ESHRE et al. 2023). The changes in clinical practice, supported by the change in reimbursement conditions, were immediately reflected in the proportion of multiple deliveries, even though the mean age at birth continued to rise (Fig. 4). After reaching a peak of 21.4 multiple deliveries per 1,000 deliveries in 2010, a significant decline was observed to 12 per 1,000 in 2023 (Fig. 4).

A closer investigation revealed that the twinning rate varied significantly across maternal age groups, and the sharp increase observed in the second half of the 1990s did not apply uniformly across all age groups (Fig. 5). The lowest twinning rates were determined for women under the age of 25, with only a minimal increase during the 1990s. This pattern reflects the fact that the use of ART by this age group is relatively rare (Kocourková et al. 2023). In contrast, women aged 30–39 evinced the highest twinning rates; moreover, this group also experienced the most substantial growth in the twinning rate in the late 1990s. Between the early 1990s and 1999, the twinning rate in this group almost doubled – from under 15 twin deliveries per 1,000 deliveries to 26 per 1,000 deliveries for women aged 30–34, and to 30 per 1,000 deliveries for those aged 35–39.

Regarding women aged 40 and over, the twinning rate exhibited considerable fluctuation – even when applying a three-year moving average – due to the

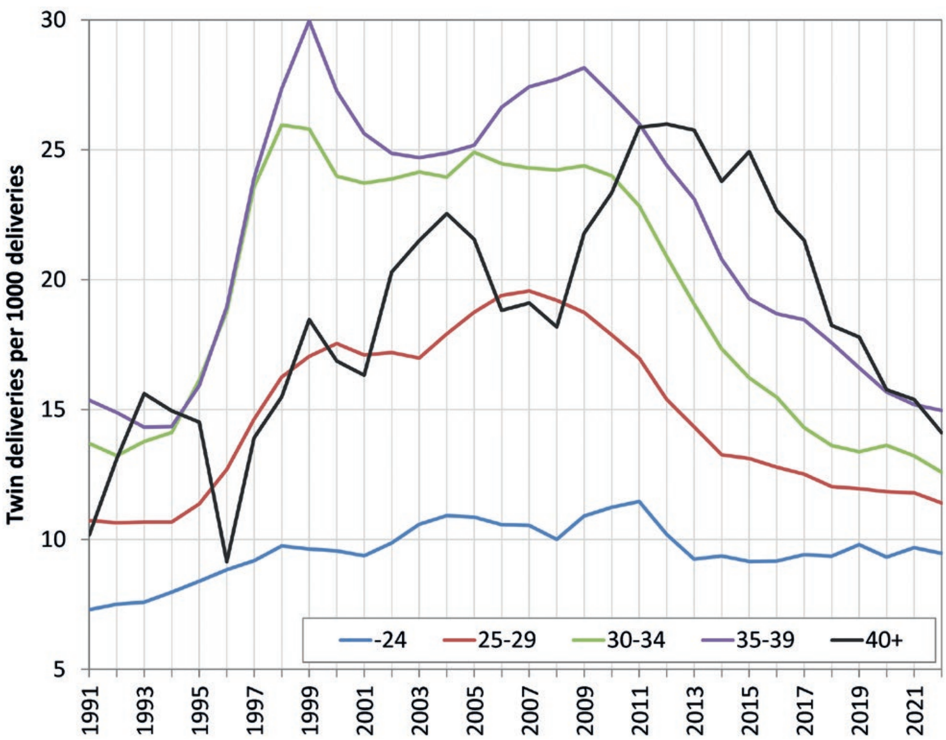

Fig. 5 Twinning rate by the age group of mothers, Czechia, 1990–2023, 3-year moving average.
Source: CZSO (2025), authors' calculations.

relatively low number of births to women in this age group. Despite the volatility, an upward trend was observed up to around 2015 driven largely by the increasing number of ART cycles provided for women aged 40+ (IHIS CZ 2022; Kocourková et al. 2023), even though this group was no longer entitled to claim reimbursement from the public health insurance system for ART treatment.

Towards the end of the observation period, a gradual decline in the twinning rate was observed across all the 25+ age groups to between 12 to 15 twin deliveries per 1,000 deliveries.

4.3 The role of ART and the maternal age in the structure of newborns from multiple deliveries

The analysis of live births provided a more detailed insight into the impact of ART. Employing anonymised individual-level data for Czechia covering live births in the period 2013–2020, we investigated the relationship between the use of ART and the share of children born from multiple pregnancies, as well as the structure of newborns by the age of the mother and the mode of conception (with and without the use of

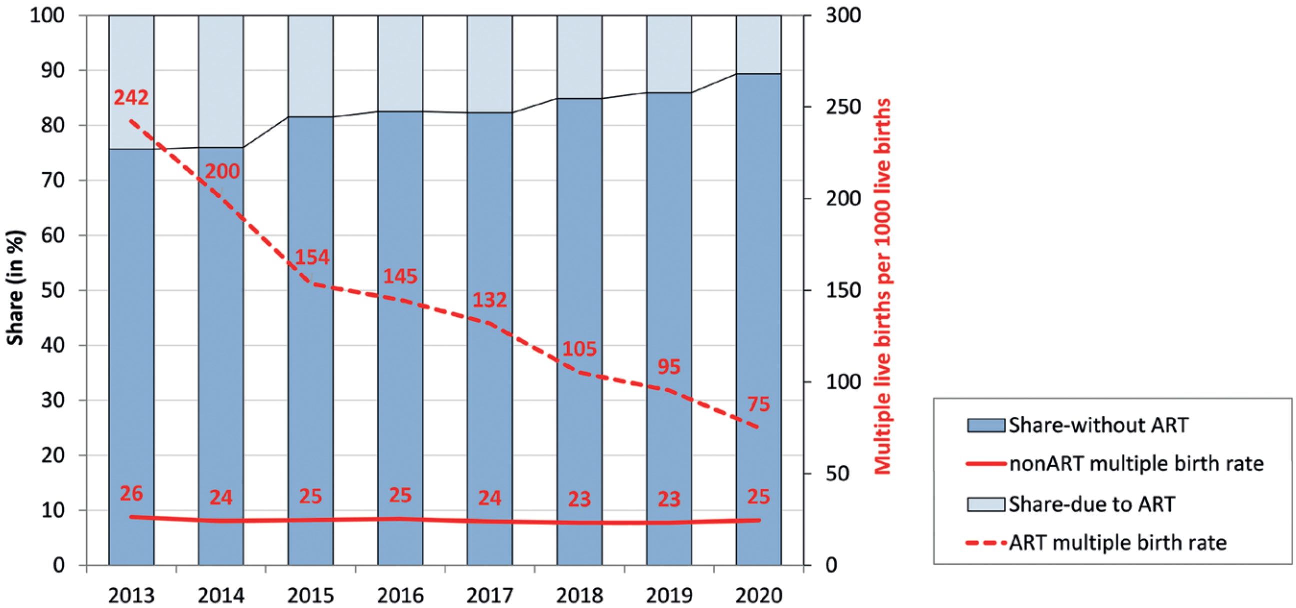

Fig. 6 Multiple live births according to ART use (in %) and the multiple birth rate (non-ART multiple births per 1000 non-ART live births and ART multiple births per 1000 ART births), Czechia, 2013–2020.
Source: authors' calculations.

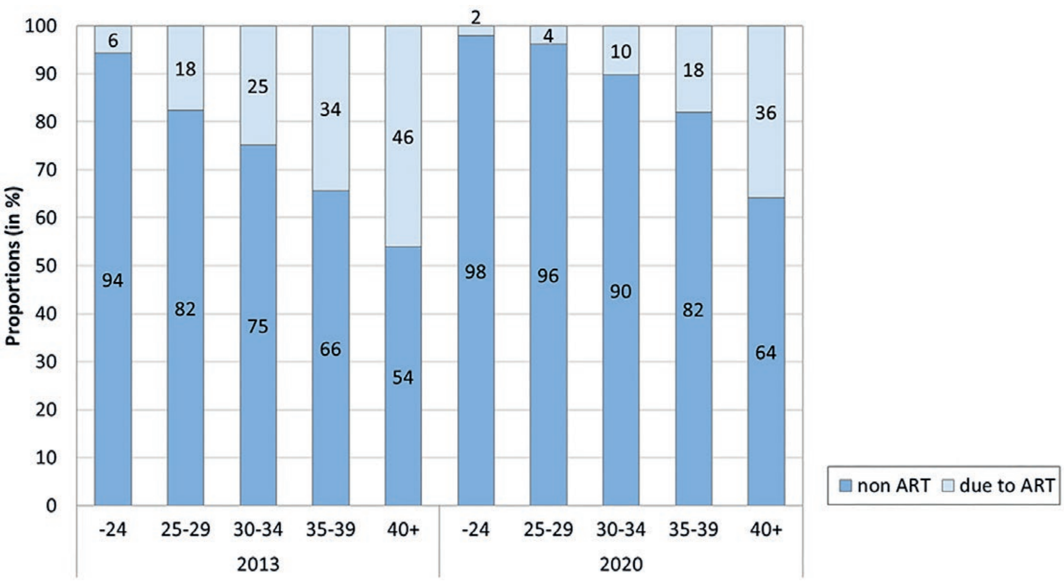

Fig. 7 Multiple live births according to ART use and age of the mother at birth, Czechia, 2013 and 2020 (in %).
Source: authors' calculations.

ART). We focused in detail on the period from 2013, for which detailed individual-level data are available, thus allowing for the analysis of live births according to whether conception followed the use of ART or not (for more details on fertility with and without ART, see Kocourková et al. 2023). This period covered those years in which significant changes in medical practice regarding embryo transfer came into full effect, supported by the amendments to public health insurance reimbursement legislation introduced in 2012.

As a result, the proportion of ART-multiple live births of all multiple live births decreased from 24.3% in 2013 to 10.7% in 2020 (Fig. 6). Similarly, the ART multiple birth rate decreased by more than two-thirds – from 242 multiple births per 1,000 total live births following ART in 2013 to 75‰ in 2020. In contrast, the non-ART multiple birth rate remained unchanged (23–26 multiple births per 1,000 total live births without ART).

Fig. 7 shows that with regard to multiple births, the proportion conceived following ART treatment varied markedly by the maternal age. Not only was the overall incidence of multiple births lower for young mothers (see Fig. 5), but more than 90% were conceived without ART treatment. With increasing maternal ages, multiple births conceived via ART accounted for a progressively higher share of all multiple live births. However, the impact of policies that limited the transfer of more than one embryo is also evident over the observed time period. In 2013, nearly half of all multiples born to women aged 40 and over were conceived via ART, and multiple births due to ART represented approximately one-third of all multiple births for women aged 35–39 years. By 2020, the proportions had declined to 36% and 18%, respectively (Fig. 7).

5. Conclusions

This study examined trends in the incidence of twin and multiple births in the international comparative context, focusing particularly on Czechia. We explored the interplay between the expansion of assisted reproductive technology (ART) since the 1990s and the subsequent promotion of single embryo transfer. The results revealed substantial heterogeneity in terms of the evolution of twinning rates across countries. While Czechia was found to be one of a number of countries in which the promotion of single embryo transfer in clinical practice was followed by a rapid decline in the twinning rate, other European countries displayed markedly different trends – ranging from stagnation (e.g. Germany) to a continuous increase in twin births up to the present day (e.g. Greece). The analysis indicated that the significant variation between countries was driven primarily by differences in national ART practices, including the prevalence of multiple embryo transfers over single embryo transfers, rather than the overall volume of ART care provided. It was also most likely influenced by the role of non-ART medically-assisted reproduction (Pison et al. 2015), which we were unable to examine due to the limited availability of statistics on MAR methods other than in vitro fertilisation.

The 2012 revision of reimbursement legislation in Czechia, which linked an increase in the number of ART cycles subsidised by the public health insurance system to single embryo transfers in the first two attempts, led to a sharp rise in the share of single embryo transfers and the almost complete elimination of transfers involving three or more embryos (Ferraretti et al. 2013; EIM for ESHRE et al. 2016a; EIM for ESHRE et al. 2016b; EIM for ESHRE et al. 2023). These changes in clinical practice were quickly reflected in a substantial decline in the twinning rate, which fell from a peak of 21.4 per 1,000 deliveries in 2010 to 12 per 1,000 in 2023.

Despite the overall decline in the number of ART-related multiples in Czechia, the results of the study highlight the marked heterogeneity of multiple birth rates across maternal age groups: the lowest and most stable rates were observed for women under 25, while women aged 30 and over consistently recorded the highest rates. The analysis of live births revealed that, despite the decline in the overall number of multiple births, women aged 35 and over accounted for a substantial share of multiple live births conceived via ART; as recently as in 2020, 18% and over one-third of multiple live births resulted from ART treatment concerning the 35–39 and 40+ age groups, respectively.

Acknowledgements

This paper was supported by the Czech Science Foundation (No. 21-31691S), the “Demographic consequences of assisted reproduction in Czechia” project.

References

Blondel, B., Macfarlane, A., Gissler, M., Breart, G., Zeitlin, J. (2006): General obstetrics: Preterm birth and multiple pregnancy in European countries participating in the PERISTAT project. An International Journal of Obstetrics & Gynaecology 113(5), 528–35, https://doi.org/10.1111/j.1471-0528.2006.00923.x.

Blondel, B., Lelong, N., Kermarrec, M., Goffinet, F. (2012): Trends in perinatal health in France from 1995 to 2010. Results from the French National Perinatal Surveys. Journal de Gynécologie Obstétrique et Biologie de la Reproduction 41(4), e1–e15, https://doi.org/10.1016/j.jgyn.2012.04.014.

Bønnelykke, B. (1990): Maternal Age and Parity as Predictors of Human Twinning. Acta geneticae medicae et gemellologiae: twin research, 39(3), 329–34, https://doi.org/10.1017/S0001566000005237.

Bortolus, R., Parazzini, F., Chatenoud, L., Benzi, G., Bianchi, M. M., Marini, A. (1999): The epidemiology of multiple

births. Human Reproduction Update 5(2), 179–187, https://doi.org/10.1093/humupd/5.2.179.
Bulmer, M. G. (1970): The Biology of Twinning in Man. Oxford: Clarendon Press, 205 p.
CZSO (2025): Demographic Yearbook, table Births and Children Born from Twins and Triplets by Sex and Mother's Age. Available online: https://csu.gov.cz/casova_rada_demografie (accessed on 1 January 2025).
ESHRE Guideline Group on the Number of Embryos to Transfer, Alteri, A., Arroyo, G., Baccino, G., Craciunas, L., De Geyter, Ch., Ebner, T., Koleva, M., Kordic, K., Mcheik, S., Mertes, H., Pavicic Baldani, D., Rodriguez-Wallberg, K. A., Rugescu, I., Santos-Ribeiro, S., Tilleman, K., Woodward, B., Vermeulen, N., Veleva, Z. (2024): ESHRE guideline: number of embryos to transfer during IVF/ICSI. Human Reproduction 39(4), 647–657, https://doi.org/10.1093/humrep/deae010.
Fait, T., Šťastná, A., Kocourková, J., Waldaufová, E., Šídlo, L., Kníže, M. (2022): Has the cesarean epidemic in Czechia been reversed despite fertility postponement? BMC Pregnancy and Childbirth 22: 469, https://doi.org/10.1186/s12884-022-04781-1.
Ferraretti, A. P., Goossens, V., Kupka, M., Bhattacharya, S., de Mouzon, J., Castilla, J. A., Erb, K., Korsak, V., Nyboe Andersen, A., The European IVF-Monitoring (EIM) Consortium for the European Society of Human Reproduction and Embryology (ESHRE), Strohmer, H., Bogaerts, K., Kyurkchiev, S., Vrcic, H., Pelekanos, M., Rezabek, K., Erb, K., Gissler, M., Royere, D., Bühler, K., Tarlatzis, B. C., Kosztolanyi, G., Bjorgvinsson, H., Mocanu, E., Scaravelli, G., Lokshin, V., Arajs, M., Gudleviciene, Z., Lazarevski, S., Moshin, V., Motrenko Simic, T., Hazekamp, J. T., Kurzawa, R., Calhaz-Jorge, C., Rugescu, I., Korsak, V., Radunovic, N., Tomazevic, T., Hernandez Hernandez, J., Karlström, P. O., Weder, M., Lambalk, C., Veselovsky, V., Baranowski, R. (2013): Assisted reproductive technology in Europe, 2009: results generated from European registers by ESHRE. Human Reproduction 28(9), 2318–2331, https://doi.org/10.1093/humrep/det278.
Frejka, T., Sobotka, T. (2008): Overview Chapter 1: Fertility in Europe: Diverse, delayed and below replacement. Demographic Research 19(3), 15–46, https://doi.org/10.4054/DemRes.2008.19.3.
Hoekstra, Ch., Zhao, Z. Z., Lambalk, C. B., Willemsen, G., Martin, N. G., Boomsma, D. I., Montgomery, G. W. (2008): Dizygotic twinning, Human Reproduction Update 14(1), 37–47, https://doi.org/10.1093/humupd/dmm036.
Human Fertility Database (2025): Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Available online: www.humanfertility.org (accessed on 5 April 2025).
Human Multiple Births Database (2025): French Institute for Demographic Studies – INED (distributor). Available online: https://www.twinbirths.org (accessed on 1 June 2025).
IHIS CZ (2022): Assisted reproduction in the Czech Republic in 2020. Available online: www.uzis.cz/res/f/008420/asistreprodukce2020.pdf (accessed on 5 April 2025).
Kocourková, J., Fait, T. (2009): Can increased use of ART retrieve the Czech Republic from the low fertility trap? Neuro Endocrinology Letters 30(6), 739–748, https://pubmed.ncbi.nlm.nih.gov/20038934.
Kocourková, J., Konečná, H., Burcin, B., Kučera, T. (2015): How old is too old? A contribution to the discussion on age limits for assisted reproduction technique access. Reproductive BioMedicine Online 30, 482–492, https://doi.org/10.1016/j.rbmo.2015.01.017.
Kocourková, J., Šťastná, A., Burcin, B. (2023): The influence of the increasing use of assisted reproduction technologies on the recent growth in fertility in Czechia. Scientific Reports 13: 10854, https://doi.org/10.1038/s41598-023-37071-7.
Marleen, S., Kodithuwakku, W., Nandasena, R., Mohideen, S., Allotey, J., Fernández-García, S., Gaetano-Gil, A., Ruiz-Calvo, G., Aquilina, J., Khalil, A., Bhide, P., Zamora, J., Thangaratinam, S. (2024): Maternal and perinatal outcomes in twin pregnancies following assisted reproduction: a systematic review and meta-analysis involving 802 462 pregnancies. Human Reproduction Update 30(3), 309–322, https://doi.org/10.1093/humupd/dmae002.
McKelvey, A., David, A., Shenfield, F., Jauniaux, E. (2009): The impact of cross-border reproductive care or 'fertility tourism' on NHS maternity services. BJOG: An International Journal of Obstetrics and Gynaecology 116(11), 1520–1523, https://doi.org/10.1111/j.1471-0528.2009.02294.x.
Němečková, M. (2016): Vícečetných porodů je již méně (Multiple births are becoming less common). Statistika&My. Available online: https://statistikaamy.csu.gov.cz/vicecetnych-porodu-je-jiz-mene (accessed on 5 April 2024).
Nygren, K. G., Nyboe Andersen, A., The European IVF-monitoring programme (EIM) for the European Society of Human Reproduction Embryology (ESHRE) (2001): Assisted reproductive technology in Europe, 1998. Results generated from European registers by ESHRE. Human Reproduction 16(11), 2459–2471, https://doi.org/10.1093/humrep/16.11.2459.
Nygren, K. G., Nyboe Andersen, A., The European IVF-monitoring programme (EIM) (2002): Assisted reproductive technology in Europe, 1999. Results generated from European registers by ESHRE. Human Reproduction 17(12), 3260–3274, https://doi.org/10.1093/humrep/17.12.3260.
Pison, G., Monden, Ch., Smits, J. (2015): Twinning Rates in Developed Countries: Trends and Explanations. Population and Development Review 41(4), 629–649, https://doi.org/10.1111/j.1728-4457.2015.00088.x.
Santana, D. S., Cecatti, J. G., Surita, F. G., Silveira, C., Costa, M. L., Souza, J. P., Mazhar, S. B., Jayaratne, K., Qureshi, Z., Sousa, M. H., Vogel, J. P. MD on behalf of the WHO Multicountry Survey on Maternal and Newborn Health Research Network (2016): Twin Pregnancy and Severe Maternal Outcomes: The World Health Organization Multicountry Survey on Maternal and Newborn Health. Obstetrics & Gynecology 127(4), 631–641, https://doi.org/10.1097/AOG.0000000000001338.
Smits, J., Monden, C. (2011): Twinning across the Developing World. PLoS ONE 6(9): e25239, https://doi.org/10.1371/journal.pone.0025239.
The European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE), Calhaz-Jorge, C., de Geyter, C., Kupka, M. S., de Mouzon, J., Erb, K., Mocanu, E., Motrenko, T., Scaravelli, G., Wyns, C., Goossens, V.,

Gliozheni, O., Strohmer, H., Petrovskaya, E., Tishkevich, O., Bogaerts, K., Antonova, I., Vrcic, H., Ljiljak, D., Rezabek, K., Markova, J., Lemmen, J., Sõritsa, D., Gissler, M., Tiitinen, A., Royere, D., Tandler-Schneider, A., Uszkoriet, M., Loutradis, D., Tarlatzis, B. C., Urbancsek, J., Kosztolanyi, G., Bjorgvinsson, H., Mocanu, E., Scaravelli, G., Lokshin, V., Ravil, V., Gudleviciene, Z., Belo Lopes, G., Moshin, V., Motrenko Simic, T., Vukicevic, D., Romundstad, L. B., Kurzawa, R., Calhaz-Jorge, C., Laranjeira, A. R., Rugescu, I., Doroftei, B., Korsak, V., Radunovic, N., Tabs, N., Tomazevic, T., Virant-Klun, I., Hernandez Hernandez, J., Castilla Alcalá, J. A., Bergh, C., Weder, M., De Geyter, C., Smeenk, J. M. J., Gryshchenko, M., Baranowski, R. (2016a): Assisted reproductive technology in Europe, 2012: results generated from European registers by ESHRE, Human Reproduction 31(8), 1638–1652, https://doi.org/10.1093/humrep/dew151.

The European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE), Kupka, M. S., D'Hooghe, T., Ferraretti, A. P., de Mouzon, J., Erb, K., Castilla, J. A., Calhaz-Jorge, C., De Geyter, C., Goossens, V. (2016b): Assisted reproductive technology in Europe, 2011: results generated from European registers by ESHRE. Human Reproduction 31(2), 233–248, https://doi.org/10.1093/humrep/dev319.

The European IVF-monitoring programme (EIM), for the European Society of Human Reproduction and Embryology (ESHRE) (2000): Assisted reproductive technology in Europe, 1997. Results generated from European registers by ESHRE. Human Reproduction 16(2), 384–391, https://doi.org/10.1093/humrep/16.2.384.

The European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE), Smeenk, J., Wyns, C., De Geyter, C., Kupka, M., Bergh, C., Cuevas Saiz, I., De Neubourg, D., Rezabek, K., Tandler-Schneider, A., Rugescu, I., Goossens, V. (2023): ART in Europe, 2019: results generated from European registries by ESHRE. Human Reproduction 38(12), 2321–2338, https://doi.org/10.1093/humrep/dead197.

Torres, C., Pison, G., Caporali, A. (2023): The Human Multiple Births Database (HMBD): An international database on twin and other multiple births. Demographic Research 48(4), 89–106, https://doi.org/10.4054/DemRes.2023.48.4.

Volejníková, A., Šťastná, A., Kocourková, J. (2025): Cross-border Reproductive Care in Czechia: Insights from National Registry Data. Comparative Population Studies 50, 183–204, https://doi.org/10.12765/CPoS-2025-12.

Whittaker, M., Greatholder, I., Kilby, M. D., Heazell, A. E. P. (2023): Risk factors for adverse outcomes in twin pregnancies: a narrative review. The Journal of Maternal-Fetal & Neonatal Medicine 36(2), https://doi.org/10.1080/14767058.2023.2240467.

Zegers-Hochschild, F., Adamson, G. D., Dyer, S., Racowsky, C., de Mouzon, J., Sokol, R., Rienzi, L., Sunde, A., Schmidt, L., Cooke, I. D., Simpson, J. L., van der Poel, S. (2017): The International Glossary on Infertility and Fertility Care. Human Reproduction 32(9), 1786–1801, https://doi.org/10.1093/humrep/dex234.

Shifting seasons: Long-term crop dynamics across agroclimatic regions of Czechia

Jiří Tomíček[1,2], Jan Mišurec[2], Markéta Potůčková[1,*]

[1] Charles University, Faculty of Science, Department of Applied Geoinformatics and Cartography, Czechia
[2] Gisat Ltd., Czechia
[*] Corresponding author: marketa.potuckova@natur.cuni.cz

ABSTRACT

This study analyzes the evolution of phenological (start-of-season, end-of-season, length-of-season, day of maximum-of-season) and productivity (small and large seasonal integrals) parameters for six major crop types in Czechia (winter cereals, spring cereals, winter rapeseed, fodder crops, sugar beetroot, and corn), using a 35-year Landsat time series (1986–2020). The leaf area index (LAI) was retrieved using an artificial neural network regression model trained on PROSAIL radiative transfer simulations and validated with extensive in situ measurements collected in 2017 and 2018 in the lowlands of Central Bohemia. The supervised classification of Landsat quarterly composites enabled the identification of crop spatial patterns for each growing season. Phenological and productivity indicators were then derived from LAI time series aggregated at the level of ten agro-climatic regions using the threshold approach. Changes in phenological and productivity parameters over the examined period were assessed through the linear least squares regression analysis and the significance of trends was tested. Results revealed significant negative trends in the end-of-season and day of maximum-of-season for winter and spring cereals, winter rapeseed (up to –0.7 days/year), and fodder crops (up to –1.6 days/year), indicating an earlier maturation and harvest. Significant differences in trends in phenological and productivity parameters were observed between agro-climatic regions in more than 40% of cases, and the response was observed to be highly crop-specific. While the shift in harvest dates and the shortening of the season for corn and fodder crops were more pronounced in warmer regions, the shift in winter rapeseed phenology occurred more rapidly in colder regions. The findings underscore the relevance of crop type and regional climate in shaping phenological responses, offering a basis for future research and planning of agricultural adaptation strategies.

KEYWORDS

Landsat; leaf area index; PROSAIL; land surface phenology; productivity; Czechia

Received: 20 June 2025
Accepted: 8 October 2025
Published online: 27 November 2025

Tomíček, J., Mišurec, J., Potůčková, M. (2025): Shifting seasons: Long-term crop dynamics across agroclimatic regions of Czechia. AUC Geographica 60(2), 233–253
https://doi.org/10.14712/23361980.2025.22

1. Introduction

Phenology studies periodic life cycles of living organisms in relation to weather, climate or other biotic and abiotic factors (Lieth 1974; Nord and Lynch 2009). It is mostly based on monitoring time occurrence of certain clearly recognizable signs of plant development (i.e. emergence, flowering, changes in coloring, etc.), which are generally known as "phenological events". The periods between them are then referred as "phenological phases (Caparros-Santiago et al. 2021). The shifts of phenological events due to climate change have been foreseen and studied for decades (Sparks and Carey 1995; Menzel 2000; Hassan et al. 2023). In the case of plants, the most evident are advancements in spring and summer but delay in autumn (Caparros-Santiago et al. 2021; Hassan et al. 2023; Campioli et al. 2025). An increasing length of the growing season on one hand causes an increase in net primary production but on the other hand has impacts on atmospheric CO2 content (positive in spring time, negative due to extending autumn phenophases), water exchange or alternation in species interaction which may cause a decrease of biodiversity (Caparros-Santiago et al. 2021; Yuan et al. 2024).

The ground phenological (GP) observations have a long history (Koch et al. 2007; Hajkova et al. 2012; Fitchett et al. 2015). Standardized procedures have been developed to date. Thus, the phenological data can be collected by volunteers in order to obtain higher spatial coverage (Kaspar et al. 2014). In spite of these efforts, the number of measurements is limited both in time and space. Satellite based monitoring of vegetation growth stages, known as Land Surface Phenology (LSP), allows for much larger spatial scale using time series of vegetation-related characteristics derived from the multispectral imagery (Caparros-Santiago et al. 2021; Gašparović et al. 2024). Unlike GP, LSP does not determine the phenological events based on the presence of specific signs of plant development. Instead, it defines the date on which a certain level of the vegetation-related characteristics under consideration is achieved by observing the vegetation cover (e.g., the date on which the maximum value of a given vegetation-related indicator is reached in a given year). LSP brings advantages, such as being cost-effective and easier to relate to climatic measurements that are usually coarser in resolution and might be difficult to fit GP observations. However, LSP is also affected by noise caused by sensor and processing flaws, or mixed signals from multiple land covers. It is also better suited to community-based than individual-based observations. It is common practice to combine GP and LSP observations when GP serves as the ground truth for deriving and testing LSP models, and when LSP is used to upscale GP observations (Rodriguez-Galiano et al. 2015).

Phenological observations have long received attention in agriculture (Wielgolaski 1974; Chmielewski 2013). They are essential for crop management (e.g., efficient irrigation, fertilization, pest management), yield estimation, or controlling crop agricultural policies (Meroni et al. 2021; Pei et al. 2025). Satellite data with a wide range of spectral, spatial, and temporal resolutions is used for crop phenology mapping (Gao and Zhang 2021). The sensors on the Landsat satellites provide long-term data with high spatial (30 m) and temporal (16 days) resolution. In addition to the spatial resolution of 10 or 20 m, the Sentinel 2 MSI sensor offers a higher temporal resolution (up to 5 days) thanks to the constellation of two satellites. Compared to Landsat data, it also has additional red-edge and SWIR spectral bands suitable for vegetation monitoring. As a result, both Landsat and Sentinel data, as well as their harmonized products (HLS) (Claverie et al. 2018), are among the most widely used for monitoring agricultural crops at regional or higher levels (Chaves et al. 2020; Misra et al. 2020; Gao and Zhang 2021; Htitiou et al. 2024). They can be fused with lower spatial but higher temporal resolution data, such as Moderate Resolution Imaging Spectroradiometer (MODIS) or Advanced Baseline Imager (ABI) to densify the time series for modeling crop grows in near real time (Schreier et al. 2021; Sisheber et al. 2022; Dhillon et al. 2023; Shen et al. 2023). Some authors have attempted to solve the problem of clouds in optical data by fusion with SAR data. Meroni et al. (2021) showed the complementarity of Sentinel-1 and Sentinel-2 data for LSP retrieval, especially for winter crops.

Key LPS parameters derived from satellite images include start of the season (SOS), end of the season (EOS), length of the season (LOS), peak of the season (POS), mild greenup and mild greendown (Hanes et al. 2014). The common way to determine them is to apply curve-based or trend-based approaches to the generated time series of the selected vegetation index (VI) such as normalized differential vegetation index (NDVI), enhanced vegetation index (EVI), or leaf area index (LAI). Curve-based approaches fit phenological curves derived from historical time series of VIs to current observations to predict current and future crop growth stages. They are robust and reliable for crops with consistent growth cycles. Trend-based approaches detect upward or downward trends from current time series data using momentum and VI thresholds. They are simpler to implement and more flexible to unexpected changes in crop growth patterns. However, they are less effective for forecasting future phenological stages and are more susceptible to noise and anomalies in the data (Eklundh and Jönsson 2016; Gao and Zhang 2021). Based on VI time series and LSP parameters, crop biomass (Dong et al. 2020), gross and net primary production (Gitelson et al. 2012), or yield (Skakun et al. 2019; dos Santos Luciano et al. 2021; Dhillon et al. 2023; Zhang et al. 2023; Řezník et al. 2020) can be modeled and estimated.

Long-term satellite observations make it possible to track changes in LSP over recent decades and place them in the context of global change. Studies of global trends of LSP parameters, mainly SOS and EOS, based on MODIS and the Advanced Very High Resolution Radiometer (AVHRR) data showed variances across climatic regions over the Northen Hemisphere (Jeong et al. 2011; Zhang et al. 2014) and the worldwide (Zhang et al. 2014). While the first study points to significant shifts in SOS and EOS in Europe, the latter one finds the overall trends in Europe, and especially in its temperate climate region, generally insignificant. Yuan et al. (2024) provide an overview of the impacts of global climate change on agricultural production. Bartošová et al. (2025) present differences in long-term GP observations (1961–2021) from Czechia for wild plants and agricultural crops across three altitude intervals. They observed some asynchrony in phenological shifts, with agricultural crops showing more pronounced shifts towards the beginning of the season compared to wild plants especially in low (0–299 m) and mid altitudes (300–499 m).

The present study uses a Landsat time series spanning over 30 years (1986–2020) to investigate long-term trends in crop development and productivity across Czechia. Specifically, we focus on the following objectives:

1. Extracting selected crop phenological (start-of-season: SOS, end-of-season: EOS, length-of-season: LOS, and day of maximum-of-season: MAX_DOY) and productivity characteristics (namely the small and large integral of the seasonal curve, SINT and LINT, respectively).
2. Analysing crop-specific temporal patterns to understand how these phenological and productivity metrics evolve over time for different crop types.
3. Assessing the role of natural conditions by examining how the observed trends in crop development and productivity vary across agro-climatic regions of Czechia.

The innovative element lies in linking long-term satellite-derived crop metrics with regional agro-climatic variability, offering new insights into the spatial and temporal dynamics of agricultural systems under changing environmental conditions.

Based on GP observations (Bartošová et al. 2025), we formulate the following hypotheses:

H1: Significant temporal trends in LSP and productivity characteristics (SOS, EOS, LOS, MAX_DOY, SINT, LINT) are expected over the observed period (1986–2020).

H2: Variability in trends will be detectable across agro-climatic regions of Czechia, enabled by the use of high spatial resolution Landsat data and detailed knowledge of crop distribution.

The following crops (or groups of crops) were taken into account for the analysis 1) winter cereals (including winter wheat, winter barley, winter rye etc.), 2) spring cereals (including spring wheat, spring barley, oat, spring rye etc.), 3) winter rapeseed, 4) fodder crops (including alfalfa, clover etc.), 5) sugar beetroot and 6) corn. The reasons for this selection were following: 1) the selected crops are the most frequent ones in the conditions of Czechia as they represent ca. 93% of the arable land in the country, and 2) they represent crops with different requirements for growing conditions. For the purpose of the study, the definitions of the Vegetation Phenology and Productivity parameters by Copernicus Land Monitoring Service is used (HR-VPP: User Manual). The study builds on previous work of the authors when a radiative transfer model-based algorithm for retrieval of LAI from Sentinel-2 and Landsat data for dominant crop types in Czechia was proposed and implemented (Tomíček et al. 2021; Tomíček et al. 2022).

2. Study area

The area of interest covers the entire Czechia. We used the Czech national agroclimatic regionalization,

Tab. 1 Characteristics of climatic regions according to decree No. 327/1998 Coll. issued by the Ministry of Agriculture.

Region code	Region characteristic	Sum of temp. above 10 °C	Mean annual temp.	Mean annual precipitation
0:VT	very warm, dry	2800–3100	9–10 °C	500–600 mm
1:T1	warm, dry	2600–2800	8–9 °C	< 500 mm
2:T2	warm, mildly dry	2600–2800	8–9 °C	500–600 mm
3:T3	warm, mildly humid	2500–2800	(7)8–9 °C	550–650 (700) mm
4:MT1	mildly warm, dry	2400–2600	7–8.5 °C	450–550 mm
5:MT2	mildly warm, mildly humid	2200–2500	7–8 °C	550–650 (700) mm
6:MT3	mildly warm (to warm), humid	2500–2700	7.5–8.5 °C	700–900 mm
7:MT4	mildly warm, humid	2200–2400	6–7 °C	650–750 mm
8:MCh	mildly cold, humid	2000–2200	5–6 °C	700–800 mm
9:CH	cold, humid	< 2000	< 5 °C	> 800 mm

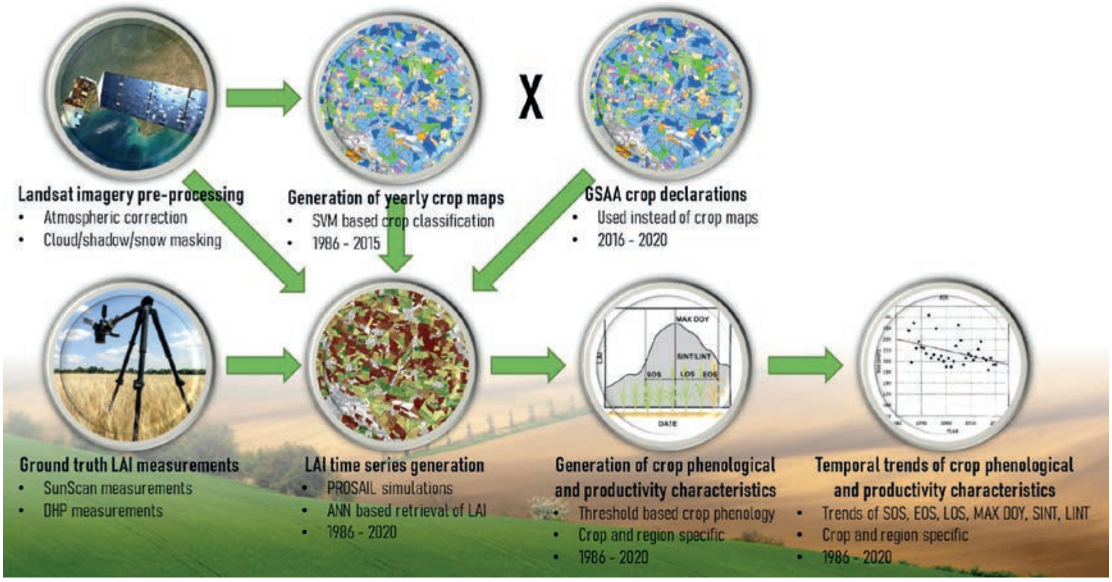

Fig. 1 Methodology and processing workflow comprising Landsat time series pre-processing, ground leaf area index (LAI) measurements, LAI retrieval based on the PROSAIL radiative transfer model and an artificial neural network (ANN) regression model, crop classification, derivation of crop phenological and productivity characteristics and their analysis over agroclimatic zones.

as defined by Decree No. 327/1998 Coll. issued by the Ministry of Agriculture (Decree No. 327/1998), which divides the territory of Czechia into 10 agroclimatic zones based on temperature and humidity characteristics. Table 1 contains the values of the key climatic characteristics of individual agroclimatic regions.

3. Materials and methods

The Landsat series of satellites has been providing high-resolution multispectral data for more than three decades. For the present study, the Landsat time series covered the period from 1986 to 2020. The overall methodology and processing workflow is depicted in Fig. 1 and described in detail in the following subsections 3.1–3.5.

3.1 Landsat imagery preprocessing

The Landsat spectral bands with native spatial resolution of 30 m were used in this study. The raw Level 1 scenes were processed to Level 2 (top-of-canopy reflectance) in the ARCSI (Atmospheric and Radiometric Correction of Satellite Imagery) software (ARCSI GitHub). Invalid or defective pixels (such as snow, cloud and shadows, saturated pixels, etc.) were masked using the FMask algorithm (Zhu and Woodcock 2012; Zhu et al. 2015).

3.2 Development of the Leaf Area Index retrieval model

3.2.1 Ground-truth LAI measurements

Ground-based LAI measurements were collected in Elbeland, a fertile lowland area in central Bohemia belonging to agroclimatic region T2 (average annual temperature 8–9 °C, precipitation 500–600 mm; Tab. 1). this area is considered one of the most fertile in the Czechia. Reference LAI values were measured using two methods: (1) with a Delta-T SunScan instrument (Webb et al. 2016), and (2) through digital hemispherical photography (DHP). At each sampling point (an area of approximately 20 × 20 m), either five SunScan measurements, eight DHP images, or both were collected – depending on site conditions – and averaged as reference values. Points where both DHP and SunScan LAI measurements were collected simultaneously allowed for direct comparison between the two methods. To maximize the consistency between the two LAI datasets, a simple linear transformation was applied to the SunScan-derived LAI values; see Tomíček et al. (2021) for more detailed description. In total, 432 points were measured on 39 plots in 2017 and 2018 in Elbeland, central Bohemia (Fig. 2). Reference data were used to calibrate and validate the LAI estimation model (section 3.2.2. for details). The campaigns were scheduled to cover key phenological phases of the growing season (campaign dates together with reference Landsat scenes are listed in Tab. 2).

3.2.2 LAI retrieval approach

The applied approach of LAI retrieval from high resolution satellite data was proposed in our previous studies (Tomíček et al. 2021; Tomíček et al. 2022). The developed algorithm uses crop-optimized PROSAIL radiative transfer model (RTM) to generate a database of simulations for training of the regression model. Ranges and distribution functions of biophysical, biochemical and structural parameters (the input parameters of the PROSAIL RTM) for individual crops of interest were derived based on an extensive dataset

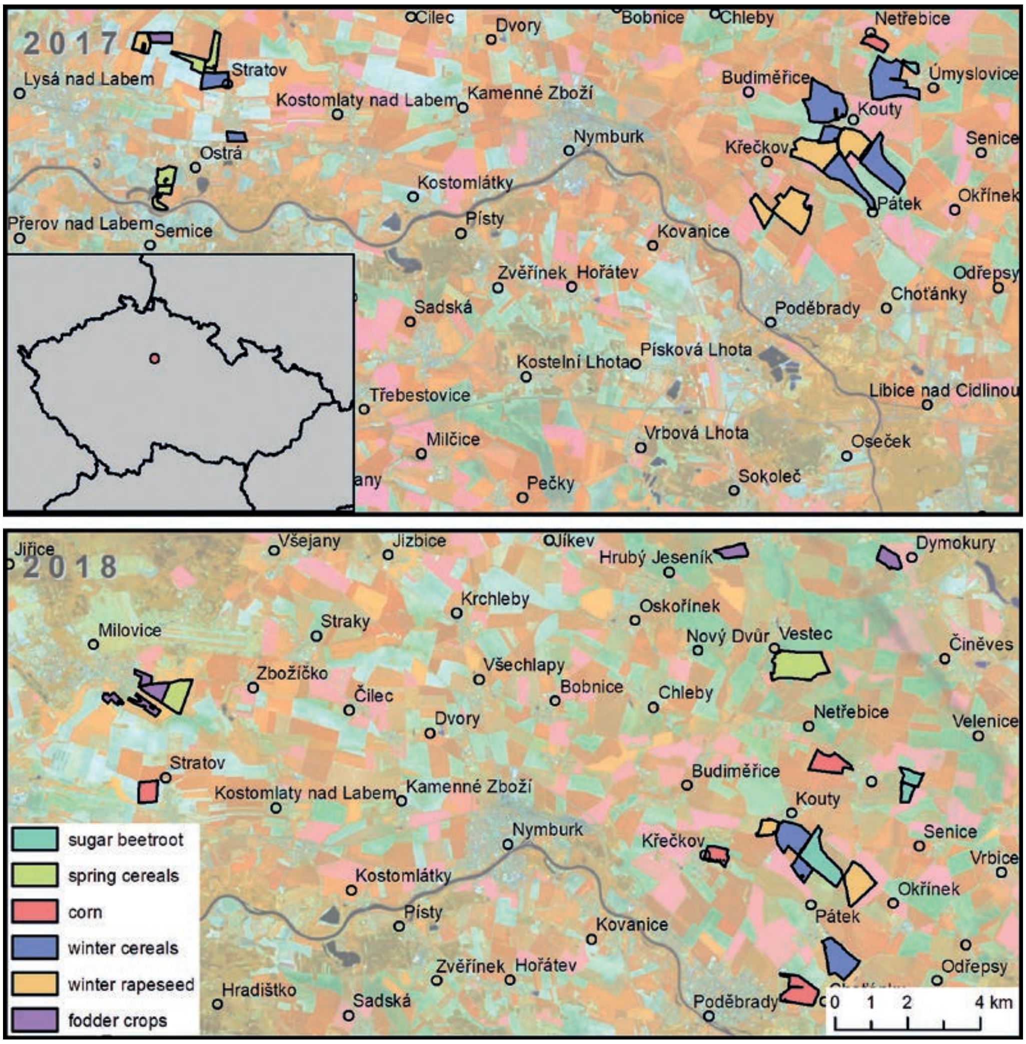

Fig. 2 Agricultural parcels where field campaigns took place in 2017 and 2018.

Tab. 2 Summary on field sampling dates and the reference Landsat scenes.

Date of field sampling	29.–31. 3. 2017	17.–19. 5. 2017	19.–21. 6. 2017	27.–30. 4. 2018	21. 5. 2018	26. 7. 2018
Reference Landsat scene	1. 4. 2017	19. 5. 2017	20. 6. 2017	28. 4. 2018	22. 5. 2018	25. 7. 2018

of field measurements and an empirical parametrization procedure (Tomíček et al. 2021).

We used an artificial neural network (ANN) approach as the regression model for LAI quantitative estimation. Despite its "black-box" nature, this approach provides the ability to implicitly model complex nonlinear relationships between model inputs and outputs (Richter et al. 2012). Using the TensorFlow python library, a feed-forward neural network with one hidden layer was implemented, the widely used rectified linear unit (ReLu) was chosen as the activation function (Wolanin et al. 2019; Xu et al. 2022). To evaluate model performance, the training dataset were divided into calibration (80%) and validation (20%) subsets, and mean squared error was tracked as the loss function within an early stopping mechanism (Tomíček et al. 2021).

The accuracy assessment was performed on cloud-free images with a maximum time delay of 5 days from the collection of ground-truth reference data. For most crops of interest, RMSE was below 1 (except for spring cereals, RMSE = 1.36 and winter rapeseed, RMSE = 2.38) and R^2 was above or equal to 0.7 (except for spring cereals, R^2 = 0.48), Tomíček et al. (2022) for detailed validation results.

3.3 Generation of yearly crop maps

Since changes in phenological and productivity characteristics are monitored specifically for different crops, it was first necessary to know spatial pattern of the crops for the considered growing seasons. Unfortunately, a systematic registry of crop type cultivated on particular agricultural parcels is available only from 2016 onwards; LPIS/GSAA registry data was provided by the State Agricultural Intervention Fund. Therefore, the spatial patterns of the considered crops for the period before 2016 had to be obtained by an alternative way – in this case by supervised classification of Landsat satellite data.

Multitemporal composites were generated first from the source imagery using quarterly time step. The main aim of such temporal aggregation was to cover the entire area of interest by valid data for the given period with no (or at least minimum) occurrence of "nodata gaps" caused by clouds, shadows or snow. The aggregation was based on calculating weighted average of the input reflectance values taking into account 1) spatial distance of the given pixel to the nearest cloud/shadow/snow (the further the pixel is, the higher weight it gets) and 2) temporal distance of the given data acquisition to the mid-date of the used compositing period (scenes acquired closer to the mid-date are preferred over those acquired at the beginning or the end of the period).

Support Vector Machine (SVM) classifier was used for classification of the crop classes: 1) winter cereals, 2) spring cereals, 3) winter rapeseed, 4) fodder crops, 5) sugar beetroot, 6) corn and 7) other crops. The two input parameters of the SVM classifier (C and gamma) were automatically tuned by repeated training to find the best performing configuration. Also, two different kernels (linear and RBF) of the algorithm were considered. The described crop classification was applied under a cropland mask derived from (a) archival LPIS data (available from 2004 onwards) and (b) an internal land cover classification, which accounts for cropland areas not included in the 2004 LPIS dataset. The crop classification procedure then resulted in a thematic raster layer (crop map) and a pixel-based probability layer. The last step was postprocessing, including thematic filtering (pixels with a probability below 70% were reassigned to the "other crops" class) and spatial filtering using a sieve filter (the minimum mapping unit was set at 10 pixels).

Reference data used for training the SVM classifier as well as for validation of the output crop classification maps were obtained by visual interpretation of Landsat images. The visual interpretation was performed using false-colour RGB combination of NIR, SWIR-1 and SWIR-2 bands, which was found to be the most suitable spectral combination for identification of the different crop types. Plots for this visual interpretation were selected randomly across the entire Czechia, with particular attention given to cases where there was high certainty regarding the assigned/interpreted crop type. In addition, independent validation dataset was created as well. However, as the visual interpretation of crop types was highly time demanding, validation data were interpreted only for some years (1986, 1993, 2000, 2002, 2011, 2012). Selection of the reference years took into account two different aspect: 1) quality of the input imagery used for visual interpretation and 2) main phases of the economic development of Czechia (1986: late phase of the socialistic regime, 1993: beginning of economic transformation, restitutions and privatization of agricultural land, 2000 and 2002: preparation for the EU membership, 2011, 2012: EU membership). For each validation year, between 400 and 500 parcels were analysed.

3.4 LAI time series generation

Successful determination of phenological and productivity parameters from remote sensing imagery requires a relatively dense time series of observations. However, for the vast majority of the period of interest

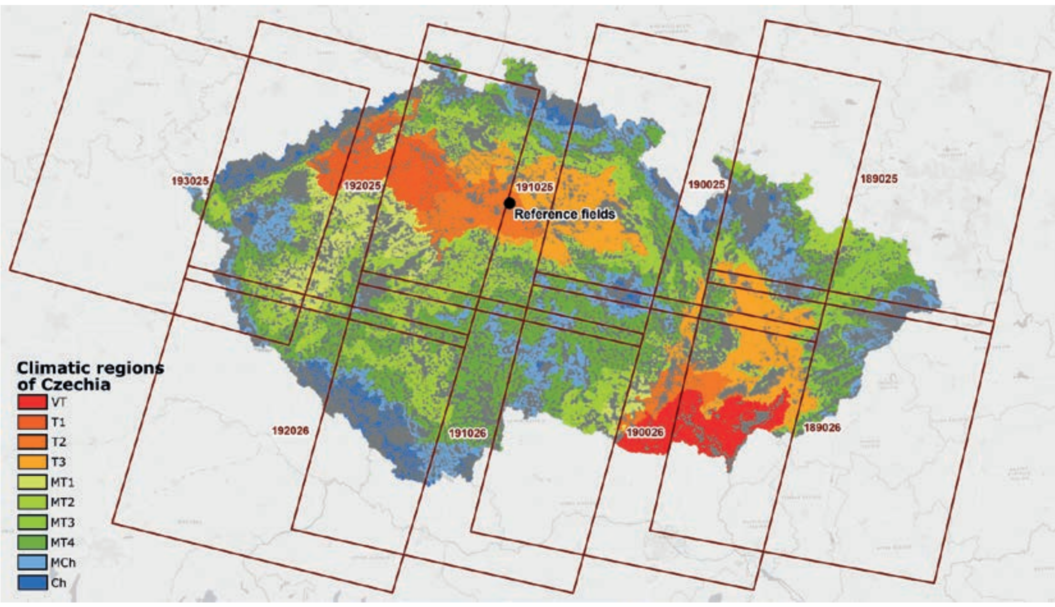

Fig. 3 Landsat tiles covering the territory of Czechia together with agroclimatic regions.

between 1986 and 2020, only Landsat data are available, with a revisit time of 16 days. To overcome this problem, Landsat-derived LAI products were spatially aggregated based on agroclimatic regions (described in section 2), where each of the regions spans two or more Landsat tiles (Fig. 3). The key assumption was that climatic conditions play a fundamental role in the timing of the phenological cycle of vegetation.

Zonal median of LAI was calculated for each single acquisition date of the source imagery and for all existing combination of agroclimatic region and crop. The number of dates for which aggregated LAI values are available is then considerably higher compared to the situation when LAI is considered locally (e.g. on parcel or point level).

3.5 Determination of crop phenological and productivity characteristics

Annual time series of the LAI values were filtered by the Savitzky-Golay filter as a first step to suppress influence of noise in the source data. In the next step, a radial basis function (RBF) was fitted to interpolate the LAI seasonal profile within a 1-day step. A 'threshold approach' was then applied to extract phenological (SOS, EOS, LOS, MAX_DOY) and productivity (SINT, LINT) parameters. The threshold approach uses a certain percentage of the annual LAI amplitude (i.e., the difference between the annual maximum and minimum LAI) as a threshold to determine the timing of phenological phases.

In the case of our study, 25% of the LAI annual amplitude was set as the threshold value. Date, when LAI first reaches such threshold is then considered as the SOS date, whereas EOS date is then considered as the day when LAI drops below the threshold. The date of reaching the annual LAI maximum is MAX_DOY, the period between SOS and EOS is then considered as LOS. Smoothed and interpolated LAI profiles were also used for extraction of vegetation productivity indicators (small integral, SINT and large integral, LINT). Both of these indicators represent area under the LAI temporal curve. The difference is that the LINT takes into account full area under curve (i.e. above LAI = 0), whereas the SINT indicator takes into account only the area under curve above the baseline defined as the annual LAI minimum. The principle of determining phenological and productivity parameters is shown in the diagram in Fig. 4. The described approach is used for example in case of TIMESAT software (Eklundh and Jönsson 2015) which is applied for production of the High Resolution Vegetation Phenology and Productivity (HR-VPP) products under the Copernicus Land Monitoring Service (CLMS). The SINT and LINT indicators show a strong correlation specifically with the SPROD (seasonal productivity) and TPROD (total productivity) parameters produced in the HR-VPP dataset.

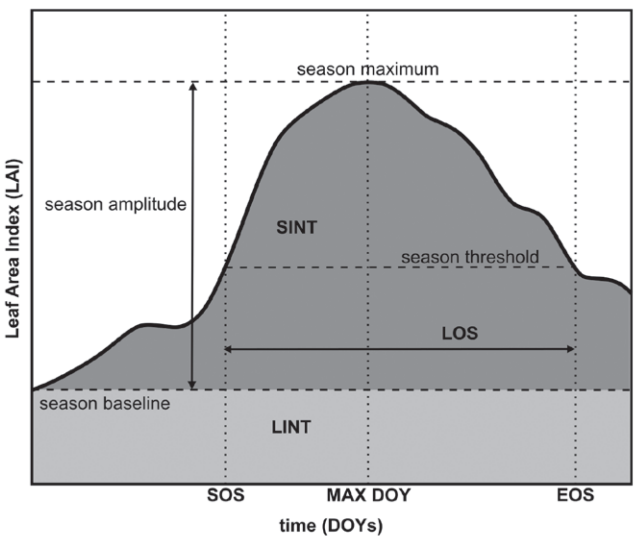

Fig. 4 Diagram illustrating the principle used to determine phenological and productivity parameters.

The 35-year evolution of phenological and productivity parameters was then examined using linear least squares regression. For each combination of a) climatic region, b) crop and c) phenological or productivity parameter, the slope of the regression line was calculated. The statistical significance of the trend was verified using the Wald test with the t-distribution of the test statistic. The Wald statistic results from dividing the regression coefficient by its corresponding standard error; the null hypothesis states that the slope is equal to zero.

In order to compare the influence of the agroclimatic regions, the differences in slope values between regions were calculated for each LSP, productivity parameter and crop individually. To ensure consistency in the sign of the differences, the slope value corresponding to the colder agroclimatic region was always subtracted from that corresponding to the warmer region. A Wilcoxon signed-rank test was used to test the null hypothesis that the median of differences in slope values equals zero for six crops and six parameters (36 cases in total). Rejecting the null hypothesis indicated a systematic shift in the slope of the given combination of parameters and crops across the agroclimatic regions.

4. Results

4.1 Crop classification accuracy assessment

Two raster layers (thematic crop map and probability map) were obtained for each year of the 1986–2015 period as the output of the SVM classification model. These layers were further the subject of quality assessment based on a standardized validation workflow calculating class-related accuracy indicators (user's and producer's accuracy and F-1 score) as well as the overall accuracy. For all validation years, the overall accuracy was higher than 80%. We expect similar accuracy characteristics for the rest of the years since the same crop classification workflow was applied there. The results of the crop classification accuracy assessment are shown in Tab. 3.

4.2 Aggregated LAI time series

The set of six plots in Fig. 5 shows examples of seasonal time series for six crops of interest: 1) winter cereals, 2) spring cereals, 3) winter rapeseed, 4) fodder crops, 5) sugar beetroot and 6) corn derived by spatial aggregation of the original LAI layers within the extent of agro-climatic regions (data for the year 2018 and agro-climatic region MT1 were used here). The original LAI values derived from the spectral information of the Landsat data (hollow circles) were smoothed using the Savitzky-golay filter (solid black dots) and, as a final step, interpolated with a 1-day step using the RBF function (black line). The quality of the created seasonal LAI time series is crucial for the correct determination of the phenological and productivity characteristics of the stand.

4.3 35-year development of phenology and productivity

Fig. 6 shows an example of the 35-year evolution of the six phenological and productivity parameters of interest (SOS, EOS, LOS, MAX_DOY, SINT and LINT) for winter cereals and climate region VT. Values for individual years were fitted with a regression line determined by the method of least squares for trend evaluation. The p-value determines the significance of the observed trend (in this case, the trend is significant for parameters EOS and MAX_DOY).

The magnitude of the trend (slope of the regression line) of the 35-year evolution of the phenological and productivity parameters of interest is visualized in Fig. 7. The exact values together with their significance are then summarized in Appendix 1. Only a few significant trends were documented in the case of the phenological parameters SOS and LOS. Moreover, the variability of slope values was relatively high across crops and climatic regions. However, a significant negative trend occurred in the case of EOS and MAX-DOY phenological parameters in most climatic regions of crops: winter cereals, spring cereals, winter rapeseed and fodder crops, i.e. all crops with the exception of the so-called summer crops (sugar beetroot and corn). Specifically, for EOS, a significant trend was demonstrated for winter cereals in 10, spring cereals in 7, winter rapeseed in 8 and fodder crops in 7 climate regions out of 10; for MAX-DOY there were 10 occurrences for winter cereals, 8 for spring cereals, 7 for winter rapeseed and 6 for fodder crops. The slope of the regression line (the magnitude of the trend) for the parameter MAX-DOY was in the range of −0.4 and −0.7 in 90% of significant cases. In

Tab. 3 Crop classification accuracy assessment metrics (W.C. = winter cereals, S.C. = spring cereals, W.R. = winter rapeseed, F.C. = fodder crops, S.B. = sugar beetroot, C. = corn, OA = overall accuracy, UA = user's accuracy, PA = producer's accuracy, F1 = F-1 score).

Parameter		W.C.	S.C.	W.R.	F.C.	S.B.	C.	OA
1986	UA	86.6	40.7	96.0	92.5	93.4	78.5	80.1 (n = 433)
	PA	56.9	68.6	95.1	86.0	91.9	86.4	
	F1	68.6	51.1	95.6	89.1	92.7	82.3	
1993	UA	96.2	92.3	100.0	100.0	75.9	97.2	93.3 (n = 434)
	PA	100.0	96.0	98.1	100.0	96.9	60.3	
	F1	98.0	94.1	99.0	100.0	85.1	74.5	
2000	UA	100.0	77.3	100.0	100.0	96.4	94.8	95.3 (n = 424)
	PA	91.0	98.1	97.1	93.0	96.4	98.2	
	F1	95.3	86.4	98.5	96.4	96.4	96.5	
2002	UA	96.7	72.9	97.4	100.0	87.8	94.1	92.2 (n = 503)
	PA	87.3	94.4	99.1	93.8	93.5	84.2	
	F1	91.8	82.3	98.2	96.8	90.6	88.9	
2011	UA	93.3	73.8	100.0	100.0	98.2	100.0	94.4 (n = 430)
	PA	95.1	86.5	97.1	100.0	100.0	85.5	
	F1	94.2	79.6	98.5	100.0	99.1	92.2	
2012	UA	75.4	96.7	96.7	97.5	97.4	69.7	85.4 (n = 397)
	PA	98.0	56.9	89.0	92.9	67.9	95.8	
	F1	85.2	71.6	92.7	95.1	80.0	80.7	

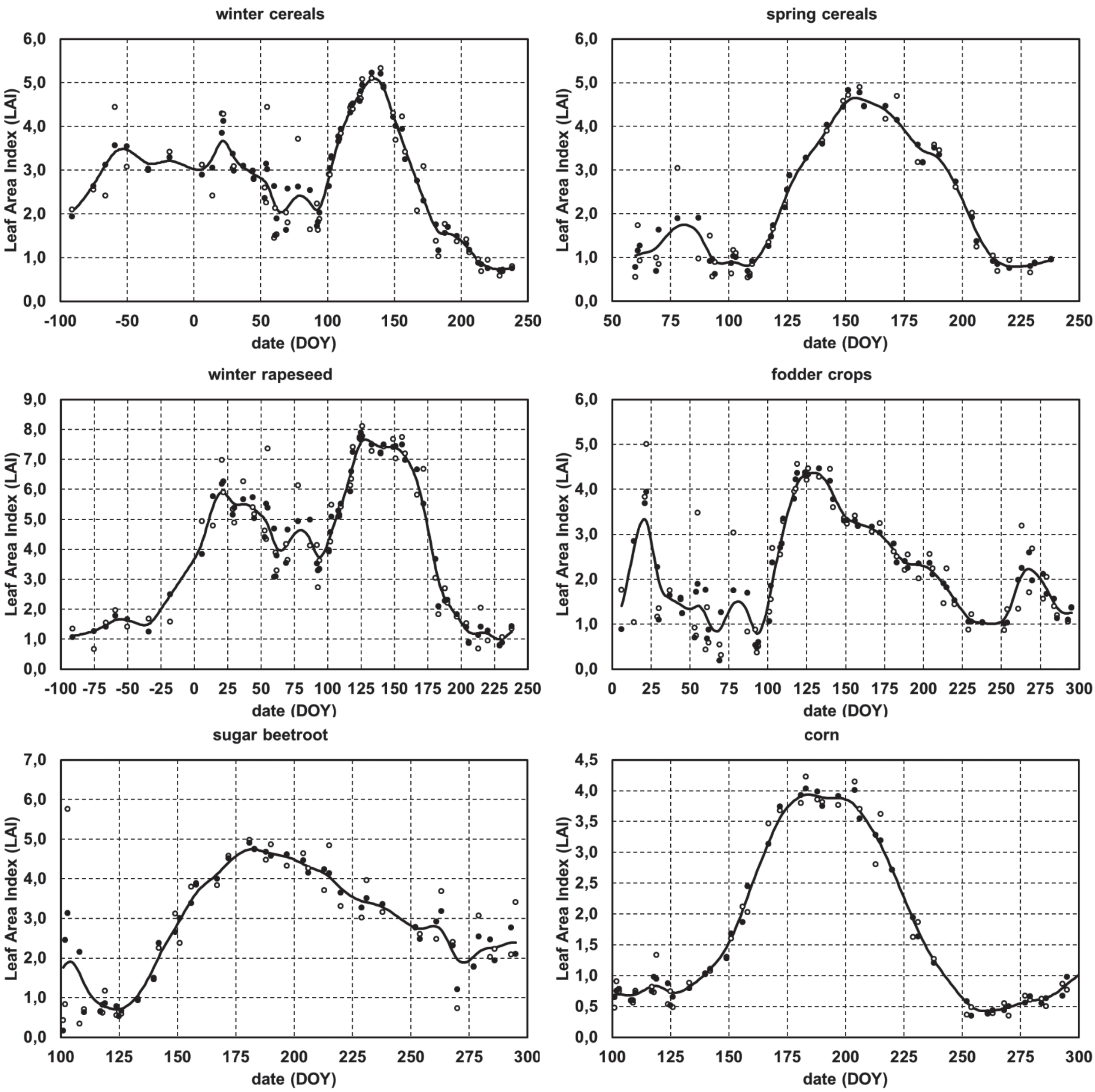

Fig. 5 Example of the aggregated LAI time series for six crops of interest. The original retrieval value is shown by the hollow circles, the solid black dots represent the filtered values, and the values interpolated by the RBF function to a 1-day step are symbolized by the black line.

contrast, slope values varied more between crops in the EOS case. While for winter cereals, spring cereals and winter rapeseed, the slope was between −0.3 and −0.5 in 92% of significant cases, for the fodder crops the slope was between −0.7 and −1.6 in all significant cases.

In the case of productivity parameters, a significant trend in more than half of the climate regions occurred only for spring cereals and sugar beetroot SINT (8 and 6 occurrences out of 10, respectively).

Fig. 8 shows boxplots depicting the mean and dispersion statistics of the differences in LSP and productivity parameter slope values between agroclimatic regions for each crop. The null hypothesis revealed that the median difference was significant in 15 out of 36 cases: six of these were positive and nine were negative. In the majority of cases, the slope values themselves were consistently either positive or negative for a given crop type and observed parameter (Fig. 7 and Appendix 1). Therefore, a positive median difference indicates that the slope value decreases when moving from warmer to colder regions. This implies that shifts in the LSP and productivity parameters were more pronounced in warmer regions during the observed period, provided the slope values were both positive. Conversely, when both slopes were negative, stronger shifts in the observed parameters occurred in the colder regions.

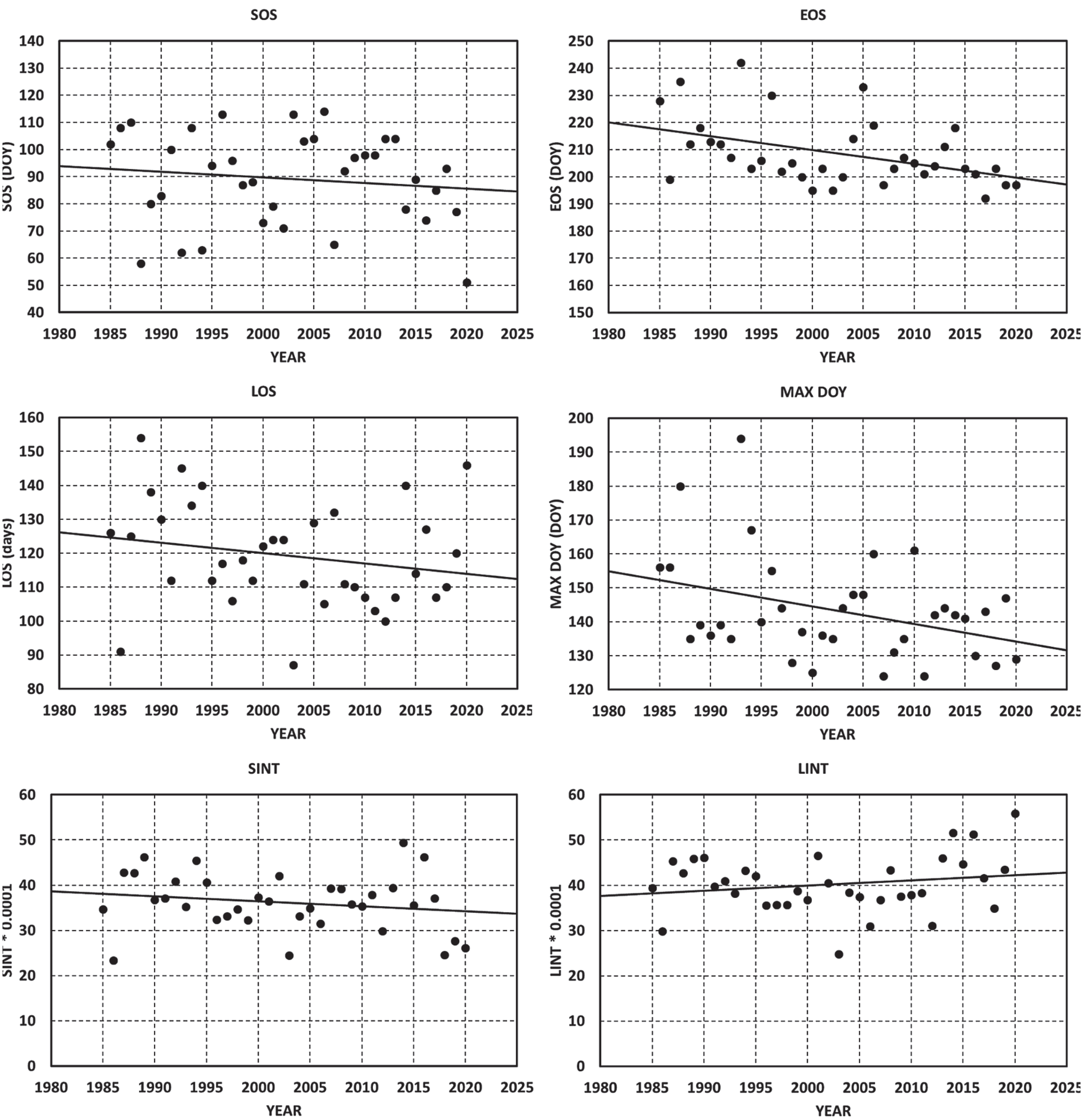

Fig. 6 Example for the 35-year evolution of SOS, EOS, LOS, MAX-DOY, SINT a LINT for winter cereals and climate region VT.

5. Discussion

5.1 Retrieval of LSP parameters

Different vegetation spectral indices such as NDVI or EVI are used for LSP due to their ease of calculation (Misra et al. 2020; Zhang et al. 2023). In accordance with (Lu et al. 2025), the choice of LAI can be supported by its biophysical relevance. By measuring the total leaf area per unit ground area, LAI directly reflects vegetation structure and function. It is closely linked to photosynthesis, transpiration, and carbon fluxes, making it a more meaningful physiological indicator than purely spectral indices. Moreover, LAI is more sensitive than NDVI to canopy development and senescence, particularly in areas with high vegetation. It provides a quantitative measure of vegetation growth stages, which is crucial for modeling ecosystem processes. Furthermore, significant progress has been made in recent years in retrieving LAI based on machine learning and radiative transfer models (Tomíček at al. 2021; Qin et al. 2024).

The accuracy of derived LAI with RMSE = 1.36 for spring cereals, RMSE = 2.38 for winter rapeseed, and RMSE < 1 for other crops (Tomíček et al. 2022) is comparable with other models. (Mourad et al. 2020)

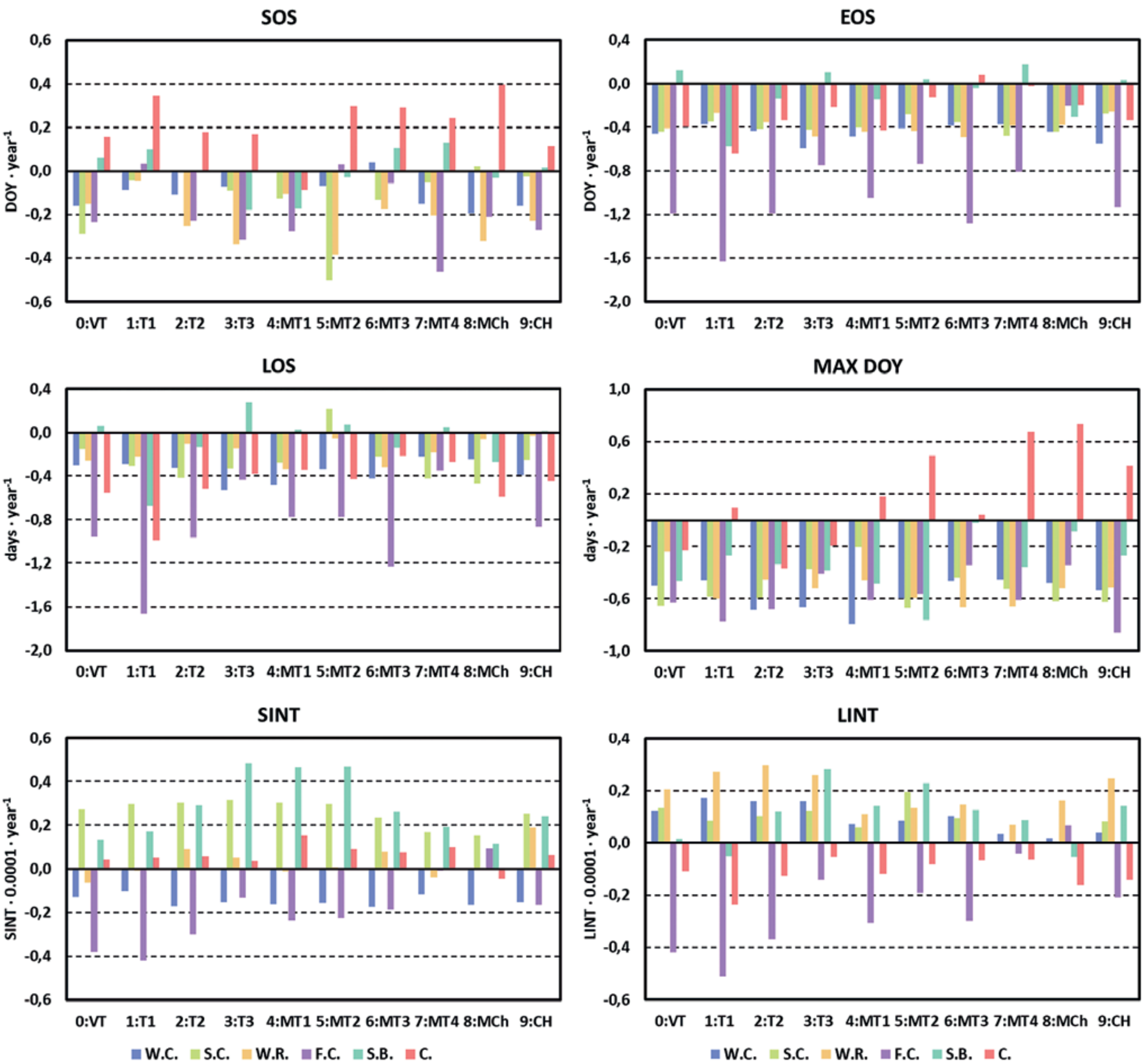

Fig. 7 Slope of individual phenological and productivity parameters derived over all combinations of climatic regions and crops (W.C. = winter cereals, S.C. = spring cereals, W.R. = winter rapeseed, F.C. = fodder crops, S.B. = sugar beetroot, C. = corn).

evaluated empirical models for LAI derivation from NDVI, EVI2, and soil adjusted vegetation index (SAVI) as well as a biophysical model based on ANN embedded to the ESA Sentinel Application Platform (SNAP) software applied on Snetinel-2 and HLS product. After comparison to in-situ measurement, the best models revealed RMSE between 0.65 and 0.89 for corns (barley and wheat). Similarly to our approach, Dhakar et al. (2021) conducted retrieval of wheat LAI by LUT-based inversion of PROSAIL-5B model using atmospherically corrected Landsat-8 OLI reflectance. They achieved a good agreement with the in-situ observed LAI having RMSE of 0.70.

Two main methods are often used to determine Land Surface Phenology (LPS) metrics from satellite data: the threshold method (used for example in Timesat tool; Jönsson and Eklundh 2004) and the derivative method (used for example in HANTS workflow; Zhou et al. 2015). The threshold method determines phenological events (such as SOS or EOS) based on crossing a fixed level of the used variable (e.g. LAI) typically corresponding to a certain percentage of its seasonal maximum or seasonal amplitude. The threshold method is simple, intuitive and easy to implement. It is also more tolerant to presence of moderate noise in the input data if the thresholds are chosen properly. Another advantage can be also seen in its flexibility since the thresholds can be adjusted independently for different types of vegetation. On the contrary, the biggest disadvantages of this method include primarily its sensitivity to the threshold choice when there is generally no exact clue on what percentage of the season maximum (or amplitude) is appropriate to be considered as start/end of

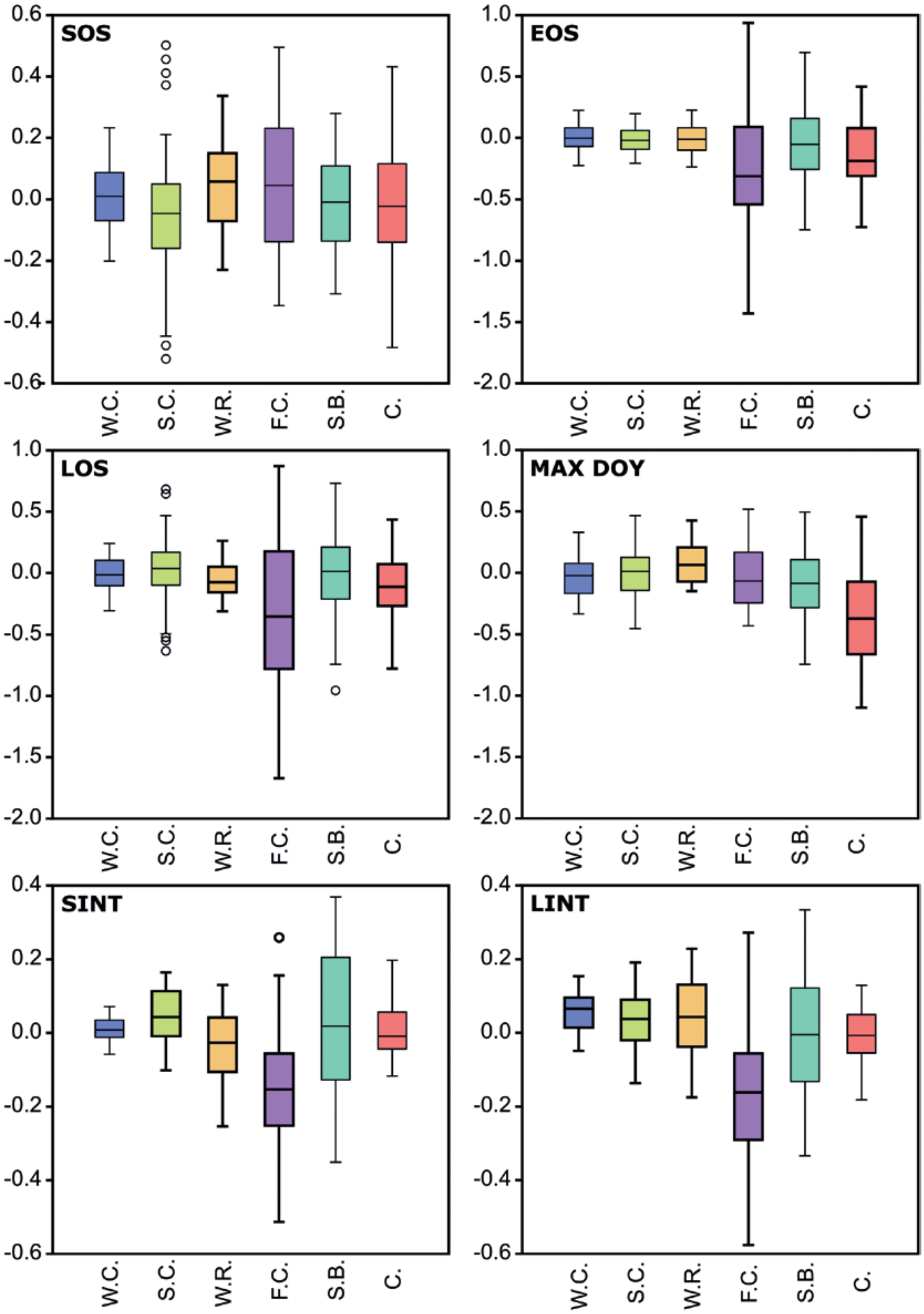

Fig. 8 Statistics of differences in LSP and productivity parameters slope values between agroclimatic regions calculated for each crop. Significant cases are highlighted in bold.

the season (e.g. Huang et al. 2019). In addition, the threshold method is not very suitable for such case typical for flat phenological profiles (i.e. low seasonal amplitudes). It may also miss the full shape and complexity of the phenological curve (this happen especially in cases where multiple growing season are present at the given place). In summary, the threshold method seems to be ideal for rapid applications or in cases when lower-quality of the data is expected on the input. The derivative method identifies phenological events based on the rate of change in the input variable (e.g. LAI) typically by finding the inflection points using first derivative. Thus, there is no need to define any thresholds and the condition for detection of the main phenological events (such as SOS and EOS) are there clearly defined. It is also better capturing full dynamics of the phenological profile and thus it is more suitable in in cases typical for rapid LAI transitions (i.e. as a spring "green-up"). On the other hand, performance of the derivative method is highly sensitive to noise in the input data and thus needs high quality and well-smoothed inputs (e.g. de Beurs and Henebry 2009). From implementation perspective, it is more complex (taking into account all the fitting and smoothing steps before applying derivative) and maybe less intuitive compared to the threshold method.

5.2 Crop types classification and its accuracy

Knowledge of crop types and their spatial distribution was a crucial to the present study. As the information from the LPIS/GSAA registry was unavailable before 2016, supervised classification based on quarterly cloud free mosaics of Landsat imagery was carried out instead. A comparative study performed by (Pluto-Kossakowska 2021) showed that there are no significant differences in accuracy when utilizing different machine learning (ML) algorithms for the multitemporal classification of satellite images for crop and arable land recognition. According to their findings, the ANN classifiers perform just a few percent better than ML. Among ML algorithms, SVM and Random Forest (RF) are commonly used. User's accuracy achieved by our approach based on SVM algorithm is comparable to or outperforms the results collected by (Pluto-Kossakowska 2021) as shown in Tab. 4. (Van Tricht et al. 2023) developed an open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping. Their classification approach is based on decision trees and Landsat and Sentinel-2 imagery. They claim user's (and producer's) accuracy of 94 and 86 (78 and 76)% for cereals and corn, respectively. (Huang et al. 2022) focused on winter cereals in Europe between 2016 and 2020. They combined Landsat and Sentinel-2 imagery with Sentinel-1 SAR data in order to discriminate between winter cereals and winter rapeseed. They implemented a time-weighted dynamic time warping (TWDTW) method, based on the comparison of seasonal changes in NDVI with standard seasonal changes, as well as RF classification, achieving overall accuracies of 91 and 81%, respectively. Specifically, in Czechia, they reached equal user's and producer's accuracies of 87% in discriminating winter cereals. Thus, the classification accuracy based on the SVM algorithm as applied in our study is among the best achieved using comparable methods on Landsat imagery.

5.3 Phenological and productivity trends and their relation to the crop types and climate regions

5.3.1 General trends of LSP and productivity parameters

Over the last 40 years, LSP has undergone changes that vary by climate region and fluctuate over time. Jeong et al. (2011) observed an increase in LOS of temperate vegetation in the Northern Hemisphere between 1982 and 2008. They based their results on analyzing NDVI derived from AVHRR and temperatures. However, the SOS advance of 5.2 and 0.2 days and the EOS delay of 4.3 and 2.3 days differed between the 1982–1999 and 2000–2008 periods, respectively. Specifically in Europe, the delayed EOS of 8.2 days was more significant than the advanced SOS of 3.2 days in the latter period. Global LSP based on AVHRR and MODIS data from 1982 to 2010 was also studied by (Zhang et al. 2014). The seasonal vegetative trajectory was derived from daily EVI across Köppen's climate regions. The analysis showed that SOS generally shifted early in temperate, cold and polar climates in the Northern Hemisphere. However, areas with a significantly earlier SOS decreased in number between 2000 and 2010 compared to the period between 1982 and 1999, and LOS also increased. Notably, the overall trends in Europe were generally insignificant.

In our study, we calculated and analyzed trends in the six LSP and productivity characteristics of six crops over ten agroclimatic regions of Czechia between 1986 and 2020. Overall, there was no significant trend of SOS for the studied agricultural crops except for winter rapeseed in regions T3 and MCh. Nevertheless, there was a general trend towards earlier SOS for most crops except for sugar beetroot and corn. On average, the shift was 1.1 day per decade for winter and spring cereals, and 2.1 day for winter rapeseed and fodder crops. Similar SOS behavior of rainfed and irrigated crops was observed in Spain (Michavila et al. 2024). The later occurrence of SOS of corn, which has a later emergence, is in accordance with study carried out the Midwest of the United States (Zhang et al. 2019). The EOS trend was significant in a greater number of crops and agroclimatic regions, with an average shift of 5.6 days per decade for the four most significant crops. Contrary to the findings of (Jeong et al. 2011) but in accordance with those of (Michavila et

Tab. 4 Average user's accuracy (UA) achieved by SVM classification of Landsat images compared to accuracies reported for the same crops by (Pluto-Kossakowska, 2021). In the case of rapeseed and sugar beetroot only RF classification results were available in the reference literature.

	Winter cereals	Spring cereals	Rapeseed	Sugar beetroot	Corn
UA our study %	91	76	98	92	89
UA P-K %	72	53	96	79	90

al. 2024), this shift is negative, i.e. towards an earlier DOY. It should be noted that both Jeong et al. (2011) and Zhang et al. (2014) focused on vegetation in general. The LSP parameters of agricultural crops are specific because, in addition to natural processes such as temperature, day length or precipitation, they are also determined by human management decisions such as the actual planting and harvest days, irrigation, fertilization and pests and weed control (Bartošová et al. 2025). Moreover, in the case of agricultural crops, the EOS does not represent the biological termination of the vegetation curve, as it does for natural vegetation. Instead, the senescence phase of the curve is artificially shortened by harvesting, with the harvest date primarily depending on the timing of crop maturity. Therefore, they are not fully comparable. Even when only agricultural plots are considered, the coarser resolution of MODIS or AVHRR causes slight changes to SOS and EOS trends due to the surrounding natural vegetation mixing with crops in the fragmented agricultural landscape (Sisheber et al. 2023).

In general, the significance of the LOS trend was low in our study. On average, it tended towards a shortening of 3.5 days per decade which is again contrary to the findings of Jeong et al. (2011). The shortening of LOS is mainly caused by the advance of EOS. The LAI maximum (MAX_DOY) trends show similar patterns of significance to those of EOS (Appendix 1). Also, its average advance of 5.5 days per decade for the four most significant crops is almost equal to the shift in EOS. On the other hand, the negative trends of the four LSP parameters discussed are in accordance with predictions of changes in the time of sowing, flowering and maturity of cereals in Europe due to climate change (Olesen et al. 2012)but also affected by day length and potential physiological stresses. Responses may vary between species and varieties. Climate change will affect the timing of cereal crop development, but exact changes will also depend on changes in varieties as affected by plant breeding and variety choices. This study aimed to assess changes in timing of major phenological stages of cereal crops in Northern and Central Europe under climate change. Records on dates of sowing, flowering, and maturity of wheat, oats and maize were collected from field experiments conducted during the period 1985–2009. Data for spring wheat and spring oats covered latitudes from 46 to 64°N, winter wheat from 46 to 61°N, and maize from 47 to 58°N. The number of observations (site-year-variety combinations, given an average temperature increase of 0.35°C per decade in Czechia (Crhová et al. 2022.

Examining crop types, most significant trends in LSP parameters relate to winter and spring cereals, winter rapeseed, and fodder crops. All these crops show a trend of EOS and MAX_DOY advancement but they do not differ considerably across the agroclimatic regions except for fodder crops that exhibit the most pronounced trend, as well as greatest variability in EOS (Fig. 7). Fodder crops are harvested at least twice during the season (Springer and Aiken, 2015)theoretical ethanol yield, crude protein (CP. The timing and frequency are determined by natural conditions, such as soil type, temperature and precipitation. However, they mainly depend on whether the crop is harvested for forage or seed. Thus, increasing temperatures might explain the advance in harvest and different management practices the higher variability.

The productivity parameters vary in terms of their signs and values. A significant increase in the SINT can be observed for spring cereals and for sugar beetroot in 9 and 6 of the agroclimatic regions, respectively. An opposite trend, though significant only in two warmest regions, is evident for fodder crops and winter cereals (with no significant cases). For the LINT parameter, significant trends appear mostly for non-cereals in warmer regions. Again, LINT trends are not significant for winter cereals, but they have an opposite sign, which can be interpreted as an increase in total LAI while the area above the seasonal baseline was decreasing (Fig. 4).

5.3.2 Relation of LSP and production parameters to agroclimatic regions

Evaluating the differences in the slopes of the LSP parameters between the agroclimatic regions revealed a significant dependence in one-third of the 24 cases (i.e. four parameters times six crops), but this did not apply to cereals. The significant positive median slope differences in SOS and MAX_DOY for winter rapeseed indicate that the advance of SOS and MAX_DOY was higher in the colder regions, of around 0.5 days per decade for both parameters. The other significant shifts (fodder crops – EOS and LOS; corn – EOS, LOS, and MAX_DOY; and winter rapeseed – LOS) exhibited negative median differences. These observations suggest that during the observed period, the advancement of the harvesting of fodder crops and corn and the shortening of the production season were larger in warmer regions than in colder ones. When interpreting the negative median difference in the MAX_DOY parameter for corn, the sign of the slope must be considered. While it shows slight advancement in warmer regions, there is a trend towards later DOY in colder regions. This is consistent with the idea of using colder regions to grow crops that were previously only suitable for warmer regions, as discussed in relation to northern Europe, for example (Unc et al. 2021).

Bartošová et al. (2025) examined the relationship between elevation and the phenological phases of winter wheat. They analyzed in-situ observations of registered winter wheat cultivars at 17 experimental stations in Czechia between 1961 and 2021. The three stages evaluated were: i) jointing (first node at least 1 cm above the node); ii) heading (beginning of heading); and iii) ripening (fully ripe, hardened grains). The observations were grouped into

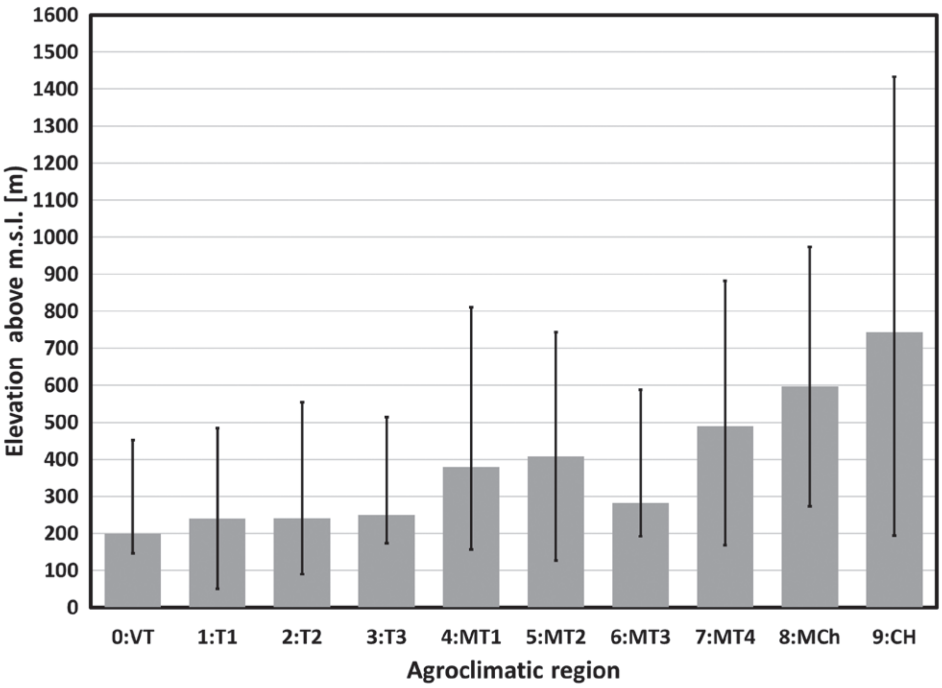

Fig. 9 Median, minimum, and maximum elevation calculated within the agroclimatic regions.

three elevation categories: 0–299 m, 300–499 m, and 500–750 m above mean sea level (a.m.s.l.). The agroclimatic regions used in our study cover a variety of elevations, as can be seen in Fig. 9.

For purposes of comparison, we aggregated the agroclimatic regions and used the most suited slope values for winter cereals, since winter wheat dominates this category. If the SOS is neglected, as it did not demonstrate a significant trend in our study, the MAX_DOY parameter exhibits a steeper negative slope of approximately 2.5 days per decade in the two lower elevation intervals and 1 day per decade in the highest interval, as shown in Table 5. The EOS shows a good fit with ripening, except in the high elevation interval where it indicates an advance to an earlier DOY of 1.7 days per decade. Despite these slight differences, both types of observation indicate the same trend direction and magnitude. The reasons for these differences are: i) the parameters are defined differently, although they are closely related; ii) winter cereals also include other crops (e.g. winter barley and winter rye); iii) the start and therefore the length of the time series differ (1961–2021 vs. 1986–2020). Nevertheless, the overlap between the two studies is the greatest among existing literature in terms of research objectives, covered territory and crop type.

Regarding the productivity parameters, the significant positive median slope differences in SINT for spring cereals and in LINT for winter and spring cereals indicate that these parameters increased more in warmer regions than in colder regions during the observed period. Conversely, the significant

Tab. 5 Comparison of trends in selected phenological parameters of winter cereals (winter wheat) in relation to elevation above mean sea level (a.m.s.l.). All parameters exhibited a significant trend, except SOS (in italics).

Elevation interval a.m.s.l. AGC regions	Results Bartošová et al. (2025)		LSP observations	
		Days per decade		Days per decade
0–299 m VT, T1–T3, MT3	jointing	−4.5	*SOS*	*−0.8*
	heading	−3.3	MAX_DOY	−5.6
	ripening	−4.4	EOS	−4.5
300–499 m MT1, MT2, MT4	jointing	−6.9	*SOS*	*−0.8*
	heading	−3.6	MAX_DOY	−6.2
	ripening	−4.8	EOS	−4.2
> 500 m MCH, CH	jointing	−5.7	*SOS*	*−1.8*
	heading	−4.0	MAX_DOY	−5.1
	ripening	−3.3	EOS	−5.0

negative median for both SINT and LINT for fodder crops shows that productivity loss is smaller in colder regions during the observed period (as the slope is negative in the majority of agroclimatic regions).

Our first hypothesis was formulated as follows: "There are significant trends in LSP and productivity characteristics over the observed period." The results of the present analysis partially confirmed this. EOS and MAX_DOY were the only two parameters with a significant negative trend; however, this was only observed for cereals, rapeseed and fodder crops. Our study thus confirmed the existence of LSP trends and their direction over the observed 35-year period. At the same time, however, it showed that despite a significant trend in some LSP characteristics, the productivity parameters revealed negligible significance. The LSP of agricultural crops differs and should be considered in regional studies on crop productivity (Lobell and Gourdji 2012), land use and land cover changes (Zhang et al. 2019), and the impacts of global change on agriculture (Brown et al. 2012; Yuan et al. 2024) and vice LSP feedback to global change (Liu et al. 2017).

The results of the discussion also proved the validity of the second hypothesis: "Using high spatial resolution Landsat data and knowledge of the distribution of specific crops will enable us to observe differences in LSP and productivity trends between agroclimatic regions". Significant differences in trends in the LSP and productivity parameters were observed between the agroclimatic regions in 15 out of the 36 studied cases.

6. Conclusions

This study demonstrates the potential of long-term, high-resolution optical satellite data series for monitoring and analyzing trends in crop phenology and productivity at a regional scale. Using a 35-year Landsat time series, we evaluated four key phenological (SOS, EOS, LOS, MAX_DOY) and two productivity (SINT, LINT) parameters for six major crop types (winter cereals, spring cereals, winter rapeseed, fodder crops, sugar beetroot, and corn) in ten agro-climatic regions in Czechia. The methodology combined robust ground-based LAI measurements, advanced radiative transfer modeling (PROSAIL), machine learning-based LAI retrieval, and supervised crop classification to calculate dense LAI time series for each climatic region and crop. The annual LAI time series were then smoothed using a Savitzky-Golay filter and interpolated with RBF to produce daily profiles, from which phenological (SOS, EOS, LOS, MAX_DOY) and productivity (SINT, LINT) parameters were extracted using a threshold approach based on 25% of the annual LAI amplitude.

Accurate knowledge of crop types and their spatial distribution was essential for this study. Supervised classification of Landsat images using the SVM algorithm provided results with an overall accuracy of higher than 80% for all validated years. The achieved accuracy is comparable or better than other approaches in recent regional studies based on Landsat data. The phenology and productivity parameters in Czechia have shown different trends over the last four decades across climatic regions and crop types. While global vegetation studies often find a lengthening of the growing season due to earlier spring and later autumn, our analysis of six key crop types between 1986 and 2020 in Czechia tends to show only a significant negative shift of EOS and MAX-DOY – especially for winter and spring cereals, winter rapeseed (up to −0.7 days/year) and fodder crops (up to −1.6 days/year). For agricultural crops in general, EOS reflects the harvest date rather than natural senescence and is thus strongly related to MAX-DOY because the growing season is artificially terminated by harvest as the crop matures. Productivity trends varied by crop and climatic region, with SINT increasing significantly for spring cereals and sugar beetroot, while significant LINT trends were observed mainly for non-cereals in warmer regions. Approximately 40% of the crop and phenological/productivity parameter combinations show significant differences in trends between agro-climatic regions, with the shift in SOS and MAX_DOY for winter rapeseed occurring more rapidly in colder regions, while the shift in harvest dates and shortening of the season for corn and fodder crops is more pronounced in warmer regions. Comparison with phenological study based on altitude (Bartošová et al. 2025) confirmed similar directions and magnitudes of trends for winter cereals. Productivity parameters (SINT and LINT) for cereals increased more in warmer regions, while losses in fodder crops productivity were less pronounced in colder regions.

The study was based entirely on satellite data, and, to the authors' knowledge, it is unique in the level of detail of the performed analysis. In order to elaborate on differences in LSP and productivity trends between agroclimatic regions, detailed meteorological observations such as temperature and precipitation on a monthly or quarterly basis are necessary.

Acknowledgments

We would like to thank the Charles University Grant Agency for the support of this project in the form of grant number 293121: "Long-term development of phenological parameters of agricultural crops based on a 35-year time series of Landsat multispectral data". In addition, we would like to thank Gisat, Ltd. for providing the data and computing resources needed to successfully complete the project. We also appreciate the valuable comments and recommendations provided by the reviewers, which helped improve the quality of this manuscript.

Appendix 1 Trend slope values for individual phenological and productivity parameters derived over all combinations of climatic regions and crops of interest (W.C. = winter cereals, S.C. = spring cereals, W.R. = winter rapeseed, F.C. = fodder crops, S.B. = sugar beetroot, C. = corn). The slopes for SINT and LINT were calculated for values scaled by 0.0001.

	Crop	0:VT	1:T1	2:T2	3:T3	4:MT1	5:MT2	6:MT3	7:MT4	8:MT	9:MCH
Phenological parameters											
SOS	W.C.	−0.200	−0.100	−0.100	−0.100	−0.010	−0.068	0.040	−0.150	−0.190	−0.160
	S.C.	−0.300	−0.000	0.000	−0.100	−0.130	−0.500	−0.132	−0.053	0.020	−0.024
	W.R.	−0.200	−0.000	−0.300	**−0.300**	−0.100	−0.384	−0.175	−0.203	**−0.320**	−0.228
	F.C.	−0.200	0.030	−0.200	−0.300	−0.280	0.032	−0.056	−0.462	−0.210	−0.272
	S.B.	0.060	0.100	0.000	−0.200	−0.170	−0.029	0.104	0.128	−0.030	0.016
	C.	0.160	0.340	0.180	0.170	−0.090	0.298	0.293	0.242	0.396	0.114
EOS	W.C.	**−0.460**	**−0.370**	**−0.430**	**−0.600**	**−0.480**	**−0.410**	**−0.380**	**−0.370**	**−0.440**	**−0.550**
	S.C.	**−0.440**	**−0.350**	**−0.410**	−0.420	**−0.400**	−0.280	−0.350	**−0.480**	**−0.440**	−0.270
	W.R.	**−0.410**	−0.270	**−0.350**	**−0.480**	**−0.440**	**−0.440**	**−0.490**	**−0.380**	**−0.380**	−0.250
	F.C.	**−1.190**	**−1.630**	**−1.190**	**−0.750**	−1.050	−0.740	**−1.280**	**−0.810**	−0.200	**−1.140**
	S.B.	0.120	**−0.570**	−0.140	0.100	−0.150	0.040	−0.040	0.170	−0.300	0.030
	C.	−0.400	**−0.650**	−0.340	−0.210	−0.430	−0.130	0.080	−0.020	−0.190	−0.340
LOS	W.C.	−0.300	−0.300	−0.300	−0.500	−0.480	−0.338	−0.422	−0.220	−0.240	−0.392
	S.C.	−0.200	−0.300	−0.400	−0.300	−0.270	0.220	−0.220	−0.425	−0.460	−0.248
	W.R.	−0.300	−0.200	−0.100	−0.100	−0.340	−0.053	−0.316	−0.180	−0.060	−0.026
	F.C.	**−1.000**	**−1.700**	**−1.000**	−0.400	−0.770	−0.770	**−1.227**	−0.346	0.008	**−0.865**
	S.B.	0.060	**−0.700**	−0.100	0.280	0.025	0.069	−0.142	0.045	−0.270	0.017
	C.	−0.600	**−1.000**	−0.500	−0.400	−0.340	−0.425	−0.212	−0.266	−0.590	−0.450
MAX-DOY	W.C.	**−0.500**	**−0.500**	**−0.700**	**−0.700**	**−0.790**	**−0.599**	**−0.466**	**−0.457**	**−0.480**	**−0.536**
	S.C.	**−0.700**	**−0.600**	**−0.600**	**−0.400**	−0.200	**−0.669**	−0.441	**−0.525**	**−0.620**	−0.624
	W.R.	−0.200	**−0.600**	−0.500	**−0.500**	−0.460	**−0.595**	**−0.664**	**−0.659**	**−0.520**	**−0.515**
	F.C.	**−0.600**	**−0.800**	**−0.700**	**−0.400**	−0.610	**−0.563**	**−0.342**	−0.610	−0.340	−0.860
	S.B.	−0.500	−0.300	−0.300	−0.400	−0.480	−0.765	−0.019	−0.357	−0.090	−0.272
	C.	−0.200	0.090	−0.400	−0.200	0.181	0.491	0.044	**0.677**	0.732	0.414
Productivity parameters											
SINT	W.C.	−0.129	−0.103	−0.169	−0.151	−0.160	−0.156	−0.174	−0.117	−0.163	−0.152
	S.C.	**0.273**	**0.295**	**0.304**	**0.316**	**0.304**	**0.297**	**0.236**	0.167	0.152	**0.254**
	W.R.	−0.064	−0.006	0.091	0.050	−0.012	0.008	0.077	−0.039	−0.003	0.190
	F.C.	**−0.379**	**−0.419**	−0.299	−0.131	−0.236	−0.226	−0.185	−0.008	0.094	−0.165
	S.B.	0.133	0.170	**0.290**	**0.484**	**0.465**	**0.469**	0.262	0.191	0.115	**0.241**
	C.	0.041	0.050	0.056	0.036	0.153	0.092	0.077	0.101	−0.044	0.063
LINT	W.C.	0.123	0.172	0.101	0.160	0.158	0.071	0.084	0.102	0.033	0.018
	S.C.	0.133	0.083	0.060	0.103	0.123	0.059	0.196	0.095	0.005	0.006
	W.R.	0.205	0.272	0.246	**0.297**	**0.259**	0.110	0.133	0.145	0.070	0.162
	F.C.	**−0.419**	**−0.511**	**−0.361**	**−0.369**	−0.141	−0.306	−0.191	−0.299	−0.042	0.065
	S.B.	0.016	−0.053	0.078	0.120	**0.281**	0.142	0.229	0.125	0.087	−0.053
	C.	−0.108	**−0.237**	−0.136	−0.125	−0.054	−0.118	−0.082	−0.068	−0.066	−0.163

References

ARCSI GitHub. Available online: https://github.com/remotesensinginfo/arcsi (accessed on 20 June 2025).

Bartošová, L., Hájková, L., Pohanková, E., Možný, M., Balek, J., Zahradníček, P., Štěpánek, P., Dížková, P., Trnka, M., Žalud, Z. (2025): Differences in phenological term changes in field crops and wild plants – do they have the same response to climate change in Central Europe? International Journal of Biometeorology 69, 659–670, https://doi.org/10.1007/s00484-024-02846-8.

Brown, M. E., de Beurs, K. M., Marshall, M. (2012): Global phenological response to climate change in crop areas using satellite remote sensing of vegetation, humidity and temperature over 26 years. Remote Sensing of Environment 126, 174–183, https://doi.org/10.1016/j.rse.2012.08.009.

Campioli, M., Marchand, L. J., Zahnd, C., Zuccarini, P., McCormack, M. L., Landuyt, D., Lorer, E., Delpierre, N., Gričar, J., Vitasse, Y. (2025): Environmental Sensitivity and Impact of Climate Change on leaf-, wood- and root Phenology for the Overstory and Understory of Temperate Deciduous Forests. Current Forestry Reports 11: 1, https://doi.org/10.1007/s40725-024-00233-5.

Caparros-Santiago, J. A., Rodriguez-Galiano, V., Dash, J. (2021): Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review. ISPRS Journal of Photogrammetry and Remote Sensing 171, 330–347, https://doi.org/10.1016/j.isprsjprs.2020.11.019.

Chaves, E. D., Picoli, M. C. A., Sanches, I. D. (2020): Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. Remote Sensing 12(18): 3062, https://doi.org/10.3390/rs12183062.

Chmielewski, F.-M. (2013): Phenology in Agriculture and Horticulture. In Schwartz, M. (Ed.), Phenology: An Integrative Environmental Science. Dordrecht: Springer, 539–561, https://doi.org/10.1007/978-94-007-6925-0_29.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., Justice, C. (2018): The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sensing of Environment 219, 145–161, https://doi.org/10.1016/j.rse.2018.09.002.

Crhová, L., Kliegrová, S., Lipina, P., Tolasz, R., Valeriánová, A. (2022): Klimatologická ročenka České republiky 2021. Praha: Český hydrometeorologický ústav. Available online: https://info.chmi.cz/rocenka/meteo2021/ (accessed on 20 March 2025, in Czech).

de Beurs, K. M., Henebry, G. M. (2010): Spatio-Temporal Statistical Methods for Modelling Land Surface Phenology. In Hudson, I., Keatley, M. (Eds.), Phenological Research. Dordrecht: Springer, https://doi.org/10.1007/978-90-481-3335-2_9.

Decree No. 327/1998. Ministry of Agriculture of the Czech Republic. Available online: https://faolex.fao.org/docs/pdf/cze124894.pdf (accessed on 20 June 2025).

Dhakar, R., Sehgal, V. K., Chakraborty, D., Sahoo, R. N., Mukherjee, J. (2021): Field scale wheat LAI retrieval from multispectral Sentinel 2A-MSI and LandSat 8-OLI imagery: effect of atmospheric correction, image resolutions and inversion techniques. Geocarto International 36(18), 2044–2064, https://doi.org/10.1080/10106049.2019.1687591.

Dhillon, M. S., Dahms, T., Kübert-Flock, C., Liepa, A., Rummler, T., Arnault, J., Steffan-Dewenter, I., Ullmann, T. (2023a): Impact of STARFM on Crop Yield Predictions: Fusing MODIS with Landsat 5, 7, and 8 NDVIs in Bavaria Germany. Remote Sensing 15(6): 1651, https://doi.org/10.3390/rs15061651.

Dhillon, M. S., Kübert-Flock, C., Dahms, T., Rummler, T., Arnault, J., Steffan-Dewenter, I., Ullmann, T. (2023b): Evaluation of MODIS, Landsat 8 and Sentinel-2 Data for Accurate Crop Yield Predictions: A Case Study Using STARFM NDVI in Bavaria, Germany. Remote Sensing 15(7): 1830, https://doi.org/10.3390/rs15071830.

Dong, T., Liu, Jiangui, Qian, B., He, L., Liu, Jane, Wang, R., Jing, Q., Champagne, C., McNairn, H., Powers, J., Shi, Y., Chen, J. M., Shang, J. (2020): Estimating crop biomass using leaf area index derived from Landsat 8 and Sentinel-2 data. ISPRS Journal of Photogrammetry and Remote Sensing 168, 236–250, https://doi.org/10.1016/j.isprsjprs.2020.08.003.

dos Santos Luciano, A. C., Picoli, M. C. A., Duft, D. G., Rocha, J. V., Leal, M. R. L. V., le Maire, G. (2021): Empirical model for forecasting sugarcane yield on a local scale in Brazil using Landsat imagery and random forest algorithm. Computers and Electronics in Agriculture 184: 106063, https://doi.org/10.1016/j.compag.2021.106063.

Eklundh, L., Jönsson, P. (2015): TIMESAT: A Software Package for Time-Series Processing and Assessment of Vegetation Dynamics. In Kuenzer, C., Dech, S., Wagner, W. (Eds.), Remote Sensing Time Series. Remote Sensing and Digital Image Processing 22. Cham: Springer, https://doi.org/10.1007/978-3-319-15967-6_7.

Eklundh, L., Jönsson, P. (2016): TIMESAT for Processing Time-Series Data from Satellite Sensors for Land Surface Monitoring. In Ban, Y. (Ed.), Multitemporal Remote Sensing. Remote Sensing and Digital Image Processing 20. Cham: Springer, https://doi.org/10.1007/978-3-319-47037-5_9.

Fitchett, J. M., Grab, S. W., Thompson, D. I. (2015): Plant phenology and climate change: Progress in methodological approaches and application. Progress in Physical Geography 39(4), 460–482, https://doi.org/10.1177/0309133315578940.

Gao, F., Zhang, X. (2021): Mapping Crop Phenology in Near Real-Time Using Satellite Remote Sensing: Challenges and Opportunities. Journal of Remote Sensing 2021: 8379391, https://doi.org/10.34133/2021/8379391.

Gašparović, M., Pilaš, I., Radočaj, D., Dobrinić, D. (2024): Monitoring and Prediction of Land Surface Phenology Using Satellite Earth Observations – A Brief Review. Applied Sciences 14(24): 12020, https://doi.org/10.3390/app142412020.

Gitelson, A. A., Peng, Y., Masek, J. G., Rundquist, D. C., Verma, S., Suyker, A., Baker, J. M., Hatfield, J. L., Meyers, T. (2012): Remote estimation of crop gross primary production with Landsat data. Remote Sensing of Environment 121, 404–414, https://doi.org/10.1016/j.rse.2012.02.017.

Hajkova, L., Nekovar, J., Richterova, D., Koznarova, V., Sulovska, S., Vavra, A., Vozenilek, V. (2012): Phenological Observation in the Czech Republic – History and Present. Phenology and Climate Change. InTech, 71–100, https://doi.org/10.5772/35256.

Hanes, J. M., Liang, L., Morisette, J. T. (2014): Land Surface Phenology. In Hanes, J. M. (Ed.), Biophysical Applications of Satellite Remote Sensing. Springer Remote Sensing/Photogrammetry. Berlin, Heidelberg: Springer, https://doi.org/10.1007/978-3-642-25047-7_4.

Hassan, T., Gulzar, R., Hamid, M., Ahmad, R., Waza, S. A., Khuroo, A. A. (2023): Plant phenology shifts under climate warming: a systematic review of recent scientific literature. Environmental Monitoring and Assessment 196: 36, https://doi.org/10.1007/s10661-023-12190-w.

HR-VPP: User Manual. Available online: https://land.copernicus.eu/en/technical-library/product-user-manual-of-seasonal-trajectories/ (accessed on 20 June 2025).

Htitiou, A., Möller, M., Riedel, T., Beyer, F., Gerighausen, H. (2024): Towards Optimising the Derivation of Phenological Phases of Different Crop Types over Germany Using Satellite Image Time Series. Remote Sensing 16(17): 3183, https://doi.org/10.3390/rs16173183.

Huang, X., Fu, Y., Wang, J., Dong, J., Zheng, Y., Pan, B., Skakun, S., Yuan, W. (2022): High-Resolution Mapping of Winter Cereals in Europe by Time Series Landsat and Sentinel Images for 2016–2020. Remote Sensing 14(9): 2120, https://doi.org/10.3390/rs14092120.

Huang, X., Liu, J., Zhu, W., Atzberger, C., Liu, Q. (2019): The optimal threshold and vegetation index time series for retrieving crop phenology based on a modified dynamic threshold method. Remote Sensing 11(23): 2725, https://doi.org/10.3390/rs11232725.

Jeong, S.-J., Ho, C.-H., Gim, H.-J., Brown, M. E. (2011): Phenology shifts at start vs. end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982–2008. Global Change Biology 17(7), 2385–2399, https://doi.org/10.1111/j.1365-2486.2011.02397.x.

Jönsson, P., Eklundh, L. (2004): TIMESAT – a program for analyzing time-series of satellite sensor data. Computers & Geosciences 30(8), 833–845, https://doi.org/10.1016/j.cageo.2004.05.006.

Kaspar, F., Zimmermann, K., Polte-Rudolf, C. (2014): An overview of the phenological observation network and the phenological database of Germany's national meteorological service (Deutscher Wetterdienst). Advances in Science and Research 11, 93–99, https://doi.org/10.5194/asr-11-93-2014.

Koch, E., Bruns, E., Chmielewski, F. M., Defila, C., Lipa, W., Menzel, A. (2007): Guidelines for plant phenological observations. World Climate Data and Monitoring Programme 1484, 1–41. Available online: https://library.wmo.int/viewer/51138/download?file=wmo-td_1484_en.pdf&type=pdf&navigator=1 (accessed on 20 June 2025).

Köppen climate classification. Available online: https://www.britannica.com/science/Koppen-climate-classification (accessed on 20 June 2025).

Lieth, H. (1974): Purposes of a Phenology Book. In Lieth, H. (Ed.), Phenology and Seasonality Modeling. Ecological Studies 8. Berlin, Heidelberg: Springer, https://doi.org/10.1007/978-3-642-51863-8_1.

Liu, F., Chen, Y., Shi, W., Zhang, S., Tao, F., Ge, Q. (2017): Influences of agricultural phenology dynamic on land surface biophysical process and climate feedback. Journal of Geographical Sciences 27, 1085–1099, https://doi.org/10.1007/s11442-017-1423-3.

Lobell, D. B., Gourdji, S. M. (2012): The Influence of Climate Change on Global Crop Productivity. Plant Physiology 160(4), 1686–1697, https://doi.org/10.1104/pp.112.208298.

Lu, L., Yu, L., Li, X., Gao, L., Bao, L., Chang, X., Gao, X., Cai, Z. (2025): Assessing Vegetation Canopy Growth Variations in Northeast China. Plants 14(1): 143, https://doi.org/10.3390/plants14010143.

Menzel, A. (2000): Trends in phenological phases in Europe between 1951 and 1996. International Journal of Biometeorology 44, 76–81, https://doi.org/10.1007/s004840000054.

Meroni, M., d'Andrimont, R., Vrieling, A., Fasbender, D., Lemoine, G., Rembold, F., Seguini, L., Verhegghen, A. (2021): Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and -2. Remote Sensing of Environment 253: 112232, https://doi.org/10.1016/j.rse.2020.112232.

Michavila, M. A., Vicente-Serrano, S. M., Llovería, R. M., Cai, Z., Eklundh, L. (2024): Evaluación espacialmente continua de la dinámica de la fenología vegetal en España entre 1983 y 2020 a partir de imágenes de satélite. Cuadernos de Investigación Geográfica 50(1), 145–178, https://doi.org/10.18172/cig.5739.

Misra, G., Cawkwell, F., Wingler, A. (2020): Status of Phenological Research Using Sentinel-2 Data: A Review. Remote Sensing 12(17): 2760, https://doi.org/10.3390/rs12172760.

Mourad, R., Jaafar, H., Anderson, M., Gao, F. (2020): Assessment of Leaf Area Index Models Using Harmonized Landsat and Sentinel-2 Surface Reflectance Data over a Semi-Arid Irrigated Landscape. Remote Sensing 12(19): 3121, https://doi.org/10.3390/rs12193121.

Nord, E. A., Lynch, J. P. (2009): Plant phenology: a critical controller of soil resource acquisition. Journal of Experimental Botany 60(7), 1927–1937, https://doi.org/10.1093/jxb/erp018.

Olesen, J. E., Børgesen, C. D., Elsgaard, L., Palosuo, T., Rötter, R. P., Skjelvåg, A. O., Peltonen-Sainio, P., Börjesson, T., Trnka, M., Ewert, F., Siebert, S., Brisson, N., Eitzinger, J., van Asselt, E. D., Oberforster, M., van der Fels-Klerx, H. J. (2012): Changes in time of sowing, flowering and maturity of cereals in Europe under climate change. Food Additives & Contaminants: Part A 29(10), 1527–1542, https://doi.org/10.1080/19440049.2012.712060.

Pei, J., Tan, S., Zou, Y., Liao, C., He, Y., Wang, J., Huang, H., Wang, T., Tian, H., Fang, H., Wang, L., Huang, J. (2025): The role of phenology in crop yield prediction: Comparison of ground-based phenology and remotely sensed phenology. Agricultural and Forest Meteorology 361: 110340, https://doi.org/10.1016/j.agrformet.2024.110340.

Pluto-Kossakowska, J. (2021): Review on Multitemporal Classification Methods of Satellite Images for Crop and Arable Land Recognition. Agriculture 11(10): 999, https://doi.org/10.3390/agriculture11100999.

Qin, G., Wu, J., Li, C., Meng, Z. (2024): Comparison of the hybrid of radiative transfer model and machine learning methods in leaf area index of grassland mapping.

Theoretical and Applied Climatology 155, 2757–2773, https://doi.org/10.1007/s00704-023-04779-5.
Richter, K., Hank, T. B., Vuolo, F., Mauser, W., D'Urso, G. (2012): Optimal Exploitation of the Sentinel-2 Spectral Capabilities for Crop Leaf Area Index Mapping. Remote Sensing 4(3), 561–582, https://doi.org/10.3390/rs4030561.
Rodriguez-Galiano, V. F., Dash, J., Atkinson, P. M. (2015): Intercomparison of satellite sensor land surface phenology and ground phenology in Europe. Geophysical Research Letters 42(7), 2253–2260, https://doi.org/10.1002/2015GL063586.
Řezník, T., Pavelka, T., Herman, L., Lukas, V., Širůček, P., Leitgeb, Š., Leitner, F. (2020): Prediction of Yield Productivity Zones from Landsat 8 and Sentinel-2A/B and Their Evaluation Using Farm Machinery Measurements. Remote Sensing 12(12): 1917, https://doi.org/10.3390/rs12121917.
Schreier, J., Ghazaryan, G., Dubovyk, O. (2021): Crop-specific phenomapping by fusing Landsat and Sentinel data with MODIS time series. European Journal of Remote Sensing 54(sup1), 47–58, https://doi.org/10.1080/22797254.2020.1831969.
Shen, Y., Zhang, X., Yang, Z., Ye, Y., Wang, J., Gao, S., Liu, Y., Wang, W., Tran, K. H., Ju, J. (2023): Developing an operational algorithm for near-real-time monitoring of crop progress at field scales by fusing harmonized Landsat and Sentinel-2 time series with geostationary satellite observations. Remote Sensing of Environment 296: 113729, https://doi.org/10.1016/j.rse.2023.113729.
Sisheber, B., Marshall, M., Mengistu, D., Nelson, A. (2022): Tracking crop phenology in a highly dynamic landscape with knowledge-based Landsat–MODIS data fusion. International Journal of Applied Earth Observation and Geoinformation 106: 102670, https://doi.org/10.1016/j.jag.2021.102670.
Sisheber, B., Marshall, M., Mengistu, D., Nelson, A. (2023): Detecting the long-term spatiotemporal crop phenology changes in a highly fragmented agricultural landscape. Agricultural and Forest Meteorology 340: 109601, https://doi.org/10.1016/j.agrformet.2023.109601.
Skakun, S., Vermote, E., Franch, B., Roger, J.-C., Kussul, N., Ju, J., Masek, J. (2019): Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating Surface Reflectance, Through Phenological Fitting, into Regression Yield Models. Remote Sensing 11(15): 1768, https://doi.org/10.3390/rs11151768.
Sparks, T. H., Carey, P. D. (1995): The Responses of Species to Climate Over Two Centuries: An Analysis of the Marsham Phenological Record, 1736–1947. Journal of Ecology 83(2), 321–329, https://doi.org/10.2307/2261570.
Springer, T. L., Aiken, G. E. (2015): Harvest frequency effects on white clover forage biomass, quality, and theoretical ethanol yield. Biomass and Bioenergy 78, 1–5, https://doi.org/10.1016/j.biombioe.2015.04.003.
Tomíček, J., Mišurec, J., Lukeš, P. (2021): Prototyping a Generic Algorithm for Crop Parameter Retrieval across the Season Using Radiative Transfer Model Inversion and Sentinel-2 Satellite Observations. Remote Sensing 13(18): 3659, https://doi.org/10.3390/rs13183659.
Tomíček, J., Mišurec, J., Lukeš, P., Potůčková, M. (2022): Retrieval of Harmonized LAI Product of Agricultural Crops from Landsat OLI and Sentinel-2 MSI Time Series. Agriculture 12(12): 2080, https://doi.org/10.3390/agriculture12122080.
Unc, A., Altdorff, D., Abakumov, E., Adl, S., Baldursson, S., Bechtold, M., Cattani, D. J., Firbank, L. G., Grand, S., Guðjónsdóttir, M., Kallenbach, C., Kedir, A. J., Li, P., McKenzie, D. B., Misra, D., Nagano, H., Neher, D. A., Niemi, J., Oelbermann, M., Overgård Lehmann, J., Parsons, D., Quideau, S., Sharkhuu, A., Smreczak, B., Sorvali, J., Vallotton, J. D., Whalen, J. K., Young, E. H., Zhang, M., Borchard, N. (2021): Expansion of Agriculture in Northern Cold-Climate Regions: A Cross-Sectoral Perspective on Opportunities and Challenges. Frontiers in Sustainable Food Systems 5, https://doi.org/10.3389/fsufs.2021.663448.
Van Tricht, K., Degerickx, J., Gilliams, S., Zanaga, D., Battude, M., Grosu, A., Brombacher, J., Lesiv, M., Bayas, J. C. L., Karanam, S., Fritz, S., Becker-Reshef, I., Franch, B., Mollà-Bononad, B., Boogaard, H., Pratihast, A. K., Koetz, B., Szantoi, Z. (2023): WorldCereal: a dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping. Earth System Science Data 15(12), 5491–5515, https://doi.org/10.5194/essd-15-5491-2023.
Webb, N., Nicholl, C., Wood, J., Potter, E. (2016): SunScan Manual, Version 3.3; Delta-T Devices Ltd.: Cambridge, UK, 1–82. Available online: https://delta-t.co.uk/wp-content/uploads/2017/02/SSI-UM_v3.3.pdf (accessed on 20 January 2025).
Wielgolaski, F.-E. (1974): Phenology in Agriculture. In Lieth, H. (Ed.), Phenology and Seasonality Modeling Studies 8. Berlin, Heidelberg: Springer, https://doi.org/10.1007/978-3-642-51863-8_31.
Wolanin, A., Camps-Valls, G., Gómez-Chova, L., Mateo-García, G., van der Tol, C., Zhang, Y., Guanter, L. (2019): Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations. Remote Sensing of Environment 225, 441–457, https://doi.org/10.1016/j.rse.2019.03.002.
Xu, J., Quackenbush, L. J., Volk, T. A., Im, J. (2022): Estimation of shrub willow biophysical parameters across time and space from Sentinel-2 and unmanned aerial system (UAS) data. Field Crops Research 287: 108655, https://doi.org/10.1016/j.fcr.2022.108655.
Yuan, X., Li, S., Chen, J., Yu, H., Yang, T., Wang, C., Huang, S., Chen, H., Ao, X. (2024): Impacts of Global Climate Change on Agricultural Production: A Comprehensive Review. Agronomy 14(7): 1360, https://doi.org/10.3390/agronomy14071360.
Zhang, H., Zhang, Y., Liu, K., Lan, S., Gao, T., Li, M. (2023): Winter wheat yield prediction using integrated Landsat 8 and Sentinel-2 vegetation index time-series data and machine learning algorithms. Computers and Electronics in Agriculture 213: 108250, https://doi.org/10.1016/j.compag.2023.108250.
Zhang, X., Liu, L., Henebry, G. M. (2019): Impacts of land cover and land use change on long-term trend of land surface phenology: a case study in agricultural ecosystems. Environmental Research Letters 14(4): 044020, https://doi.org/10.1088/1748-9326/ab04d2.
Zhang, X., Tan, B., Yu, Y. (2014): Interannual variations and trends in global land surface phenology derived

from enhanced vegetation index during 1982–2010. International Journal of Biometeorology 58, 547–564, https://doi.org/10.1007/s00484-014-0802-z.
Zhou, J., Jia, L., Menenti, M. (2015): Reconstruction of global MODIS NDVI time series: Performance of Harmonic ANalysis of Time Series (HANTS). Remote Sensing of Environment 163, 217–228, https://doi.org/10.1016/j.rse.2015.03.018.
Zhu, Z., Wang, S., Woodcock, C. E. (2015): Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sensing of Environment 159, 269–277, https://doi.org/10.1016/j.rse.2014.12.014.
Zhu, Z., Woodcock, C. E. (2012): Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment 118, 83–94, https://doi.org/10.1016/j.rse.2011.10.028.

Potential migration of Czech medical students with special regard to migration drivers and barriers

Michal Šimůnek, Dušan Drbohlav*

Charles University, Faculty of Science, Department of Social Geography and Regional Development, Czechia
* Corresponding author: dusan.drbohlav@natur.cuni.cz

ABSTRACT

This article examines the phenomenon of potential international migration among Czech medical students, with the objective of identifying prospective and actual migrants and analyzing the underlying motivations, intended durations of migration, the roles of institutional structures versus social networks, the strength of social ties, preferred destinations, and perceived barriers to mobility. Anchored in selected established migration theories and conceptual frameworks, the study deployed an electronically administered questionnaire targeting 397 fourth – to sixth-year General Medicine students across four faculties of Charles University (Czechia) during March–April 2022. Data analysis was conducted using SPSS software, incorporating factor analysis and binary logistic regression. Despite a generally higher inclination toward migration, only 7% of respondents exhibited a strong likelihood of actual emigration when specific preparatory steps and intended timelines were considered. This subset of students, characterized by clearly articulated goals to enhance professional and financial prospects, reliance on transnational social networks, preference for extended stays abroad, and a diminished likelihood of return, contrasts markedly with the broader cohort of less committed potential migrants. Proficiency in German emerges as a key determinant for Czech medical students, particularly in the context of migration to German-speaking countries such as Germany, Austria, and Switzerland, while English-speaking countries like the USA and Canada remain attractive destinations too. Beyond language competencies, prior international work or study experience were found to significantly inform students' migration trajectories. Nevertheless, the most prominent deterrent to migration remains the anticipated loss of familial and social connections.

KEYWORDS

potential emigration; medical students; regression analysis; Charles University; Czechia

Received: 6 May 2025
Accepted: 26 October 2025
Published online: 27 November 2025

Šimůnek, M., Drbohlav, D. (2025): Potential migration of Czech medical students with special regard to migration drivers and barriers. AUC Geographica 60(2), 254–268
https://doi.org/10.14712/23361980.2025.21

1. Introduction

International migration and recruitment of health professionals is receiving increased attention globally. The World Health Organization reported in 2016 that there is a shortage of 2.8 million physicians worldwide (World 2016). Professions related to the health sector are in high demand even in the developed countries of the world, where this problem is mainly addressed by recruiting doctors from abroad (Ikenwilo 2007; Vavrečková 2007). The migration of health professionals is steadily increasing, mainly due to demographic problems, which include the high average age of doctors and a generally growing aging population with a greater need for medical care, generally increasing the demand for health and social services. European Union (EU) countries are no exception to these trends (Bojanic et al. 2015; Dodani and LaPorte 2005; IHS Markit 2021; Żuk et al. 2019).

The problems have been compounded by the impact of the Covid-19 pandemic, which broke out globally in the early 2020s (it did not subside until early 2022), causing an overload for physicians worldwide (Bojanic et al. 2015; IHS Markit 2021). In Czechia, students in the final years of medical faculties and secondary medical schools were called up due to the limited capacity of the medical staff (iROZHLAS 2020). The American Medical Association has stated that overload is so high that it can lead to burnout syndrome and physicians leaving practice (IHS Markit 2021).

In many EU countries (including Czechia), physician emigration is perceived as a relatively large societal problem, especially for young doctors and graduates (Bojanic et al. 2015; Humphries et al. 2021; Krajewski-Siuda et al. 2012; Żuk et al. 2019). This is mainly due to the fact that healthcare workers seek opportunities that match their skills, qualifications, individual preferences, and need for higher earnings (Brennan et al. 2023; de Vries et al. 2023; Dodani and LaPorte 2005; Lee 1966).

The main aim of this paper is to examine the potential migration of physicians, specifically medical students, in Czechia. Hence, two important gaps in the current knowledge will be addressed. First, we compare our research findings with those from the early 2000s (Vavrečková et al. 2006). Second, we introduce a comparative international perspective by examining the nature and determinants of migration preferences in Czechia in relation to other EU countries – particularly in Central and Eastern Europe, such as Poland, Croatia, and Romania (see sources below).

This study seeks to identify the importance of potential and actual (although not yet realised) migration abroad of Czech medical students studying at Charles University (the oldest Czech university with the most students[1]). It also examines the destinations, the expected length of stay, the motivations that lead medical students to leave and the barriers that prevent them from leaving. Particular attention is paid to finding out how potential and actual respondent-migrants differ and what role institutions, as well as family and other personal networks play in the potential migration movement of medics. This research also supplements studies carried out in other EU countries with similar socio-economic development (Poland – Krajewski-Siuda et al. 2012; Pszczółkowska et al. 2024; Croatia – Bojanic et al. 2015; and Romania – Apostu et al. 2022).

The paper briefly introduces several concepts, through which we can better understand and interpret our data. This is followed by an evaluation of selected relevant sources with an introduction to the health context of Czechia. The objectives and research questions precede the methodological approach and methods used. After the presentation of the research results, the paper concludes with a discussion and conclusions.

2. Theoretical framework

For better understanding and interpreting the results of our study, we make use of the following conceptual frameworks:

The migration push and pull model (Lee 1966; and many other researchers) is a concept contributing to understanding which factors influence migration processes and how. The model is the result of oppositely acting forces – the so-called "pull", which attracts a person to a given place of residence, and the push, which drives them from their initial place of residence.

Migration network theory enriches the push and pull model with linkages between the migrants themselves (Light et al. 1989). This theory is defined as a set of interpersonal ties that link prospective migrants, former migrants and non-migrants in source and destination countries in the form of kinship, friendship or community ties (Massey 1988, 1993). These migration networks provide important assistance in overcoming barriers to migration and subsequent integration. Migration networks can be distinguished according to the depth of social ties and the nature of the interaction into two basic types, namely strong ties (personal social networks) and weak ties (wider social networks) (Boyd 1989; Granovetter 1973). According to Granovetter (1973), weak ties tend to be more important in obtaining employment opportunities, while emphasizing that these weak ties create bridges between nodes of strong relationships. In addition to direct contacts, online social networks

1 In 2024, Charles University had a total of 53,219 students enrolled, 14,562 of whom were studying at medical faculties. Masaryk University, the second-largest institution, had 33,531 students enrolled, 4,786 of whom were studying at medical faculties (MŠMT 2024).

have recently emerged and play an important role in communication too.

Very simply put, "the neoclassical economics model assumes that people move abroad permanently to maximize lifetime earnings, whereas the new economics of labour migration assume they leave temporarily to overcome market deficiencies at home" (Constant and Massey 2002: 5).

Institutional theory emphasizes the role of institutional actors in influencing migration. These mostly include the government, financial institutions, real estate agencies, law firms, social and humanitarian organisations, non-governmental agencies, as well as mafia-type structures. All of these are significantly involved in shaping the migration behaviour of the population. They reinforce and benefit from migration flows.

The given conceptual framework includes selected, yet highly important aspects of migration motivation and migration mechanisms, which are reflected in both migration preferences and actual migration – at the micro level of the individual migrant as well as from the perspective of meso- and macro-structural attributes. The selection of these theories and concepts, also makes it possible to better understand the determinants of migration and migration preferences in both the source and destination countries. Equally important is the interconnection of the applied theories/concepts, which further enhances their explanatory value.

In the text below, we only discuss the factors and contingencies that emerge in the reality of potential and real migration of medics and doctors from relatively developed countries (including Czechia) to the most developed ones. First, we will characterize the motivations and barriers of the given migration, and then the positive and negative consequences.

2.1 Drivers and barriers to migration

Migration of medical professionals is driven by a complex interplay of macro-level and micro-level factors, ranging from economic incentives to personal aspirations. Countries with high economic performance tend to have higher market wages than countries with lower economic performance. This causes workers to migrate from regions with lower wages to regions with higher wages (IOM 2014; Massey 1993). Even in the medical sector, the most commonly cited economic drivers of migration are average gross wages and related taxes for physicians (Tjadens et al. 2013). Additionally, the process of medical specialization requires post-graduate examinations, which often delay the full financial benefits of the profession in the country of origin, particularly in Czechia (Tjadens et al. 2013; Sedláková 2015).

A very important aspect of migration, especially for future doctors, is the quality of postgraduate education, which can vary widely (Sedláková 2015). Studying abroad can also give future doctors an advantage over those studying in their home country, for example, in career progression, as is the case in Ireland (Akl et al. 2007; Humphries et al. 2021). Then there are the better working conditions abroad, which can include high quality equipment, less bureaucracy, shorter working hours or even less overtime, top-notch co-workers, as well as a favourable social climate. In some countries (e.g. Poland, Germany), on the other hand, it is perceived that older doctors are promoted over younger ones, which creates a bad climate in health care institutions (Tjadens et al. 2013; Docquier 2006). The recognition of diplomas and attestations throughout the EU definitely contributes to international migration in the health care sector (Vavrečková and Gazdagová 2005). Directive 2005/36/EC ensures recognition of qualifications across the EU, easing administration and enabling smooth integration into healthcare systems (Gkolfinopoulos 2016). Among geographical factors, geographical proximity to economically stronger countries is particularly important (Tjadens et al. 2013). In general, the more distant the home country is from the destination country, the more migration decreases (Fields 1979).

On the micro level, social networks are of paramount importance. For migrants, it is very important to maintain contact with someone from abroad or to know someone in their close circle who lives abroad. If such contacts are maintained, it is much easier for the person to make the decision to migrate. This is greatly helped nowadays by new technologies such as Facebook, Whatsapp etc. (Tjadens et al. 2013). Social networks can facilitate migrants' access to job opportunities and can also have a positive effect on the eventual integration process. They also support the emergence of subcultures, which can greatly benefit immigrant groups (Horáková 2007). Furthermore, social networks are considered a very reliable source of information, which is one of the key aspects in migration decision-making (Schumann et al. 2019).

Personal factors are also very important in highly skilled migration. The main determinant within this area is generally an improvement in quality of life (Horáková 2007). Long-term migration can increase the possibility of personal development, the development of language skills or the fulfillment of the desire to experience new cultures and lifestyles (Tjadens et al. 2013). Health systems are broadly similar around the world, allowing students and doctors to travel around the world without having problems finding a job in this sector (Sedláková 2015; Tjadens et al. 2013).

Despite strong push and pull factors, various barriers can deter medical professionals from emigrating. For young doctors, not knowing a foreign language, the impossibility of easy communication with patients, can be one of the major personal barriers (Bojanic et al. 2015; Krajewski-Siuda et al. 2012). Other barriers that have been frequently cited in previous research include separation from family and friends (Gouda et al. 2015). Financial factors that may prevent long-term migration are also quite numerous. The financial costs of moving and adapting to life in the destination country represent one of the key challenges (Lee 1966). One of the other significant barriers to mobility may be visa procedures, and regulations for working as a physician, such as licensing and registration fees that are in place in the USA or the UK, for example (Brennan et al. 2023; Schumann et al. 2019).

2.2 Consequences

In general, from the perspective of the country of origin, the negative consequences of migration of future doctors or highly skilled labour tend to prevail. Quite a number of studies have concluded that the outflow of highly skilled workers from medicine contributes significantly to the shortage of medical personnel, the consequences of which affect mainly the quality of medical care in countries of origin (Boncea 2013; Żuk et al. 2019). Countries lose not only investment in medical education but also the health care workforce (Dodani and LaPorte 2005). In addition, as the migration of young doctors increases, what happens is that the workforce ages faster in the source country, leading to a growing shortage of doctors – there is a lack of new doctors to replace older doctors who retire (IOM 2014). In the short term, the biggest problem is the shortage of qualified medical staff and the problems that are tied to this (Gouda et al. 2015). However, migration of doctors for work can also turn into permanent migration much more easily than before, especially due to higher wages or better working conditions abroad. This is also facilitated by the fact that it is much easier for EU residents to migrate with their families across the EU (Suciu et al. 2017).

Among the positive consequences, skilled medical professionals gain new knowledge and skills that they can bring to the domestic health system if they return to their country of origin (Docquier 2006; Suciu et al. 2017). Furthermore, these returnees can introduce new medical practices. And they can also create an imaginary bridge with those who are still abroad (Bojanic et al. 2015). Thus, return migration can, among other things, create international scientific networks, which in turn can increase and facilitate the movement of people and goods, but also ideas, and can contribute to an increase in foreign direct investment, including in the health sector (Docquier and Rapoport 2012). Return migration can also be seen as another source of "brain gain", where doctors who return to their country of origin bring new experience and knowledge that they have gained in the most developed countries (Docquier 2006; Levitt 1998 and her concept of social remittances). Financial remittances can also be counted among the positive consequences.

3. Context of Czechia

The OECD statistics (2021) show that Czechia is in a relatively good position in terms of the number of newly graduated doctors (17 per 100,000 population – STATISTA 2024, Fig. 1); in absolute terms between 2016 and 2020 – 1,035 newly registered Czech doctors in Czechia per year (Czech 2021), and in terms of the total number of doctors (4.1 per 1,000 population). According to OECD statistics (2021), the feminization of the Czech health care system continues (56% of doctors are women) and the process of aging of health professionals is underway (14% of doctors are over 65 and another 19% are aged 55–64; only 23% of doctors are under 35). In terms of equipment

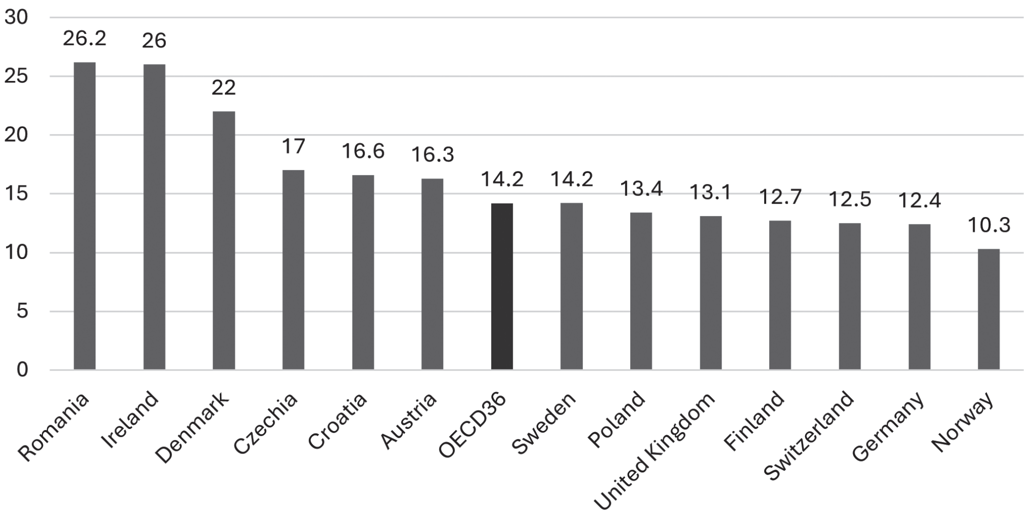

Fig. 1 Graduates of medicine in OECD countries in 2021 (per 100,000 population).
Note: The table displays a selection of OECD countries.
Source: STATISTA (2024).

with modern technology, as measured by the number of CT (computed tomography) scanners and MRI (magnetic resonance imaging) units, Czechia suffers from under-equipment in a European comparison (OECD 2021).

While in Czechia all the administrative work is done by doctors, in the most developed countries doctors are to some extent spared the administrative burden (Vavrečková et al. 2007). This mismatch reflects the structural push factors outlined in traditional migration theory, particularly those relating to workload and inadequate system resources (Lee 1966). The dissatisfaction of Czech doctors with the health care system occasionally escalates into publicized protests. Evidence of this can be seen in the “Thank You, We’re Leaving” initiative of 2010/2011, which aimed to improve working conditions, especially pay and compulsory further training for doctors, or the current disputes and rejection of the existing overtime model by many doctors (e.g. ČTK 2023). These mobilisations reflect collective discontent shaped by professional identity and sustained through interpersonal and institutional networks, as described in migration network theory (Massey et al. 1993; Granovetter 1973). Such ties enhance the perceived feasibility of migration, especially among younger doctors. Between 2015 and 2019, 1,583 Czech doctors migrated abroad, most of them to the UK (506) and Germany (340).

These patterns are not unique to Czechia. Studies from Poland, Romania, and Croatia – countries with comparable healthcare systems and socio-economic contexts – consistently report high migration intentions among medical students and junior doctors (Suciu et al. 2017; Krajewski-Siuda et al. 2012; Bojanic et al. 2015). Common push factors include low pay, limited access to postgraduate training, and dissatisfaction with domestic healthcare infrastructure. However, recent evidence from Poland indicates that systemic issues – such as political instability, poor working conditions, and a climate of public distrust toward medical professionals – now outweigh remuneration as key motives (Pszczółkowska et al. 2024). Similarly, a Romanian study found that migration propensity remained high despite recent wage increases, underscoring the persistent role of structural disparities and weak institutional support (Apostu et al. 2022).

4. Objectives and research questions

The target population of our research are students of general medicine, who were interviewed using an online questionnaire survey. Based on the sources cited above, several basic research questions were defined:

- What are the migration tendencies and what is the probability of migration intention of future doctors?
- How significant is the role of institutional actors in labour migration decisions among future physicians, compared to that of family and other personal networks?
- What motives most influence future doctors when deciding whether to work abroad?
- What is the expected length of work experience of future physicians?
- Which countries are the main migration destination territories of future doctors?
- What factors prevent future doctors from going abroad to work?

5. Methodology

We use a quantitative research approach. The method of data collection was a questionnaire survey. The sample studied is made up of Czech students (those who are citizens of the Czech Republic) of the 4th-6th years studying General Medicine at the medical faculties of Charles University. 469 respondents out of a total of 2,370 students completed the questionnaire survey, giving a return rate for the survey of 20%. In editing and coding, 72 respondents were excluded due to non-Czech citizenship or insufficient completion of key questions. This resulted in a total of 397 respondents in the final sample.

The respondents were approached through the individual study departments of the medical faculties of Charles University. The data collection and the actual filling-in took place online, through an electronic questionnaire. The survey was conducted anonymously and participation was entirely voluntary. Data was collected via an online platform that did not store any personal identifiers, thereby ensuring confidentiality. Respondents were informed about the purpose of the study and how their data would be used. Online surveying was chosen for its time efficiency, ease of access via computer or mobile device, and the ability for respondents to pause and resume later. The questionnaire survey was carried out over two months (March 1, 2022 – April 30, 2022). All medical faculties of Charles University participated in the research except the 2nd Faculty of Medicine in Prague. Their management did not specify the reason for non-participation.

The sample as such was further divided into three groups, namely 96 so-called potential migrants (with further breakdown below), 240 so-called non-migrants and 61 so-called undecided, who were not counted in most statistical analyses to avoid biasing the whole research.

The initial questionnaire was designed based on our research questions and prior literature. It was then refined in collaboration with the chairman of the Young Doctors Association, whose team has prior experience of collecting data from young doctors and maintains ongoing contact with this target group.

Furthermore, the structure of the questionnaire enables to some extent comparison with a study conducted 19 years ago (Vavrečková et al. 2006).

The questionnaire consists of 21 questions and is divided into several sections. The questionnaire included closed, semi-closed, open-ended and scale questions to combine structured, quantifiable responses with deeper qualitative insights, thereby enhancing the validity and analytical richness of the data (Creswell and Creswell 2018). Respondents did not answer all items, instead, they were directed only to those questions relevant to them based on their previous answers, using filter questions to achieve this structure. The first part was used to obtain respondents´ characteristics for further classification and provide some basic information about them. The second part asked expanding questions related to language skills and previous study or work experience. The third part explored migration itself and focused on potential migrants as well as a group of undecideds. Non-migrants answered only one follow-up question on barriers to migration.

Descriptive statistics was used in the data analysis, along with more sophisticated methods. Binary logistic regression was used to predict migration intentions. This method was chosen because it enables us to estimate the probability of an outcome (intention to migrate) based on the presence or absence of explanatory variables, while also assessing their statistical significance and controlling for other factors. Factor analysis was chosen to determine the structure among the multiple items measuring barriers that were rated by respondents on a Likert scale. Due to the high dimensionality of the barrier-related items, factor analysis was used to identify latent constructs that grouped strongly correlated items. This reduced complexity. It also retained the main patterns in respondents' perceptions. In the factor analysis, which met the condition that a larger number of variables should be correlated with a value higher than 0.3 and also established the goodness of fit of the data, using the Kaiser-Maeyer-Olkin test (0.720), we use the "Principal components" method with Varimax rotation.

6. Research results

6.1 The likelihood of migration

The migration potential of medics was investigated via using a key question, i.e.: "Do you expect to work as a doctor abroad at some point after completing your current studies?" If respondents answered "definitely yes" or "rather yes", they were classified as potential migrants. If they answered "definitely not" or "rather not", they are classified as in the non-migrant group. And a "don't know" response placed respondents in the undecided group. Of the total sample, 60% fell into the non-migrant group, 24% into the potential migrant group and the remaining 15% into the undecided group. In order to better determine how many medics, out of 96 potential migrants, want to really go abroad to work, we need to determine the degree of reality of this proclaimed departure. We determined the degree of realism based on two factors – the implementation of concrete steps for the intention to work abroad and the date of departure. Focusing on the first factor, we found that 27% of medics had done nothing towards moving abroad, 39% of medics had not yet taken any concrete steps but plan to do so, and 34% of medics had already taken some concrete steps and are thus highly likely to actually migrate. The results of the second factor show that 46% of the respondents want to move abroad immediately after the completion of their current studies and 21% of the medics want to leave after the completion of the first general block of postgraduate education (this takes 30 months). For these medics, we assume that they are fairly firmly decided about going abroad for work. Furthermore, 11% of medics want to migrate after passing the final certification exam (5–6 years) or later. 23% responded that they did not know.

In order to determine as accurately as possible the reality of migration and thus determine who is an active migrant, two conditions must be strictly met simultaneously. Firstly, the individual has already taken specific steps to realize the intention to work abroad and secondly, he/she wants to go abroad either immediately after the completion of the final certification exam or after the completion of the first general block. Given these conditions, 27% fall into the group of active migrants and 73% remain in the group of potential migrants. The degree of realistic migration intention in the context of the whole sample, i.e. 397 respondents, is reflected in Fig. 2, where we see that only some 7% of the original 24% of potential migrants are active migrants.

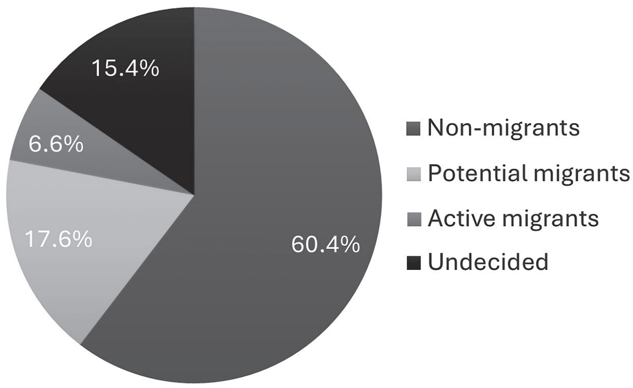

Fig. 2 Strength of the migration plan (%).
Note: N = 397; based on the question: "Do you expect to work as a physician abroad at some point after graduation?"
Options: "definitely yes", "rather yes", "I don't know", "rather no", "definitely no".
Source: own research (2022).

Tab. 1 CEFR language proficiency levels of respondents (%).

	A1	A2	B1	B2	C1	C2
English language	0.8	1.8	14.1	44.7	31.5	7.1
German language	19.9	17.9	18.2	10.3	5.5	1.0
French language	12.3	8.6	6.0	3.8	1.8	0.8
Russian language	11.8	2.3	1.5	0.3	0.0	1.3
Spanish language	12.6	5.3	4.0	1.5	0.3	0.8
other language	2.3	2.0	1.5	0.3	0.8	2.3

Note: N = 397; other languages – Italian, Polish, Norwegian, Vietnamese etc.; based on the question: "What is the level of proficiency in each language?" – For available options see more in the table.
Source: own research (2022).

6.2 Knowledge of languages and foreign experience

Participants in the study were aged 21 to 26 and 34% were male and 66% female. Knowledge of world languages is crucial for gaining a job abroad, especially knowledge of English. In the case of Czechia, German is also important due to its geographical location. The medics assessed their language skills in five basic world languages. Moreover, they also had the opportunity to supplement and rate other languages they know (Tab. 1). They assessed their proficiency on the basis of the CEFR – Common European Framework of Reference, which divides three levels of language proficiency, which are further subdivided into six levels. These levels are 'Basic' (A1, A2), 'Independent' (B1, B2), and 'Proficient' (C1, C2) (CAMBRIDGE 2022). It can be noted that 97% of medics are proficient in English at levels B1–C2, of which the B2 level is the most common (45%). However, it is worth mentioning the C1 (32%) and C2 (7%) levels too. German tends to be a secondary language for most respondents in Czechia, with 35% of medics proficient at the B1–C2 level.

When dividing the sample into potential and active migrants, the level of English is over 96% at B1–C2 for both groups. C1 is the most common level. The difference, however, can be seen in the knowledge of German language, where 46% of the active migrants have a knowledge of German at the B2–C2 level, compared to only 27% of the potential migrants. For Czech medical students, immediate readiness to speak German well seems to be one of the key factors in the decision to actually migrate to a geographically close German-speaking country.

In addition to language, study or work experience abroad can also significantly influence respondents' decision to go abroad. It is mainly about creating a network of social relationships or gathering valuable information about the country. In our research we looked at experience of more than 3 months over the last 10 years. From a total sample of 397 respondents, 15% of respondents declared that they already had such experience.[2]

2 Based on the question: "Have you studied or worked abroad in the last 10 years?" – Options: "yes", or, "no"; N = 397.

6.3 Destinations

Across the whole sample of 96 potential migrants, geographically close countries such as Germany and Austria were most frequently mentioned, followed by Switzerland. The island countries of the United Kingdom and Ireland performed slightly worse. Comparing the destination country decisions of potential and active migrants, some differences can be noticed, except for the most frequently mentioned Germany, which is represented by 25% and 27% respectively in both groups (Fig. 3). Active migrants most frequently mentioned geographically close countries, i.e. Austria and Switzerland. Potential migrants most frequently mentioned the Nordic countries, with Sweden being the most frequently mentioned, followed by the United Kingdom and Ireland.

6.4 Length of stay abroad

In terms of length of stay abroad, in the total sample of 96 potential migrants, more than half are considering return migration, with 36% of respondents expecting to return within 5 years and 17% considering staying abroad for more than 5 years. A quarter of respondents intend to stay abroad permanently. If we examine potential and active migrants separately, we can see significant differences (Fig. 4). A huge difference exists for permanent migration, with 42% of active migrants wanting to stay abroad permanently, which is a significant difference compared to 19% of potential migrants. Return migration is expected by 54% of potential migrants and 50% of active migrants, but there is a significant difference in the length of stay abroad before the expected return. 42% of active migrants expect to return after more than 5 years, compared to 7% of potential migrants.

6.5 Motives for leaving

We next explored motives for leaving, where respondents were offered 15 different options, (including an "other" option) while evaluating the importance of any presented option (all options were rated on a five-point rating scale, ranging between 1 being completely

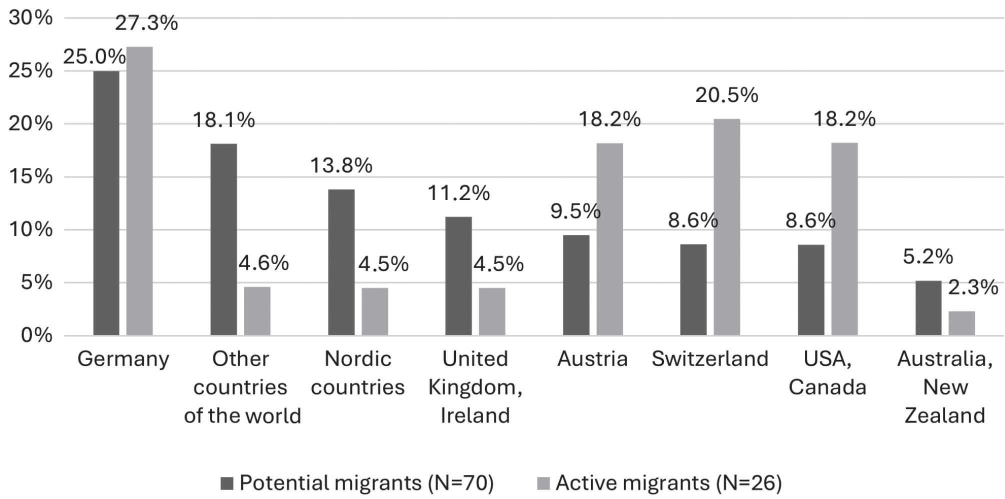

Fig. 3 Final destinations (%).
Note: Total N = 159; Potential migrants N = 70; Active migrants N = 26; based on the question: "To which country are you most likely to go? (Only one field needs to be filled in, i.e. country. If you do not know yet, leave the fields blank)."
Source: own research (2022).

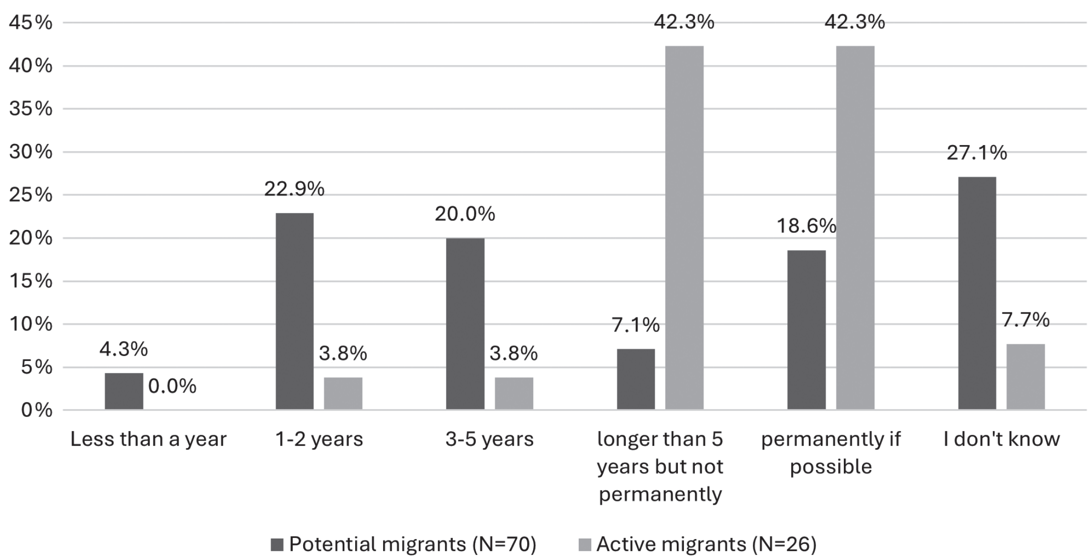

Fig. 4 Planned length of stay abroad (%).
Note: Potential migrants N = 70; Active migrants N = 26; based on the question: "How long are you likely to stay abroad?" For available options, see more in the table.
Source: own research (2022).

unimportant and 5 being completely important). An arithmetic mean was used in the evaluation, which was calculated for each motive separately. Across all 96 potential migrants combined, medics placed the greatest importance on gaining international insight and contacts. The other two important factors are higher financial remuneration and better working conditions. Focusing on potential and active migrants separately (Tab. 2), we found that for active migrants, better working conditions are more important, while for potential migrants the most important motive is to gain international insight. Both groups consider a higher financial reward to be a highly important motive. By contrast, whereas the motive of improving knowledge of a foreign language is ranked very high by potential migrants, for active migrants this motive is one of the least important (they already know the language well).

6.6 Factors behind medics' migration plan

The likelihood or risk of migration of medics to work abroad was analysed using binary logistic regression. Tab. 3 shows the fully adjusted odds ratios (Odds ratio = OR) as the output of the model at 95% confidence level (95% Confidence Interval = 95% CI).

All variables were entered into the model simultaneously. The dependent variable was the probability of movement, based on the question: "Do you expect to work as a doctor abroad at some point after

Tab. 2 Individual motives for going abroad to work.

N = 96	Potential migrants N = 70	Active migrants N = 26
Gaining international insight and contacts	4.07	3.85
Higher financial remuneration	3.86	4.00
Improvement of foreign language	3.80	2.58
Better working conditions	3.67	4.15
Increased opportunity for career growth in Czechia after returning from abroad	3.56	2.92
Better quality of postgraduate education	3.50	3.77
Work in a renowned medical facility	3.44	2.81
Opportunity to work with cutting-edge technology	3.43	3.23
Higher opportunity for career growth abroad	3.27	2.85
Higher level of healthcare in the destination country	3.26	3.46
Learning about a different cultural environment	3.24	2.77
Become independent	3.19	2.58
More opportunities to work in your specialised field	3.00	2.69
Geographical proximity of the destination country to the home country	2.51	2.12

Note: Min. = 1.00, Max. = 5.00; based on the question: "Please indicate how important motivation to work abroad is to you and rate the motives" (1 means completely unimportant and 5 means completely important) – For available options see more in the table.
Source: own research (2022).

completing your current studies?", which was measured in our study on a five-point scale ranging from "definitely yes" to "definitely no". This variable was dichotomized as "I plan to migrate" ("definitely yes", "rather yes") and "I do not plan to migrate" ("definitely no", "rather no"). We did not include the answer "don't know" in either variable and considered them as missing values. The variables gender, year of the study, faculty, previous work or study experience in the last few years abroad, and knowledge of English and German language were included in the analysis as independent variables. 336 respondents out of a total sample of 397 were included in the analysis. Nagelkerk's R Square, is equal to 0.297, which can be considered as a fairly competent (but not very high) ability of the model to explain the given reality.

The results show that there was no statistically significant relationship between the variables faculty, year and gender and the plan to go abroad. In the case of gender, where we chose women as the reference category, men are 1.3 times more likely (OR = 1.323; 95% CI 0.733–2.387) to go abroad for work than women. Within year groups, concerning the migration abroad for work, both Year 5 and Year 4 were equally 1.2 times more likely to migrate (OR = 1.233; 95% CI 0.621–2.446 and OR = 1.158; 95% CI 0.577–2.326, respectively) than Year 6, which was defined as the reference category. As for faculties, medics from the Faculty of Medicine of the Charles University in Hradec Králové were 2.7 times more likely (OR = 2.736; 95% CI 0.854–8.765) to go abroad for work than medics from the 1st Faculty of Medicine of the Charles University in Prague, which was the reference category.

In each of the variables describing previous work or study experience from abroad and the level of English and German language proficiency, some categories were detected as statistically significant in relation to the plan to migrate abroad for work. If a medical student had study or work experience abroad in previous years, he/she is 2.7 times more likely (OR = 2.680; 95% CI 1.276–5.627) to migrate for work abroad than those who do not have this experience. According to our model, a higher English language level plays no role in the decision to pursue migration. The opposite is true for the German language, where the higher the level of proficiency, the more likely it is to implement a plan to go abroad for work; for example, if a medical student knows German at C1 level, he/she is 19.5 times more likely to migrate (OR = 19.504; 95% CI 5.181–73.424) than one who does not know it at all.

6.7 Factors hindering migration

Barriers to medics going abroad to work were measured via using the question, "If you do not plan to go abroad to work, please indicate how important each of the factors that prevent you from doing so are to you." Medics rated each factor on a five-point scale from not at all important to very important. To better understand the variables analyzed, factor analysis was used where factor extraction was performed using the Principal components method, the result of which showed us three factors capturing 57% of the total variability. Based on the representation of each variable in the newly extracted factors, we found that

Tab. 3 Regression analysis – adjusted odds ratios (OR) and 95% confidence interval for finding the probability of migration.

Dependent variable: 1 = has got plans to migrate; 0 = has not got plans to migrate	N = 336		
Independent variables	Adjusted odds ratios (OR)	95% confidence interval	
		Lower	Upper
Females – REF	1		
Males	1.323	0.733	2.387
6th year – REF	1		
5th year	1.233	0.621	2.326
4th year	1.158	0.577	2.326
Experience abroad – no – REF	1		
Experience abroad – yes	**2.680**	**1.276**	**5.627**
1st Medical Faculty of Charles University in Prague – REF	1		
3rd Medical Faculty of Charles University in Prague	0.862	0.440	1.691
Faculty of Medicine of Charles University in Pilsen	0.940	0.456	1.938
Faculty of Medicine of Charles University in Hradec Králové	2.736	0.854	8.765
English language – A1 – REF	1		
English language – A2	0.080	0.003	2.531
English language – B1	**0.063**	**0.004**	**0.912**
English language – B2	0.097	0.008	1.206
English language – C1	0.261	0.021	3.218
English language – C2	0.621	0.046	8.477
German language – Does not know – REF	1		
German language – A1	2.033	0.846	4.882
German language – A2	**2.678**	**1.122**	**6.396**
German language – B1	1.753	0.745	4.124
German language – B2	**5.010**	**1.897**	**13.235**
German language – C1	**19.504**	**5.181**	**73.424**
German language – C2	14.146	0.796	251.528

Note: Bold when $p < 0.05$, REF = reference category.
Source: own research (2022).

the most information was retained in the variable of separation from family and friends (72%). On the other hand, the least information was retained in the variables of partner disapproval (31%).

Tab. 4 shows us the factor loadings. These factors give us the degree of correlation between the original and newly created variables. Based on the rotated matrix of factor loadings with respect to the variables saturated by the aforementioned factors characterizing the barriers to the migration movement of medical professionals, we chose to name them as follows:

- Factor 1: Psychological-administrative barrier.
- Factor 2: Social barrier.
- Factor 3: Professional barrier.

As part of the analysis, we created individual scores for each variable based on the factor score matrix. This shows us how each socio-demographic factor corresponds to each factor.

The psychological-administrative barrier is mainly related to self-esteem and the psychological burden that is exerted on medics. The social barrier, which plays the most important role means that medics are concerned about the lack of social contact with their family and close friends. Both of these factors are particularly connected to women. The professional barrier means that medics do not feel the need to go abroad to work because they are convinced that they will find a very good job in their field in Czechia. This factor corresponds mainly with men.

6.8 Influence of institutional actors vs family/ personal actors

We also investigated the role of institutions vs. family and other personal networks in the potential migration movement of medics, i.e., which actors medics intend to use to gain employment opportunities or who may be instrumental in their eventual integration in the destination country. We first focused on the interpersonal ties that might link potential migrants

Tab. 4 Rotated matrix of factor loadings.

	Factor		
	1	2	3
Comparative exam in some countries	0.775		
Doubts about the fulfilment of professional requirements in a given country	0.718		−0.209
Problems with attestation examinations	0.711		0.247
High costs associated with living abroad	0.662		0.258
Excessive paperwork and difficulty in dealing with various documents before departure	0.606		
Fear of language problems / I don't speak the language	0.492	0.289	−0.475
Separation from family and friends		0.851	
Loneliness and lack of social contacts	0.295	0.754	
Partner's disapproval		0.553	
Assumption of a good job in Czechia			0.819

Note: Extraction method: Principal components; the rotation method: Varimax (5 iterations).
Source: own research (2022).

Tab. 5 Importance of actors in getting a job abroad.

N = 96	Potential migrants N = 70	Active migrants N = 26
Internet sites for searching for job opportunities	3.20	3.31
Compatriot community	2.84	2.62
Family relatives in the destination country	2.74	2.77
Acquaintances from the country of origin	2.73	2.31
Family relatives from the country of origin	2.73	2.73
Chamber of Medicine in the country of destination	2.73	2.15
Specialised job placement agencies	2.69	2.19
Other NGOs in the destination country	2.61	2.27
Medical chamber from the country of origin	2.56	1.85
Acquaintances of my family in the destination country	2.47	2.42

Note: Min. = 1.00, Max. = 5.00; based on the question: "Please indicate how important the following actors would be for you in getting a job." (1 means completely unimportant and 5 means completely important) – For available options see more in the table.
Source: own research (2022).

to their possible human networks abroad. The results across 96 potential migrants illustrate that 38% of medics have some ties – acquaintances, friends or family members who live abroad.

When asked about the importance of these actors, respondents were presented with 10 different options (including an "other" option). They rated answers on a five-point scale, with 1 being completely unimportant and 5 being completely important and an arithmetic mean was used in the evaluation.

First, we will focus on the actors that medics intend to use to make use of job opportunities in the target country (Tab. 5).

Potential and active migrants perceive institutions as less important than human relations networks, except for job search websites, which are perceived as the most important within both groups, with values over 3.00. Potential migrants then perceive the medical chamber in the destination country as an important actor, and we see that it is on par with most family and other personal relationships. The differences in weights between actors within groups are minimal.

Next, we look at the importance of institutions or family and other personal networks in the potential integration in the destination country.

In Tab. 6, which also divides the set into potential and active migrants, we can see that the first two most important actors are self-initiated (own initiative) and the future employer. Potential migrants attach more importance to institutions, where the help of other institutions and programs besides the future employer is provided. Weak ties seem to be slightly more important for potential migrants, where

Tab. 6 Importance of actors at the beginning of the possible integration in the destination country abroad.

N = 96	Potential migrants N = 70	Active migrants N = 26
Own initiative	3.90	3.92
Future employer	3.36	3.50
Acquaintances from the country of origin	3.10	2.54
My acquaintance from the destination country	3.09	2.88
Help from other institutions and programmes	3.00	2.69
Family relatives in the destination country	2.93	3.19
Compatriot community	2.86	2.81
Chamber of medicine in the country of destination	2.71	2.19
Family relatives from the country of origin	2.64	2.73
Various non-profit organisations	2.31	1.88

Note: Min. = 1.00, Max. = 5.00; based on the question: "Please indicate how important the following actors would be for you in the beginning of your integration in the destination country abroad." (1 means completely unimportant and 5 means completely important) – For available options see more in the table.
Source: own research (2022).

mainly acquaintances from the country of origin and acquaintances from the country of destination dominate. Networks seem to be more important for active migrants, where a slight dominance of strong ties is seen, specifically family relatives in the destination country. Various non-profit organizations and the medical chamber in the destination country are perceived as the least important in both groups.

7. Discussion

The quality of the Czech healthcare system is relatively high, but it faces several serious challenges. These include, for example, the aging medical workforce, a lack of the most advanced modern equipment, overwork and overtime, a heavy administrative and bureaucratic burden on doctors, complicated specialization/attestation process, and comparatively lower salaries - especially in relation to wealthier EU countries. These issues contribute to growing dissatisfaction among medical staff, occasionally culminating in nationwide protests. This overall discontent can, and sometimes does, result in international migration, particularly among younger doctors.

Our research has helped to shed light on some important aspects behind the potential threat of increased emigration of medical students from Czechia. The study confirmed a fairly wide range of reasons underlying the declared preference for emigration (among active migrants), which correspond directly to the previously mentioned "push" factors – further underlining the importance of this finding. These include the migration appeal of destination countries, which offer not only higher individual rewards for doctors' work, but also gaining international insight and contacts, and superior quality of structural and institutional attributes – such as better working conditions, better quality of postgraduate education, and opportunities to work with cutting-edge technology. In summary, the desire for a higher salary was important, but the significance of other structural attributes was equally evident. In this respect, Czechia resembles several other Central and Eastern European countries (e.g. Poland – Krajewski-Siuda et al. 2012; Croatia – Bojanic et al. 2015; and Romania – Suciu et al. 2017).

Compared to a similar study from 2006 (Vavrečková et al. 2006), our findings confirm the important and growing role of German language proficiency, which is closely linked to respondents' preference for migration to German-speaking countries – especially Germany, Austria, and Switzerland. Conversely, the importance of Ireland and the United Kingdom as destination countries appears to have declined significantly, likely due to Brexit – particularly the end of free movement and the accompanying administrative and bureaucratic burdens.

As expected, social networks emerged as a very important factor, taking many different forms among both potential and active migrants. (By the way, an important deterrent to potential emigration is the separation from family, along with all the related emotional and practical implications – see also Vavrečková et al. 2006). No significant differences were found in the preferences for support from so-called strong versus weak ties (in Granovetter's 1973 terms). Compared to the role of institutional actors, social networks did not appear to be either dominant or insignificant. To reach more substantial conclusions in this area, it would be necessary to gain deeper insight into the essence of both of these important factors through more detailed research.

An important perspective in the analysis of the situation is brought by economic concepts of migration. Constant and Massey (2002) highlight that neoclassical economics and the new economics of labor migration propose differing perspectives on both migration duration and motivation. The neoclassical approach aligns more closely with the behavior of "our active migrants" – those intending to stay abroad permanently or return only after more than five years – who are primarily motivated by profit maximization.

On the other hand, the new economics of labour migration in our case aligns more closely with the motivations of our "potential migrants" – those considering temporary migration - most of whom intend to return within five years. This group uses migration as a strategy to overcome market failures in their home country. In addition to higher financial rewards, they seek to gain an international perspective, improve their foreign language skills, and pursue better postgraduate education. These motivations are ultimately framed by their higher expectations for career advancement in Czechia upon their eventual return from abroad.

Although the migration of young doctors from Czechia to other countries does not currently pose a critical threat to the functioning of the healthcare system, it remains a latent issue. This problem could escalate rapidly, particularly in response to the growing global demand to strengthen healthcare systems in many developed countries within the EU and beyond, or due to worsening conditions within Czechia itself.

The first priority should be to stabilize the domestic medical workforce by improving working conditions – specifically through better salaries, modernized equipment, reduced bureaucratic burdens, and an overall decrease in workload. Equally important is encouraging Czech doctors abroad to return, bringing with them valuable social remittances such as new professional experiences and related know-how (e.g. Drbohlav and Dzúrová 2023).

To achieve these goals, a range of targeted recruitment programs must be developed (e.g. Kostelecký et al. 2008). In addition, further research into both the potential and actual migration of doctors is essential to addressing these urgent practical challenges.

8. Conclusions

Below, we summarize the main findings by addressing the research questions:

The findings suggest that the outflow of doctors seeking employment abroad is probably not large or significant enough to seriously impact the Czech healthcare system. Specifically, 17% were identified as "potential" migrants, while only 7% were classified as "actual" migrants.

For potential migrants – and overall – the most influential factor was the opportunity to gain an international perspective and build professional contacts. In contrast, for active migrants, better working conditions were the primary motivator.

We assumed that the majority of future doctors would undertake only temporary work stays abroad, leading primarily to return migration – an assumption that was confirmed by our analysis. Most potential and active migrants expect to eventually return to their home country. However, notable differences between potential and active migrants emerged and should be taken into consideration. The analysis suggests that social ties are often more influential in helping medical professionals secure employment abroad. On the other hand, when it comes to integration, it appears that medical professionals tend to rely more heavily on institutional actors (Massey et al. 1993). Furthermore, the influence of strong and weak ties, as defined by Granovetter (1973), was demonstrated. The results of this analysis confirmed that, overall, it is difficult to clearly determine which type of tie holds greater significance. However, strong ties appear to play a more important role for "active migrants" – those who are firmly committed to migration. In contrast, weak ties seem to exert greater influence among "potential migrants".

Germany plays a key role across the entire research sample, as well as within both respondent groups (potential and active migrants). In addition to Germany, potential migrants also show a preference – albeit to a lesser extent – for "other countries of the world" and the Nordic countries. Among active migrants, Germany remains a top destination, alongside Switzerland, Austria, and the USA/Canada, which are also considered highly attractive migration choices.

A more nuanced perspective on the issue is provided by the results of the regression analysis. These results indicate that having work or study experience abroad, as well as proficiency in the German language, significantly increases the likelihood of declaring plans to emigrate in the future. In contrast, no significant relationships were found between emigration intentions and factors such as gender, academic year, or the specific faculty where respondents studied.

The most important factor discouraging future doctors from working abroad is the prospect of being separated from family and friends – a social barrier. Other key deterrents include concerns about insufficient language skills (a psychological-administrative barrier) and the belief that they will be able to secure a very good job within Czechia (a professional barrier).

This research has its limitations. The sample of participating General Medicine students, although relatively large and drawn from the country's largest university, is limited to four faculties of Charles University and therefore does not represent a nationwide perspective. Second, differences across medical specialties were not captured, and the absence of qualitative data prevents deeper insight into individual motivations and strategies. Furthermore, the migration intentions of young physicians may also change over time due to professional and personal circumstances.

Future research should combine larger, more complex surveys with qualitative approaches to improve our understanding of specialty-specific patterns and the institutional, familial and social factors that influence physicians' behaviour and migration.

References

Akl, E. A., Maroun, N., Major, S., Afif, C., Chahoud, B., Choucair, J., Sakr, M., Schünemann, H. J. (2007): Why are you draining your brain? Factors underlying decisions of graduating Lebanese medical students to migrate. Social Science & Medicine 64(6), 1278–1284, https://doi.org/10.1016/j.socscimed.2006.10.021.

Apostu, S. A., Vasile, V., Marin, E., Bunduchi, E. (2022): Factors Influencing Physicians Migration – A Case Study from Romania. Mathematics 10(3): 505, https://doi.org/10.3390/math10030505.

Bojanic, A., Bojanic, K., Likic, R. (2015): Brain drain: final year medical students intentions of training abroad. Postgraduate Medical Jurnal 91(1076), 315–321, https://doi.org/10.1136/postgradmedj-2014-132908.

Boncea, I. (2013): Medical Brain Drain – A Theoretical Approach. University of Oradea. Faculty of Economics 1(1), 64–71.

Boyd, M. (1989): Family and Personal Networks in International Migrations: Recent Developments and New Agendas. The International Migration Review 23(3), 638–670, https://doi.org/10.2307/2546433.

Brennan, N., Langdon, N., Bryce, M., Burns, L., Humphries, N., Knapton, A., Gale, T. (2023): Drivers and barriers of international migration of doctors to and from the United Kingdom: a scoping review. Human Resources for Health 21: 11, https://doi.org/10.1186/s12960-022-00789-y.

CAMBRIDGE (2022): Common European Framework of Reference for Languages (CEFR). Available at: https://www.cambridgeenglish.org/exams-and-tests/cefr/ (accessed on 19 June 2022).

Constant, A., Massey, D. (2002): Return Migration by German Guestworkers: Neoclassical versus New Economic Theories. International Migration 40(4), 5–38, https://doi.org/10.1111/1468-2435.00204.

Creswell, J. W., Creswell, J. D. (2018): Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (3rd ed.). Thousand Oaks, CA: SAGE Publications. Available online: https://www.ucg.ac.me/skladiste/blog_609332/objava_105202/fajlovi/Creswell.pdf (accessed on 5 November 2023).

ČTK, iDnes.cz (2023): Změňte novelu o přesčasech, vyzývají lékaři. Při stávce mohou omezit tisíce služeb. Available online: https://www.idnes.cz/zpravy/domaci/novela-zakoniku-prace-zdravotnictvi-prescasy-lekari-sluzby.A230918_160600_domaci_vajo (accessed on 5 November 2023).
Czech Medical Chamber (2021): Internal data sources – data provided on the basis of email correspondence.
de Vries, N., Boone, A., Godderis, L., Bouman, J., Szemik, S., Matranga, D., de Winter, P. (2023): The Race to Retain Healthcare Workers: A Systematic Review on Factors that Impact Retention of Nurses and Physicians in Hospitals. INGUIRY: The Journal of Health Care Organization, Provision, and Financing 60, 1–21, https://doi.org/10.1177/00469580231159318.
Docquier, F. (2006): Brain Drain and Inequality Across Nations. IZA Discussion Papers. No. 2440, Boon, https://doi.org/10.2139/ssrn.947463.
Docquier, F., Rapoport, H. (2012): Globalization, Brain Drain, and Development. Journal of Economic Literature 50(3), 681–730, http://dx.doi.org/10.1257/jel.50.3.681.
Dodani, S., LaPorte, M. (2005): Brain Drain from Developing Countries: How can Brain Drain be Converted into Wisdom Gain? Journal of the Royal Society of Medicine 98(11), 487–491, https://doi.org/10.1258%2Fjrsm.98.11.487.
Drbohlav, D., Dzúrová, D. (2023): Social remittances and interpersonal communication: Moldovans in Prague and Turin. Geografie 128(1), 25–48, https://doi.org/10.37040/geografie.2023.002.
Fields, G. S. (1979): Place-To-Place Migration: Some New Evidence. The Review of Economics and Statistics 61(1), 21–32, https://doi.org/10.2307/1924827.
Gkolfinopoulos, A. (2016): The Migration of Greek Physicians to Germany: Motivations, Factors and the Role of National Health Sectors. European Policy Analysis 2(2), 136–157, https://doi.org/10.18278/epa.2.2.8.
Gouda, P., Kitt, K., Evans, D. S., Goggin, D., McGrath, D., Last, J., Hennessy, M., Arnett, R., O'Flynn, S., Dunne, F., O'Donovan, D. (2015): Ireland's medical brain drain: migration intentions of Irish medical students. Human Resources for Health 13(11), 19–29, https://doi.org/10.1186/s12960-015-0003-9.
Granovetter, M. (1973): The Strength of Weak Ties. American Journal of Sociology 78(6), 1360–1380, https://doi.org/10.1086/225469.
Horáková, M. (2007): Zpráva o vývoji pracovních migrací 1993–2006. VÚPSV, Praha, 32 p. Available online: https://katalog.vupsv.cz/Fulltext/vz_240.pdf (accessed on 4 November 2021, in Czech).
Humphries, N., Creese, J., Byrne, J.-P., Connell, J. (2021): COVID-19 and doctor emigration: the case of Ireland. Human Resources of Health 19: 29, https://doi.org/10.1186/s12960-021-00573-4.
IHS Markit (2021): The Complexities of Physician Supply and Demand: Projections from 2019 to 2034. Washington, DC: AAMC. Available online: https://digirepo.nlm.nih.gov/master/borndig/9918417887306676/9918417887306676.pdf (accessed on 4 October 2023).
Ikenwilo, D. (2007): Brain Drain: Painting a Picture for Africa. ATPS Special PaperSeries 34, African Technology Policy Studies Network, Nairobi. Available online: https://atpsnet.org/wp-content/uploads/2017/05/special_paper_series_34.pdf (accessed on 10 January 2025).
International Organization for Migration (IOM) (2014): Mobility of Health Professionals to, from and within the European Union. IOM Migration Research Series 48, Geneva. Available online: https://publications.iom.int/system/files/pdf/mrs48_web_27march2014.pdf (accessed on 12 February 2025).
iROZHLAS (2020): Stát povolal na pomoc mediky. Někteří pracují zatím zadarmo, před státnicemi a na druhé straně republiky. Available online: https://www.irozhlas.cz/zpravy-domov/koronavirus-covid-nemocnice-pracovni-povinnost-studenti_2010230628_sto (accessed on 4 October 2021, in Czech).
Kostelecký, T., Bernard, J., Patočková, V. (2008): Jak změnit "brain drain" v "brain gain". Politiky k ovlivňování migrace vysoce vzdělaných. Výzkumná zpráva pro MŠMT, Sociologický ústav AV. Available online: https://www.soc.cas.cz/images/drupal/soubory/1712091356373d68c659384f0381883.doc (accessed on 12 February 2025, in Czech).
Krajewski-Siuda, K., Szromek, A., Romaniuk, P., Gericke, C. A., Szpak, A., Kaczmarek, K. (2012): Emigration preferences and plans among medical students in Poland. Human Resources for Health 10: 8, https://doi.org/10.1186/1478-4491-10-8.
Lee, E. S. (1966): A theory of migration. Demography 3(1), 47–57, https://doi.org/10.2307/2060063.
Levitt, P. (1998): Social Remittances: Migration Driven Local-Level Forms of Cultural Diffusion. The International Migration Review 32(4), 926–948, https://doi.org/10.2307/2547666.
Light, I., Bhachu, P., Karageorgis, S. (1989): Migration Networks and Immigrant Entrepreneurship. In California Immigrants in World Perspective: The Conference Papers. Paper 1. Institute for Social Research, Los Angeles: University of California. Available online: https://escholarship.org/content/qt50g990sk/qt50g990sk.pdf (accessed on 5 October 2023).
Massey, D. S. (1988): Economic Development and International Migration in Comparative Perspective. Population and Development Review, 14 (3), 383–413, https://doi.org/10.2307/1972195.
Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., Taylor, J. E. (1993): Theories of International Migration: A Review and Appraisal. Population and Development Review 19(3), 431–466, https://doi.org/10.2307/2938462.
Ministerstvo školství, mládeže a tělovýchovy (MŠMT) (2024): Výkony vysokých škol – počty studentů a absolventů. Available online: https://statis.msmt.cz/statistikyvs/vykonyVS1.aspx (accessed on 10 June 2025, in Czech).
OECD (2021): OECD data: Health. Available online: https://data.oecd.org/health.htm; profile Health resources (accessed on 22 September 2021).
Pszczółkowska, D., Bojarczuk, S., Duszczyk, M., Matuszczyk, K., Szyszkowska, E. (2024): Did Covid-19 make things worse? The pandemic as a push factor stimulating the emigration intentions of junior doctors from Poland: A mixed methods study. PLoS One 19(4): e0301757, https://doi.org/10.1371/journal.pone.0301757.
Sedláková, K. (2015): Cesta za uznáním: Důvody k migraci do Německa u budoucích absolventů pražských

lékařských fakult Univerzity Karlovy. Master's thesis, Faculty of Social Sciences, Charles University, Prague (in Czech).
Schumann, M., Maaz, A., Peters, H. (2019): Doctors on the move: a qualitative study on the driving factors in a group of Egyptian physicians migrating to Germany. Globalization and Health 15: 2, https://doi.org/10.1186/s12992-018-0434-x.
STATISTA (2024). Medical graduates per 100,000 inhabitants in Europe 2006–2021. Statista. Available online: https://www.statista.com/statistics/283241/medical-graduates-per-100-000-inhabitants/ (accessed on 6 July 2025).
Suciu, Ş. M., Popescu, C. A., Ciumageanu, M. D., Buzoianu, A. D. (2017): Physician migration at its roots: a study on the emigration preferences and plans among medical students in Romania. Human Resources for Health 15: 6, https://doi.org/10.1186/s12960-017-0181-8.
Tjadens, F., Weilandt, C., Eckert, J. (2013): Mobility of Health Professionals: Health Systems, Work Conditions, Patterns of Health Workers' Mobility and Implications for Policy Makers. Berlin: Springer.
Vavrečková, J., Gazdagová, M. (2005): Poptávka po českých kvalifikovaných odbornících v zahraničí: Kvalitativní průzkum řízených rozhovorů s představiteli soukromých zprostředkovatelen práce. Praha: VÚPSV.
Vavrečková, J., Dobiášová, K., Hlinicová, H. (2006): Migrační sklony lékařů a studentů medicíny. Terénní šetření, Svazek II. Praha: VÚPSV.
Vavrečková, J. (2007): Mezinárodní migrace odborníků a riziko odchodu českých lékařů do zahraničí. FÓRUM sociální politiky 1(1), 6–10.
Vavrečková, J. et al. (2007): Riziko odlivu lékařů a odborníků IT/ICTV z České republiky do zahraničí. Praha: VÚPSV. Available online: https://katalog.vupsv.cz/fulltext/vz_214.pdf (accessed on 22 September 2021, in Czech).
World Health Organization (2016): Global strategy on human resources for health: Workforce 2030. Geneva: WHO. Available online: https://iris.who.int/server/api/core/bitstreams/3ef6ee65-42fa-4d2b-9c75-d55b2df17f9a/content (accessed on 10 March 2025).
Żuk, P., Żuk, P., Lisiewicz-Jakubaszko, J. (2019): Labour migration of doctors and nurses and the impact on the quality of health care in Eastern European countries: The case of Poland. The Economic and Labour Relations Review 30(2), 307–320, https://doi.org/10.1177/1035304619847335.

Family planning programs in South Asia: Developments and effectiveness

Sylvie Chrtková*, Jiřina Kocourková

Charles University, Faculty of Science, Department of Demography and Geodemography, Czechia
* Corresponding author: sylvie.chrtkova@natur.cuni.cz

ABSTRACT

This paper examines the development and effectiveness of Family planning programs (FPPs) in three South Asian countries – India, Pakistan, and Bangladesh – that were among the first to adopt such initiatives in response to rapid population growth. The analysis focuses on the evolution of program objectives, implementation strategies, contraceptive prevalence, and financial support, offering comparative insights into national contexts and outcomes. Despite shared goals – namely reducing fertility rates, expanding contraceptive access, and improving reproductive health – the trajectories of FPPs differ significantly across the three countries. India pioneered early implementation and initially emphasized demographic control, including controversial measures such as compulsory sterilization. Bangladesh focused on community-based approaches and benefited from strong donor support, while Pakistan faced persistent challenges due to political instability, religious opposition, and underfunding. The study finds that India and Bangladesh have reached contraceptive prevalence rates comparable to global averages, largely due to sustained investments, community outreach, and integration with public health systems. In contrast, Pakistan continues to lag behind, with limited progress and ongoing dependence on foreign aid. Financial commitment from national budgets has proven to be a key factor in the long-term success of FPPs, as demonstrated by India and Bangladesh. The comparison highlights that while structural design matters, program success ultimately depends on cultural sensitivity, gender equity, and adaptability to local conditions. Future policies should prioritize inclusive reproductive education, sustainable financing, and engagement of both women and men to strengthen the impact of family planning programs on population dynamics and socio-economic development.

KEYWORDS

family planning; India; Pakistan; Bangladesh; contraceptive prevalence

Received: 30 April 2025
Accepted: 10 October 2025
Published online: 10 December 2025

Chrtková, S., Kocourková, J. (2025): Family planning programs in South Asia: Developments and effectiveness.
AUC Geographica 60(2), 269–284
https://doi.org/10.14712/23361980.2025.24

1. Introduction

Rapid population growth has a high impact on the success of sustainable development in the world, and although many regions are experiencing a gradual reduction in population growth rates, concerns about global overpopulation persist (United Nations 2024). The world population is projected to reach 8.5 billion by 2030 and 9.7 billion by 2050 (World Bank 2018, United Nations 2024). The most populous regions of the world are East and South-East Asia, with both regions exceeding 2 billion inhabitants and representing the regions with the fastest population growth. The continuing population growth in these regions poses a threat to their economic, health and social growth. These are countries whose infrastructure and instability are unable to provide adequate living standards for their populations, widening the gap between the Western and Eastern worlds (United Nations 2022).

An effective mechanism for slowing population growth is to reduce fertility levels, which should be achieved through Family Planning programs (FPP). FPP are a structured public health initiative designed to provide individuals and couples with the information, services, and resources they need to make informed decisions about reproduction and childbearing. These programs typically include contraception education, reproductive health services, counseling, and access to contraceptive supplies (Donaldson and Tsui 2018). They aim to lower fertility, promote women's health, diversify contraceptive options, and strengthen reproductive health education (United Nations 2022).

Family Planning programs were created to reduce population growth and alleviate concerns about the unsustainable development of the world after World War II. The countries in South Asia, mainly India, Pakistan and Bangladesh, posed the greatest threat, which is why the programmes were first implemented in these countries (Robinson and Ross 2007).

While the programs slowed growth to some degree, their long-term effectiveness varied, and these countries remain among the fastest growing worldwide (United Nations 2024).

The paper examines the topic of family planning programmes in the countries concerned through several factors and sets two main objectives, namely:

1. Description and comparison of the Family Planning programs their development and main ideas focusing on their similarities and differences. This will show how the plans were gradually adapted or not adapted and what results this led to (Visaria and Ved 2016; Robinson 2023).
2. Comparing the Effectiveness of Family Planning Programs using an indicator of contraceptive prevalence and taking into account financing of FPP. Mapping program trajectories clarifies which approaches promoted or hindered progress, while contraceptive prevalence offers a comparable metric of success (FP2030, 2023). Given that financial resources influence service quality and method mix, we relate changes in contraceptive prevalence to funding levels and sources.

India, Pakistan, and Bangladesh launched their programs in similar periods and under global policy frameworks, analyzing differences in planning and implementation may explain variations in outcomes. The findings can contribute to understanding the impact of family planning policies on socioeconomic development and the effectiveness of government decisions.

2. Background

Family planning programs refer to organized efforts, typically led or supported by governments and international organizations, aimed at enabling individuals and couples to control the number and spacing of their children (Cleland et al. 2006). These programs provide access to contraceptive methods, reproductive health education, and related services, with the broader goal of improving health outcomes, empowering women, and supporting sustainable development.

The need to introduce family planning programs arose from the recognition that the extremely rapid population growth, which began to take place in the developing world since the mid-twentieth century, posed a serious threat – primarily due to its potential consequences for economic and social development. It has been presented as a possible brake on society that would cause a slowdown in the growth of living standards for most people in low-income countries (Seltzer 2002). World population conferences have played a crucial role in shaping family planning programs by fostering international dialogue, building political consensus, and redefining policy priorities (Sinding 2000). The 1974 Bucharest conference marked a shift by advocating for the integration of family planning into broader development efforts, challenging the Western emphasis on population control. In 1984, the Mexico City conference, influenced by Cold War geopolitics, promoted private sector involvement and reframed population growth as an obstacle to development. The most transformative was the 1994 Cairo conference, which redefined family planning as a matter of reproductive rights and individual autonomy, emphasizing women's empowerment, education, and access to voluntary contraception – principles that continue to guide global policy today.

Thus, the main objective was to reduce fertility levels, which would lead to a gradual population reduction, economic and social growth and consequently to an improvement in the standard of living of low-income populations (Seltzer 2002). These programmes target three basic premises – demographic, health and human rights – that essentially imply family planning. The demographic rationale primarily addresses the consequences of the delayed onset of the demographic transition, particularly the persistence of high fertility and infant mortality rates. The health perspective reflects the inadequate level of health care with a strong focus on maternal and childcare, and the human rights perspective focuses on women's and men's rights in the context of family planning. If all these foundations are pursued in a balanced manner, they can contribute to the overall improvement of societal functioning and accelerate national development.

Over the decades, the rationale behind family planning programs has evolved (Bongaarts and Sinding 2011). While demographic and economic motivations were dominant in the early phases, later decades brought increased emphasis on women's rights, health equity, and individual reproductive autonomy. Today, family planning is widely recognized as a critical component of global health and development strategies.

The potential impact of family planning programs, which have been widely implemented across the developing world since the 1960s, has been a central topic of numerous studies (Bongaarts 2025; Bongaarts 2020; Günther and Harttgen 2016; Robinson and Ross 2007; Casterline and Sinding 2000; Cleland et al. 1994; Pritchett 1994). A World Bank review conducted in the early 1990s concluded that these programs had been successful as regards to address the unmet need for contraception in many countries and recommended their expansion (World Bank 1993).

Although some studies have questioned the significance of family planning programs, arguing that they played only a minor role in fertility decline (Pritchett 1994), other research suggests that their effect was indeed significant, though relatively modest (Günther and Harttgen 2016). In countries where high-quality programs were implemented, contraceptive use increased, and fertility rates declined (Robinson and Ross 2007). More recent evidence continues to emphasize that family planning programs have been among the most important determinants of fertility decline (Bongaarts 2020). However, current findings also highlight the importance of the broader context in which these programs operate and stress the need for alignment with national policies and local conditions (Finlay 2024).

Despite the scope of family planning programs has broadened over time in low- and middle-income countries, the primary focus has remained on the availability of contraception and the extent to which it meets the growing demand for birth control that has emerged over recent decades (Kantorová and Bongaarts 2024). Therefore, this paper focuses on trends in contraceptive use in relation to family planning initiatives. The analysis focuses on three countries in South Asia, as this region was among the first to implement such programs, allowing for an in-depth assessment of their long-term development and impact.

3. Methodology

For the purpose of comparison of the Family planning programmes in the surveyed countries, the method of literature review was used. A literature review is a systematic process of locating, collecting, evaluating and synthesizing expert sources (Snyder 2019). For it allows, based on available information and the results of previous empirical studies, a comprehensive assessment and comparison of historical developments in terms of political, economic and social context. In the search analysis, the objectives were first set and then used as a basis for the selection of relevant sources related to Family planning programs. Given the historical nature of the research, it was determined that the source had to have been published between the 1960's and the present. In addition, key terms were then identified based on which information was selected. The synthesis of information was based on the chronology of the programmes, for the purpose of comparing and finding differences and similarities in the different FPP's.

The comparison of similarities and differences between the programmes was based on the identification of thematic headings, which were answered for each country based on selected sources. The headings were:

1. Government policies – positions in the government system
2. Objectives – the objectives set by the programs
3. Focus – country specific requirements or objectives
4. Availability of services – health facilities, availability of contraception
5. Achievements – significant advances related to FPP
6. Challenges – limits that countries are addressing
7. Cultural and social factors – attitudes towards contraception and the influence of religion
8. Population – demographic characteristics of the population

On the basis of these headings, a table was drawn up comparing all the countries studied across all approved programmes from the 1960's onwards.

The success of Family planning programs was further compared focusing on the aspect of contraceptive prevalence rate (CPR) (United Nations 2024). Contraceptive prevalence rate is one of the key indicators of program success and is defined as the percentage of women aged 15–49 who are currently using, or whose sexual partner is using, at least one method of contraception, regardless of method being used (Kantorová and Bongaarts 2024) i.e:

$$\text{CPR} = \left(\frac{\textit{Number of women aged } 15-49 \textit{ using contraception}}{\textit{Total number of women aged } 15-49 \textit{ who are married or in union}}\right) \times 100$$

Prevalence of contraception includes both modern and traditional methods of contraception. Modern ones include Female and male sterilization, Intrauterine devices (IUDs), Implants, Injectables, Oral contraceptive pills, Male and female condoms, Emergency contraception, Lactational amenorrhea method (LAM). Traditional methods are Rhythm (calendar-based) method and Withdrawal.

To illustrate the relationship between contraceptive prevalence and the share of government budget funding for family planning programs, data from the comparison of family planning funding in Chapter 4.7 (Tab. 2) was used. The share of FPP funding was divided into five categories: Exclusively donors, mainly donors (low share of the state budget), Combination of donors and state budget, mostly from state budget, mainly state budget. The category Mainly state budget was chosen as the reference value (100%) representing the highest level of government involvement. Other levels were scaled relative to this value – from 0% (for Exclusively donors) to 75% (e.g. Mostly from state budget), creating a quantitative axis of government involvement. The resulting graph was supplemented with contraceptive prevalence rates in selected countries and time periods.

Data for comparison of contraceptive prevalence were drawn from the World Bank database. The database guarantees the quality and integrity of the published data and continuously updates the available data. It works on the principle of collecting data from the statistical systems of member countries. Contraceptive prevalence is based on data from domestic surveys. It is precisely in data collection that the greatest limitations are found. The quality of data collection in the countries surveyed may not be entirely reliable. In the case of contraceptive prevalence, this most often involves discrepancies in respondents' understanding of questions relating to the definition of contraceptive use. It is therefore important to take these aspects into account when interpreting the data (World Bank 2023).

4. Results

4.1 Family planning programs in India

India was the first country in the world to officially launch a national family planning program in 1952 (Ledbetter 1984). Its creation was closely linked to concerns about a population explosion, which was seen as a major obstacle to economic and social development (Fárek 2012). The programs have changed several times over the decades, from an early behavioral approach to a client-centered model and massive campaigning to decentralized and participatory strategies that focused more on reproductive health and child survival. A critical assessment of these changes reveals tensions between technocratic, directive models and later efforts to respect individual rights and needs.

The first phase (the 1950's) was characterized by an emphasis on education and support for "responsible parenthood". In practice, this meant that the government assumed that behavior would change through information rather than through the actual availability of methods. Initially, the Ministry of Health and Family Welfare only approved the "rhythm method" because other forms of contraception were rejected for religious and cultural reasons (Connelly 2006; Ledbetter 1984). However, this approach failed because it did not take into account structural barriers, particularly low literacy among women and limited access to institutions. Critics point out that instead of changing behavior, it caused frustration and mistrust (Ram and Ram 2021).

In the 1960's, policy began to shift toward a client-centered approach, with the government expanding services in health facilities and gradually introducing intrauterine devices, sterilization, and other methods. Infrastructure was strengthened, and the program gained the support of international organizations, particularly the United Nations. This change had a positive impact on awareness of family planning, but fundamental problems remained: a shortage of trained personnel, irregular supplies of contraceptives, and unequal access to services, especially in rural areas (Robinson and Ross 2007). The program was criticized for focusing too much on the urban population and not taking social inequalities sufficiently into account.

A major turning point came in the 1970's with the so-called camp approach, i.e., mass sterilization campaigns. These had the direct support of Indira Gandhi's government, and especially during the period of emergency (1975–1977), compulsory sterilizations were introduced and quotas were set for individual states (Robinson and Ross 2007). These practices led to widespread human rights violations, including coercion of the poorest sections of the population (Kongawad and Boodeppa 2014). In the short term, there was a sharp increase in the number of sterilization procedures and an improvement in the population situation, but in the long term, this approach seriously undermined public confidence in state health programs and created a stigma associated with family planning (Pradhan and Dwivedi 2019).

Following this crisis, policy in the 1980's and 1990's gradually shifted toward voluntary programs and a broader context of overall family health. The goal was to link family planning to a broader concept of social care with a focus on maternal and child health (Kongawad and Boodeppa 2014). A significant change was the transition to the so-called community needs assessment approach in 1996, also known as the "no-target approach". This model aimed to decentralize decision-making, strengthen regional self-government, and take into account the specific needs of communities (Pradhan and Dwivedi 2019). This was a step forward in terms of participation, but its implementation was uneven – particularly in poorer states of India, where access to quality services remained limited.

From the late 1990's, the program began to focus more on reproductive health and child survival, linking family planning to the goals of reducing maternal and infant mortality. This approach was in line with international development agendas (United Nations 1994, Millennium Development Goals United Nations 2000) and aimed to integrate family planning into the framework of human rights and development (Chaurasia 2014). At the same time, however, Indian programs continued to focus heavily on demographic indicators, i.e., reducing fertility and achieving replacement-level fertility, highlighting the continuing tension between declared humanistic goals and technocratic practice (United Nations Population Fund n.d.).

India's family planning program underwent several phases from the 1950's onwards, each revealing its own compromises and limitations. The early behavioral model based on information campaigns and the "rhythm method" proved ineffective in an environment of low literacy and limited access for women to institutions, widening the gap between policy intentions and actual behavioral change (Connelly 2006; Ledbetter 1984; Ram and Ram 2021). The client-centered approach of the 1960's expanded the range of methods available in health facilities, but it was hampered by staff shortages, irregular supplies, and an urban bias that excluded rural and poorer populations (Robinson and Ross 2007). The subsequent approach based on sterilization camps in the 1970's brought a short-term increase in performance, but at the same time seriously damaged the legitimacy of the program, stigmatized family planning, and undermined public confidence in the state (Kongawad and Boodeppa 2014; Pradhan and Dwivedi 2019). Decentralization through the community needs assessment model (the so-called target-free approach) after 1996 strengthened the influence of local residents and emphasized participation, but its results were uneven and limited by weak capacities in poorer regions (Pradhan and Dwivedi 2019). The later reproductive health and child survival framework attempted to link family planning with rights-based agendas, but monitoring remained tied to demographic targets. Gender imbalances persisted, as responsibility for contraception rested almost exclusively with women, while the proportion of male sterilizations remained minimal (Chaurasia 2014; Kongawad and Boodeppa 2014; United Nations Population Fund n.d.). Persistent institutional weaknesses, a lack of qualified personnel, poor service quality, and paternalistic management undermined the effectiveness of the programs over the long term.

These experiences show that a focus on quantitative targets and technocratic fertility control was insufficient. The key lesson is to prioritize service quality, diversification of contraceptive methods, and greater involvement of men over simply monitoring demographic quotas. Although the National Reproductive Health and Family Planning Strategy has emphasized girls' rights and education since 2015, India's approach remains largely technocratic and focused on birth control rather than comprehensive reproductive health (United Nations Population Fund n.d.; Pradhan and Dwivedi 2019).

An important achievement of India's family planning program has been the creation of a stable and relatively well-functioning institutional system. Despite numerous challenges, such as limited diversity of contraceptive methods and persistent inequalities between urban and rural areas, India has managed to maintain continuity of services and build one of the most extensive family planning infrastructures in the Global South (Chaurasia 2014; United Nations Population Fund n.d.). However, recent studies point out that the approach remains largely technocratic and overly focused on reducing fertility rather than comprehensively ensuring reproductive health (Pradhan and Dwivedi 2019).

Despite overall successes, the structure of contraceptive methods in India remains unbalanced. The long-term dominance of female sterilization and minimal use of male methods point to a persistent gender imbalance and the "technocratic" way in which family planning policy is implemented (Chaurasia 2014; Green 2018; UNFPA India n.d.). For the present, this implies a need to systematically promote voluntary and reversible methods, expand male contraceptive options, and strengthen the quality of counseling to reduce dependence on one-off performance targets (Pradhan and Dwivedi 2019).

4.2 Family Planning programs in Pakistan

Pakistan was one of the first countries in South Asia to recognize rapid population growth as one of the key challenges to development. Since the 1950's, the government has formulated explicit antinatalist policies, but by international standards and even compared to India and Bangladesh, the effectiveness of these programs has been relatively low (Sathar 2007). Although the country adopted a series of five-year plans that included population policy, the results were undermined by structural factors: political instability, insufficient funding, gender inequalities, and the strong influence of conservative religious norms (Wazir et al. 2013; Goujon et al. 2020).

The first generation of programs in the 1960's was based on the assumption that women would seek out family planning services on their own. The distribution of modern methods, particularly intrauterine devices, was supported by foreign donors (Bean and

Bhatti 1969). However, this approach proved ineffective. The main problem was rural areas. Women in these areas did not receive sufficient information about family planning program options and were not mobile enough to travel to services. As a result, access to services was minimal in practice (Hakim and Mahmood 2001).

In the 1970's, programs began to strengthen the community approach. The inspiration was the Alma-Ata Conference (1978), which emphasized the importance of primary health care. Community centers were established in villages, and the government invested in training health workers. Although this model was better suited to the needs of rural areas, its implementation was limited by a lack of resources and poor coordination. Critics point out that the state invested in tertiary care in cities, while primary services in rural areas remained underfunded (Wazir et al. 2013).

Unlike India, which experimented with sterilization in the 1970's, or Bangladesh, where cooperation with non-governmental organizations prevailed, Pakistan's family planning program suffered most from political instability. Changes in government led to frequent interruptions or suspensions of programs. Under the Zia-ul-Haq government, the family planning agenda was even almost abandoned (Hakim and Mahmood 2001). Weak political anchoring meant that long-term strategies were not implemented and the demographic dynamics of rapid population growth continued.

Change came in the 1990's, when Prime Minister Benazir Bhutto actively promoted reproductive health. Her participation in the Cairo Conference (1994) led to the adoption of programs emphasizing human rights and gender equality. A key step was the creation of the Lady Health Workers (LHW) program, which focused on home visits, contraceptive distribution, and basic maternal and child care. The LHW program not only improved contraceptive distribution and maternal health services, but also played a role in empowering women in their local communities. By providing direct services in households, the program increased awareness of reproductive health and strengthened women's decision-making abilities, although its coverage remains inadequate in some rural areas (Hafeez et al. 2011; Sathar 2007). The program continues to operate today and is considered the most effective reproductive health policy tool in the country (World Health Organization n.d.; Sathar 2007).

As in other countries in the region, Pakistan has gradually shifted its focus from antinatalist goals to a broader concept of reproductive health and child survival. This shift has led to the introduction of programs focused on prenatal and postnatal care, immunization, HIV/AIDS prevention, and combating gender-based violence (Sathar 2007). However, persistent barriers remained: low female literacy, early marriage, and limited economic opportunities for women, which significantly undermined the effectiveness of the programs (Goujon et al. 2020).

After 2000, the programs were decentralized and responsibility shifted from the federal to the provincial level. The aim was to tailor services to local needs. In practice, however, it became apparent that provincial capacities were uneven: for example, Punjab was able to provide relatively accessible services, while Balochistan lagged significantly behind (Sathar 2007). Decentralization thus exacerbated regional inequalities rather than reducing them.

The 2017 census confirmed high population growth. The government responded by creating federal and provincial working groups, approving a national action plan (2019), and launching the "Tawazun" campaign (2021). The national action plan had clearly defined goals: to increase contraceptive prevalence, reduce population growth to 1.1%, and reduce total fertility. The goal is to be achieved by 2030. The Tawazun campaign links family planning with sustainable development and human rights (Ministry of National Health Services, Regulations & Coordination 2021.). However, experts warn that without massive investment in girls' education and women's empowerment, these goals will remain largely declarative (Goujon et al. 2020).

Pakistan's original technocratic model, based on promoting intrauterine devices and centralized management, proved culturally and socially ineffective because it ignored local norms, relied on intermittent political support, and neglected primary care in rural areas (Hakim and Mahmood 2001; Wazir et al. 2013; Robinson and Ross 2007). Weak political anchoring and frequent discontinuity of individual initiatives further undermined the continuity necessary for demographic and health policy (Sathar and Zaidi 2010). In contrast, the community-based model of health workers visiting households has shown that direct services can increase women's awareness of and access to contraception and empower them, but its coverage remains inadequate and uneven across provinces (Hafeez et al. 2011; Nishtar 2010). The decentralization process has further fragmented administration, leading to significant regional disparities in the quality and availability of services. Deeper barriers include high rates of consanguineous and early marriage, low literacy, and limited participation of women in the labor market, which systematically undermine their autonomy in reproductive health (Zaidi et al. 2019; Sathar and Kazi 2000; Sathar and Royan 2013). These factors explain why, even when methods are available, there is no significant increase in their use, with demand further dampened in the long term by limited communication channels – mainly leaflets or radio spots – that do not meet the needs of the population (Population Services International Pakistan 2021; World Health Organization 2015).

4.3 Family planning programs in Bangladesh

Bangladesh is a unique case in South Asia. Despite extreme poverty and high population density, it has achieved a rapid decline in fertility since the 1970's, from more than six children per woman to approximately 2.1 today (United Nations Population Fund 2023). This success is linked to a combination of strong state involvement, active participation by non-governmental organizations, and a community-based approach (Cleland et al. 1994; Alam 2017).

After gaining independence in 1971, the government adopted its first five-year plan with the goal of significantly reducing the fertility rate. The goals were unrealistically ambitious (Amin et al. 1987), but they created a basic institutional framework. Of key importance was the MCH-FP Planned Parenthood project in Matlab, launched in 1977 by the ICDDR. The program combined intensive home visits, contraceptive distribution, and basic health care. Studies showed that this community-based approach could reduce fertility even without significant improvements in socioeconomic conditions (Phillips et al. 1982; Phillips et al. 1996).

Unlike Pakistan, Bangladesh soon opened its doors to nongovernmental organizations such as BRAC and Grameen. These organizations integrated health services with microcredit, education programs, and women's empowerment (Alam 2017). The model combining government and NGO funding allowed for more flexible implementation and wider acceptance in rural communities.

The Cairo Conference (1994) prompted Bangladesh to adopt the Health and Population Strategy Program (HPSP), which combined family planning with a broader concept of reproductive and sexual health (Jahan 2007; World Health Organization 2004). The Bangladeshi government's thinking on family planning thus shifted toward an emphasis on the reproductive rights of the population. A system was created that provided not only contraception but also prenatal care, safe abortions, infectious disease prevention, and nutrition for mothers. This change led to greater legitimacy for the programs and increased their long-term sustainability.

The decline in fertility in Bangladesh is often referred to as a "demographic paradox" because the country achieved a significant reduction in fertility despite its low level of development and political instability (Cleland et al. 1994). The institutionalization of a robust community platform, home visits, satellite clinics, and community workers played a key role in ensuring contact with even the poorest segments of the population. Long-term cooperation between the state and non-governmental organizations was also important, maintaining the availability of services even during periods of weaker state administration and bringing innovation to the integration of family planning with health and social care (Phillips et al. 1988; Phillips et al. 1996; Mercer et al. 2004). In addition, reforms under the HPSP/HNPSP linked contraception with maternal and newborn care, preventive services, and counseling, thereby increasing the legitimacy and continuity of the program (Jahan 2007; World Health Organization 2004). Improvements in human development also contributed to the success: increased school enrollment of girls, a decline in adolescent marriages, and greater participation of women in society strengthened their autonomy in decision-making and the demand for smaller families (Bairagi 2001; Shabuz et al. 2022; United Nations Population Fund 2023). In sharp contrast to Pakistan, where political instability, weak primary care, and the limited status of women have hindered progress, the combination of community work, the flexibility of non-governmental organizations, and improvements in human development in Bangladesh has created a "social space" that has outweighed religious conservatism in determining outcomes (Sathar and Kazi 2000; Hardee et al. 2014).

Among Bangladesh's most important achievements is the creation of a strong, functional network of community centers that ensures the availability and continuity of family planning services even in rural and economically disadvantaged regions (Phillips et al. 1988; Cleland et al. 1994). Another notable result is the decline in adolescent marriages, which has contributed to a slowdown in population growth and improved reproductive health indicators, in line with broader goals of women's empowerment and gender equality (Shabuz et al. 2022; UNFPA 2024).

Currently, family planning programs are based on sustainable development goals. The key framework is Family Planning 2030, which emphasizes universal access to reproductive services, support for young women, and prevention of gender-based violence (UNFPA 2024; Family Planning 2030). At the same time, the programs are integrated into the broader National Sustainable Development Strategy (NDS) until 2044, which links population policy with environmental and social goals.

Critical evaluation shows that although the ambitious goals of individual plans were essentially unrealistic, they created a framework in which new approaches and innovations could emerge (Cleland et al. 1994). Of particular significance was the Matlab experiment in Bangladesh, which clearly demonstrated the effectiveness of a community-based and personalized approach to family planning and became an internationally recognized model (Phillips et al. 1988). Non-governmental organizations also played a key role here, as without their active involvement, the state would not have been able to ensure sufficient coverage and availability of services (Mercer et al. 2004). The shift towards an emphasis on reproductive rights significantly increased the legitimacy of the programs and contributed to their gradual institutionalization (Hardee et al. 2014). Nevertheless, certain limitations remain, including low service quality and persistent gender inequalities, which continue to undermine their long-term effectiveness despite the successes achieved so far (Sathar and Kazi 2000).

Bangladesh has succeeded despite political instability thanks to three factors: (i) a community-based door-to-doormodel (Matlab, FWA, satellite clinics) emphasizing continuous contact between households and providers (Phillips 1982; Phillips 1996; World Health Organization 2004; Jahan 2007), (ii) partnerships between the state and non-governmental organizations (BRAC, Grameen) linking family planning, microcredit, and women's education (Mercer et al. 2004; Alam 2017), and (iii) the gradual empowerment of women through girls' school attendance and later marriage (Bairagi 2001; United Nations Population Fund 2023). In contrast, in Pakistan, programs suffered from insufficient political support, inadequate primary care funding, and uneven coverage after decentralization; even the strong Lady Health Workers program was not sufficiently expanded (Sathar 2007; Hafeez et al. 2011; Wazir et al. 2013; Goujon et al. 2020).

4.4 Comparative analysis of Family planning programs

Family planning programs in India, Pakistan, and Bangladesh began operating in roughly the same period (second half of the 20th century) and given the considerable political and economic proximity of these countries, there are many commonalities in the programs (Tab. 1). All three countries have committed their Family planning programs to be part of the health system, leading to significant benefits. When the Family planning system and the health system function as a whole, it is easier to achieve the set goals and to change the settings of the different tools quite dynamically (Perera et al. 2024). At the same time, this brings more detailed mapping of the impact of Family planning programs on the health level of the population and consequently on changes in fertility and mortality (Robinson and Ross 2007).

Other common features are international aid and a community-based approach. In all countries, Family planning programs have been established with the support of major powers, as the World Health Organization (WHO) and the United Nations Population Fund (UNFPA). The community-based approach has emerged over the years in family planning programs in all three countries and has been a great success everywhere and has been considered the most effective tool for disseminating information (Sultan et al. 2002; Bhatia et al. 2024; Memon et al. 2024). India was the first to include the community approach in its programs, followed by Pakistan. Bangladesh introduced the community-based approach in its third programme, due to the continuity of effectiveness

Tab. 1 Common features and differences of Family planning programs in India, Pakistan and Bangladesh.

Common features	India	Pakistan	Bangladesh
1. Government commitments – all three countries have committed to Family planning programs as part of their health system 2. International aid and involvement NGO 3. Community approach 4. Linking health and care with Family planning programs in one sector			
Differences	**India**	**Pakistan**	**Bangladesh**
Government policies and priorities	Voluntariness in PPR, decentralized system	High government instability, decentralised system	Priorities within the socio-economic development, centralized system
Objectives	Slowing population growth, increased use of contraceptives, economic development	Slowing population growth, increased use of contraceptives	Slowdown in population growth, increase in contraceptive use, decrease in adolescent marriages
Focus	Emphasis on Family planning services (sterilization), initiatives for young families	Emphasis on sexual and reproductive health education and awareness	Emphasis on community engagement and women's empowerment
Availability of services	Extensive system, public and private sector involvement, strong promotion	Insufficient resources and very limited infrastructure	Actively expanding service distribution system, strong promotion
Achievements	Stable and quality system of functioning of Family planning Programs	Empowering women in the Lady Health Workers program	Strong functional network of community centres
Challenges	Low diversity of contraceptive methods	High illiteracy rates, especially among women	Women's educational attainment and lowlow socioeconomic status
Cultural and social factors (contraception)	Different cultural and social practices are reflected in attitudes to contraceptive use and family life issues		
Population	Different population sizes with different characteristics affect the overall dynamics of society		

Source: own processing.

in Pakistan, but its performance surpassed the quality of community-based functioning in Pakistan and became the most successful compared to the countries studied (Robinson and Ross 2007).

Despite these commonalities, the programs of each country differed in their outcomes and priorities. India faces limited diversification of contraceptive methods, as sterilization remains dominant despite relatively high prevalence (Chaurasia 2020). Pakistan faces persistent challenges in the form of female illiteracy and gender inequality (Ataullahjan 2018), while Bangladesh's main obstacle is the low socioeconomic status of women (Shabuz et al. 2022).

Government priorities have also differed. India has gradually emphasized voluntariness and sought to distance itself from the forced sterilization campaigns of the 1970's (Ministry of Health and Family Welfare n.d.). This shift has allowed couples to make more autonomous decisions, supported by a relatively extensive infrastructure. Bangladesh made women's empowerment and community participation central to its programs, based on the assumption that improvements in education and gender equality would stimulate socioeconomic development (World Health Organization 2018). Pakistan, on the other hand, is hampered by political instability and weak continuity. For decades, its programs relied mainly on leaflets and radio broadcasts, with broader media campaigns only emerging recently (Population Services International Pakistan 2021; World Health Organization 2015).

The availability of services shows several contrasts. India has built the largest network of clinics, health centers, and community centers, involving both the public and private sectors. Bangladesh, although less stable institutionally, is rapidly expanding its system through cooperation between the government and non-governmental organizations (Bates et al. 2003; World Health Organization 2015). Pakistan remains the weakest player, facing serious infrastructure shortages and insufficient resources (Robinson and Ross 2007).

4.5 Improvements in prevalence of contraception

The most rapid increase in contraceptive prevalence occurred in India and Bangladesh in the 1970s and 1980s. In India, the prevalence of contraception until the early 1970's was around 15%. At that time, only a very limited range of contraceptives was available, and the only officially recognized method from a religious point of view was the "rhythm" method. This method is not very effective on its own and requires regularity to work. For many women, it was therefore not an option they wanted to use (Ledbetter 1984). By 1980, the rate had risen to 35%. This means that in just 10 years, India managed to increase the prevalence of contraception by 20% points. The increase in the prevalence of contraception was mainly due to the introduction of compulsory sterilization under the Fourth Five-Year Plan (Kongawad and Boodeppa 2014, Robinson and Ross 2007). Between 1980 and 1990, there was a further increase of about 10 percentage points, reaching 45% in 1990. In this decade, the increase was slower. In the 1990s, there was a slight decline to 40% in India. At the beginning of the

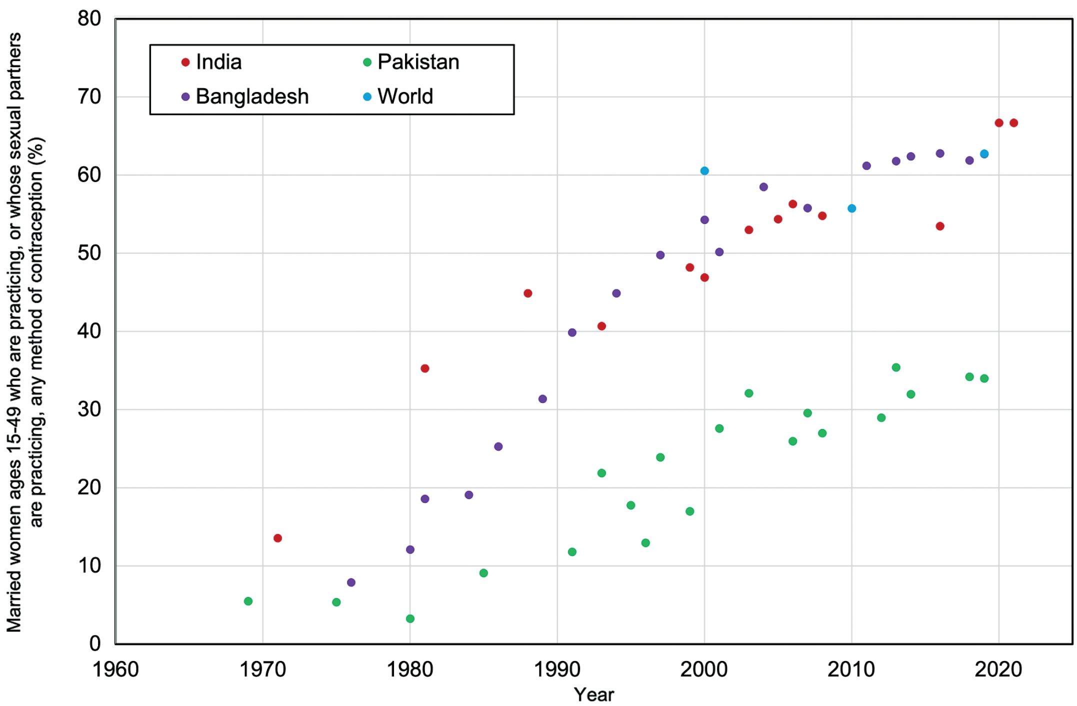

Fig. 1 Prevalence of Contraception, India, Pakistan, Bangladesh, 1960–2023.
Source: World Bank database, Prevalence of Contraception, India, Pakistan, Bangladesh, 1960–2023.

eighth Five-Year Plan, the focus shifted from expanding contraception to developing a community-based strategy. The decline was only temporary, and since the mid-1990s, contraceptive prevalence has been on the rise again. Currently, contraceptive prevalence in India is around 67%, and India's main goal is to continue increasing prevalence and, above all, to increase the diversification of modern contraceptive methods (United Nations Population Fund n.d.). Bangladesh has experienced a similar development to India. The highest increase occurred between 1970 and 1990. Bangladesh included access to contraception and the quality of its delivery in its first Five-Year Plan (Alam 2017). During this period, there was an increase from 8% in 1975 to 40% in 1991. Unlike India, however, Bangladesh also had a relatively successful period in the 1990s, when prevalence continued to rise. During this period, Bangladesh was implementing its sixth Five-Year Plan, which included the approval of a "reproductive health service package" that also focused on strengthening contraceptive use (Jahan 2007). A minor crisis did not occur until the beginning of the millennium, when there were minor fluctuations. Since 2010, the prevalence of contraception in Bangladesh has been stagnating at around 60%. In Pakistan, the increase was rather gradual at first and very inconsistent in later years, with high year-on-year fluctuations. The Pakistani government was very inconsistent in its first five-year plans, and there were significant disputes between the government and the Ministry of Health and Population Welfare. As a result, the achievement of contraceptive prevalence targets was neglected until the fourth five-year plan (Hakim and Mahmood 2001). Pakistani regulations were rather unstable, and contraceptive prevalence rose and fell regularly until the present day, which again points to strong government instability in Pakistan. A more significant increase began in 1990 after the introduction of the community system and the Lady Health Workers program in 1994. In 2019, the prevalence of contraception in Pakistan was 35%, while in Bangladesh it was around 61% and in India as high as 67%. Compared to the global prevalence of contraception, it can be seen that India and Bangladesh are approaching the global average, while Pakistan lags significantly behind (Fig 1).

4.6 Financing of Family planning programs

The success of Family planning programs is significantly influenced by the amount of money invested in their operation. All countries benefit from their own funding as well as support from foreign donors, NGO's or other countries (and later from the private sector) (Cole et al. 2019).

Funding for family planning programmes is in all the countries studied dependent on foreign aid and assistance from NGO's. However, in each of the countries studied, the pattern of funding and the extent of involvement of external funding sources varies (Tab. 2). In India, when the National Family Planning Program (was initiated 1952), most of the funding came from its own budget, with contributions from the UN and bilateral donors, of which the was the largest contributor USA. From the 1960's onwards, USAID joined in significantly and began funding most of the expenditure on technical assistance, equipment and education. However, the Government of India continued to contribute from its budget as well. In the period from 1975–1980, when forced sterilizations were taking place, many donors also withdrew, and the Government of India had to increase its share of investment again and thus more investment came from the government budget. Since the 1980's and the abolition of forced sterilizations, many donors have returned and there has been an increase in grants from USAID, UNFPA and the World Bank. However, India has steadily increased the amount of investment from its own budget, especially after the introduction of the National Rural Health Mission (2005) and the National Health Mission (2013). Today, it mostly funds Family planning programs from its own budget (United Nations Population Fund n.d.)

While India has invested significantly from its own budget, Pakistan has been the opposite. Due to significant opposition from conservatives, Family planning programs were not a priority (Guttmacher Institute 2019). Thus, until the 1990's, 80% was funded by foreign donors (USAID, UNFPA, World Bank), with Pakistan even announcing that it would establish a Population Welfare Program only if it was funded by foreign agencies. Its own participation in funding programs began to increase only after the formation of Lady Health Workers in 1993 (Oxford policy

Tab. 2 Comparison of Family planning programs funding, India, Pakistan, Bangladesh, 1950 – present day.

	1950–1980	1980–2000	2000 – present
India	Mainly donors (USAID, UN)	Combination of donors and state budget	Mainly state budget
Pakistan	Exclusively donors	Exclusively donors	Combination of donors and state budget (decentralization still key donors)
Bangladesh	Exclusively donors	Mainly donors (USAID, UN)	Mostly from state budget, donors as a complement

Source: own processing.

management 2019). Pakistan began to participate more financially and by 2010 donors were funding 60% of the programs. Currently, devolution is underway in Pakistan, an effort to shift responsibility for FPP to the provinces, which will change the overall funding system. However, foreign donors remain a necessary component of funding (UNFPA n.d.).

Bangladesh is somewhere between in terms of funding India and Pakistan. After independence in 1971, Family planning programs were entirely dependent on foreign donors. The government had invested in the infrastructure and overall functioning of the new state and there were no resources left for Family planning programs. Thus, as in Pakistan, more than 80% of the funding came from foreign donors (Cleland and Mauldin 1991). Unlike Pakistan, however, Bangladesh began to invest significantly in Family planning programs in the 1990's, particularly in the operation of services, training, and salaries of outreach workers. The share of funding from foreign donors has declined to about 50%. Today, Bangladesh is considered an ideal example of "transition financing", which consists of gradually reducing dependence on foreign donors. Today, Bangladesh finances about 70% of its costs from its own budget (UNFPA n.d.). UNFPA, USAID, NGOs and other donors then provide funding to specific projects rather than to the health system in general (Appleford and Ramarao 2019; UNFPA 2024; Cleland et al. 2006; United Nations 2011; United Nations 2016; United Nations 2005).

Fig. 2 illustrates the relationship between the share of state budget funding for FPP and contraceptive prevalence. The earlier and more extensively countries invested their own resources, the stronger their commitment to promoting contraceptive use, which is reflected in higher prevalence rates. India has financed FPP domestically since the 1960s and now covers most costs (UNFPA n.d), with steadily rising contraceptive prevalence. Bangladesh, initially dependent on foreign aid, began increasing contributions in the 1980s after the Matlab Project and the second five-year plan (Phillips et al. 1996), which was paralleled by a marked rise in contraceptive use. Since the 2000s, Bangladesh has further expanded domestic funding to achieve self-sufficiency. Pakistan, by contrast, invested little until 1994, when the Lady Health Workers program was launched by the Pakistani government; its funding increase was accompanied by growth in contraceptive prevalence (Hafeez et al. 2011). Overall, India leads in domestic funding and prevalence, Bangladesh has substantially caught up, while Pakistan still depends on foreign donors and shows lower prevalence levels.

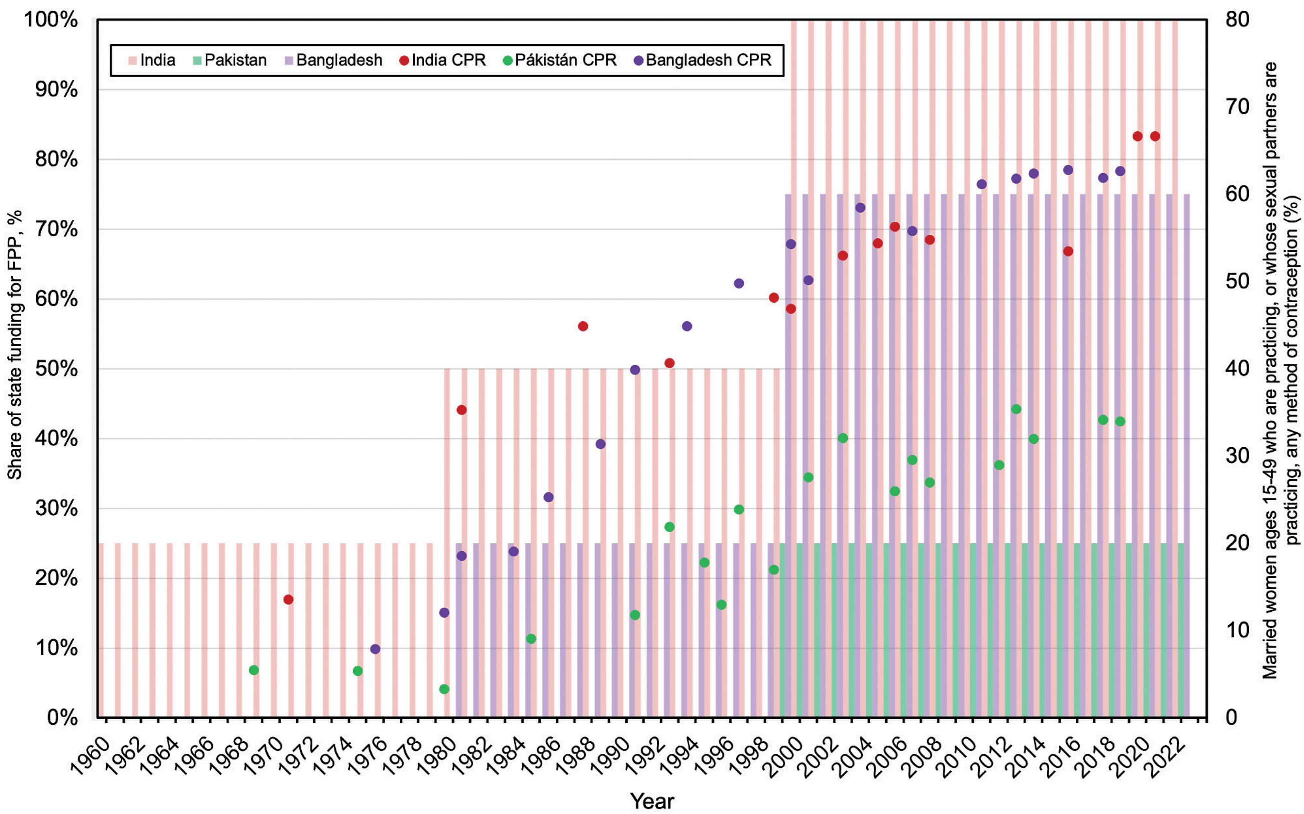

Fig. 2 Share of state funding for FPP, Prevalence of Contraception, India, Pakistan, Bangladesh, 1960–2023.
Source: World Bank database, Prevalence of Contraception, India, Pakistan, Bangladesh, 1960–2023, own processing.
Note: Share of state funding: Exclusively donors (0%), Mainly donors (25%), Combination of donors and state budget (50%), Mostly from state budget (75%), Mainly state budget (100%), Data based on Tab. 2.

5. Discussion

Comparison of Family Planning Programs in India, Bangladesh and Pakistan showed that the main objectives of FPP are the same in all three countries: slowing population growth by expanding contraceptive coverage, educating both sexes, supporting women's education, and implementing community-based systems. The main challenges are regional disparities and the influence of religion. India, Pakistan and Bangladesh all integrated FPP into health systems, relied on international aid, and used community-based approaches linking social and health care (Hay and Garcia 2007; Jha et al. 2023; Bonnifield et al. 2020).

India was the fastest in initial reforms and set the direction for Pakistan and Bangladesh. It also assumed financial responsibility early, covering most costs from its own budget (UNFPA n.d). However, the compulsory sterilization campaign of the 1970s, although it temporarily raised contraceptive prevalence, proved highly unethical and undermined public trust, leading to inconsistent trends in the 1990s after sterilizations became voluntary (Green 2018). Bangladesh, starting later due to independence, implemented FPP in the second five-year plan and, with large donor funding in the 1980s–1990s and a community-based system, achieved major increases in contraceptive prevalence (Phillips et al. 1996; Gray et al. 1997). Fluctuations in 2000–2010 were linked to political changes, health system reforms, and reduced donor subsidies (Jha et al. 2023). Persistent weaknesses are gender inequalities and lack of male involvement, which limit the broader impact of reproductive health education (Bhatia et al. 2024; Cleland and Mauldin 1991; Khan et al. 2020).

Pakistan invested little until 1994, when the Lady Health Workers program was introduced, largely financed by the government, and later reinforced by strong donor support (Hafeez et al. 2011; Phillips and Hossain 2003). Religious and cultural resistance to contraception (Azmat et al. 2015) explains its persistently low prevalence. Although community-based programs and foreign aid have helped, Pakistan still depends on external donors for more than half of FPP financing and lags behind both India and Bangladesh.

Differences are also evident in the promotion of modern contraceptive methods. India started early but narrowed its options in the 1970's and focused on sterilization (Green 2018). Pakistan faced religious resistance and did not make significant progress until the 1990's (Sathar et al. 2013). Bangladesh prioritized modern contraception from the outset and promoted it through community-based systems. As a result, it now has the highest rate of modern contraceptive use, followed by India, while Pakistan lags behind due to inconsistent policy priorities (Guttmacher Institute 2019; Azmat et al. 2015).

Overall, India and Bangladesh have now reached contraceptive prevalence close to the world average (Kantorová and Bongaarts 2024), while Pakistan remains far lower. The main differences reflect not only financing and policy commitment but also cultural and religious contexts, and societal barriers. In Pakistan, the high prevalence of consanguineous marriages, low female literacy, and early marriages have systematically reinforced the demand for larger families and limited the use of available methods (Zaidi et al. 2019; Sathar and Kazi 2000; Sathar and Royan 2013). In contrast, Bangladesh, despite political instability, has achieved a significant decline in fertility through a community-based door-to-door model (Matlab, FWA, satellite clinics), continuous contact between households and providers, and partnerships between the state and non-governmental organizations (e.g., BRAC, Grameen), which combined family planning with microcredit and women's education (Cleland et al. 1994; Phillips et al. 1988; Phillips et al. 1996; Mercer et al. 2004; World Health Organization 2004; Jahan 2007; Alam 2017). Gradual improvements in girls' school attendance, later marriages, and increasing female labor force participation further weakened traditional preferences for large families and correlated with long-term growth in the use of modern methods and a decline in TFR (Bairagi 2001; United Nations Population Fund 2023).

India presents a mixed picture: while Kerala achieved a rapid decline in fertility through education and health services, poorer states faced persistent difficulties, with programs long constrained by paternalistic and technocratic approaches that undermined women's influence (Jeffery and Jeffery 1997; Ram and Ram 2021; Robinson and Ross 2007). In Pakistan, on the other hand, stagnation in women's education and economic participation has been compounded by frequent program interruptions and weak political support, with the result that even the Lady Health Workers program has failed to achieve sufficient scale (Sathar and Zaidi 2010; Hakim and Mahmood 2001; Sathar 2007; Hafeez et al. 2011; Wazir et al. 2013; Goujon et al. 2020). Overall, these cases show that improvements in human development, education, and women's empowerment were more decisive factors for success than religious differences alone (Sathar and Royan 2013).

6. Conclusion

Despite sharing common goals and principles, family planning programs in India, Pakistan, and Bangladesh have developed along different trajectories shaped by cultural, social, economic, and political contexts. While all three countries integrated FPPs into their health systems and adopted community-based approaches with international support, India and Bangladesh have been more successful in increasing contraceptive use, particularly through early investments and policy innovations. India led in early program implementation but later faced setbacks due to coercive practices, whereas Bangladesh benefited from donor funding and structured community outreach, despite persistent gender inequalities. Pakistan has consistently lagged, primarily due to religious resistance, weak policy commitment, and continued reliance on foreign aid. The effectiveness of FPPs thus depends not only on program design but also on how well they are adapted to local social structures and values. Future strategies must prioritize context-specific approaches to ensure more inclusive and sustainable progress in reproductive health.

References

Amin, R., Choudhuri, S. R., Mariam, A. G., McCarthy, J. (1987): Family planning in Bangladesh, 1969–1983. International Family Planning Perspectives 13(1), 16–20, https://doi.org/10.2307/2948103.

Alam, N., Ali, T., Razzaque, A., Rahman, M., Haq, M. Z., Saha, S. K., Ahmed, A., Sarder, A. M., Haider, M. M., Yunus, M., Nahar, Q., Streatfield, P. K. (2017): Health and Demographic Surveillance System (HDSS) in Matlab, Bangladesh. International Journal of Epidemiology 46(3), 809–816, https://doi.org/10.1093/ije/dyx076.

Ataullahjan, A. (2018): Family planning in Pakistan: Unraveling the complexities. Master's thesis, University of Alberta. University of Alberta Education and Research Archive. Available online: https://era.library.ualberta.ca/items/6d8e9c4e-2c4b-4a5e-9b8f-1a3e5e5e5e5e (accessed on 18 February 2025).

Appleford, G., Ramarao, S. (2019): Health financing and family planning in the context of Universal Health Care: Connecting the discourse. Brief, Population Council, New York, https://doi.org/10.31899/rh6.1021.

Azmat, S. K., Ali, M., Ishaque, M., Mustafa, G., Hameed, W., Abbas, G., Ahmed, A. (2015): Assessing predictors of contraceptive use and demand for family planning services in underserved areas of Punjab province in Pakistan: Results of a cross-sectional baseline survey. Reproductive Health 12: 25, https://doi.org/10.1186/s12978-015-0016-9.

Bairagi, R. (2001): Abortion: An emerging issue to meet son preference and fertility control in Bangladesh. XXIV IUSSP General Population Conference, Salvador, Brazil, 18–24 August 2001, International Union for the Scientific Study of Population. Available online: https://iussp.org/sites/default/files/Brazil2001/s20/S21_02_Bairagi.pdf (accessed on 20 March 2025).

Bates, L. M., Islam, K., Al-Kabir, A., Schuler, S. R. (2003): From Home to Clinic and from Family Planning to Family Health: Client and Community Responses to Health Sector Reforms in Bangladesh. International Family Planning Perspectives 29(2), 88–94, https://doi.org/10.2307/3181063.

Bean, L. L., Bhatti, A. D. (1969): Three Years of Pakistan's New National Family-Planning Programme. The Pakistan Development Review 9(1), 35–57, https://doi.org/10.30541/v9i1pp.35-57.

Bhatia, B., Hossain, S., Ghosh, U., Salignac, F. (2024): Reimagining gendered community interventions: The case of family planning programs in rural Bangladesh. Global Health Research and Policy 9: 3, https://doi.org/10.1186/s41256-023-00337-8.

Bongaarts, J. (2020): Trends in fertility and fertility preferences in sub-Saharan Africa: The roles of education and family planning programs. Genus 76: 32, https://doi.org/10.1186/s41118-020-00098-z.

Bongaarts, J., Sinding, S. (2011): Population Policy in Transition in the Developing World. Science 333(6042), 574–576, https://doi.org/10.1126/science.1207558.

Bongaarts, J. (2025): Fertility Transitions in Low- and Middle-Income Countries: The Role of Preferences. Population and Development Review 51(1), 163–180, https://doi.org/10.1111/padr.12675.

Bonnifield, R., Keller, J. M., Glassman, A. (2020): Family Planning and The Global Financing Facility: Current Evidence and a Learning Agenda. Center for Global Development. Available online: https://www.cgdev.org/publication/family-planning-and-global-financing-facility-current-evidence-and-learning-agenda (accessed on 26 March 2025).

Casterline, J., Sinding, S. (2000): Unmet Need for Family Planning in Developing Countries and Implications for Population Policy. Population and Development Review 26(4), 691–723, https://doi.org/10.1111/j.1728-4457.2000.00691.x.

Cleland, J., Mauldin, W. P. (1991): The Promotion of Family Planning by Financial Payments: The Case of Bangladesh. Studies in Family Planning 22(1), 1–18, https://doi.org/10.2307/1966515.

Cleland, J., Philips, J., Amin, S., Kamal, G. (1994): The determinants of reproductive change in Bangladesh: success in a challenging environment (English). World Bank regional and sectoral studies Washington, D. C.: The World Bank. Available online: http://documents.worldbank.org/curated/en/991321468768584526 (accessed on 15 February 2025).

Cleland, J., Bernstein, S., Ezeh, A., Faundes, A., Glasier, A., Innis, J. (2006): Family planning: the unfinished agenda. The Lancet 368: 9549, 1810–1827, https://doi.org/10.1016/S0140-6736(06)69480-4.

Connelly, M. (2006): Population Control in India: Prologue to the Emergency Period. Population and Development Review 32(4), 629–667, https://doi.org/10.1111/j.1728-4457.2006.00145.x.

Cole, M. S., Boydell, V., Hardee, K., Bellows, B. (2019): The Extent to Which Performance-Based Financing Programs' Operations Manuals Reflect Rights-Based Principles: Implications for Family Planning Services. Global Health: Science and Practice 7(2), 329–339, https://doi.org/10.9745/GHSP-D-19-00007.

Donaldson, P. J., Tsui, A. O. (2018): Family Planning Programs. Oxford Research Encyclopedia of Global Public Health. Oxford: Oxford University Press.

Fárek, J. (2012): India's economy: Difficulties and remarkable rise. Institute of International Relations Prague. Available online: https://www.iir.cz/en/ekonomika-indie-obtize-i-pozoruhodny-vzestup (accessed on 26 April 2025).

Finlay, J. E. (2024): A Narrative Review of the Impact of Public Family Planning Policies and FP2030 (2023): Systematic Reviews of Financing Mechanisms for Family Planning: Evidence and Knowledge Gaps. FP2030, Washington, D. C. Available online: https://www.fp2030

.org/app/uploads/2023/07/MEETING-REPORT.-SYSTEMATIC-REVIEWS-OF-FINANCING-MECHANISMS-FOR-FAMILY-PLANNING.-EVIDENCE-AND-KNOWLEDGE-GAPS.pdf (accessed on 26 April 2025).
Programs on the Contraceptive Transition in Low- and Middle-Income Countries. Population and Development Review 50(S2), 649–674, https://doi.org/10.1111/padr.12693.
Goujon, A., Wazir, A., Gailey, N. (2020): Pakistan: A population giant falling behind in its demographic transition. Population & Societies 576(4), 1–4, https://doi.org/10.3917/popsoc.576.0001.
Gray, A., Chowdhury, J. H., Caldwell, B., Al-Sabir, A. (1997): Traditional family planning in Bangladesh. Dhaka: Population Council, https://doi.org/10.31899/rh1997.1000.
Green, H. H. (2018): The legacy of India's quest to sterilize millions of men. Pulitzer Center. Available online: https://pulitzercenter.org/stories/legacy-indias-quest-sterilize-millions-men (accessed on 26 May 2025).
Guttmacher Institute (2019): Adding it up: Investing in contraception and maternal and newborn health in Pakistan. Available online: https://www.guttmacher.org/report/adding-it-up-meeting-contraceptive-mnh-needs-pakistan (accessed on 26 May 2025).
Günther, I., Harttgen, K. (2016): Desired Fertility and Number of Children Born Across Time and Space. Demography 53(1), 55–83, https://doi.org/10.1007/s13524-015-0451-9.
Hafeez, A., Mohamud, B. K., Shiekh, M. R., Shah, S. A., Jooma, R. (2011): Lady health workers programme in Pakistan: challenges, achievements and the way forward. Journal of the Pakistan Medical Association 61(3), 210–215, PMID: 21465929.
Hakim, A., Mahmood, M. A. (2001): Population Policy Shifts and Their Implications for Population Stabilisation in Pakistan. The Pakistan Development Review 40(4), 551–573. Available online: https://www.jstor.org/stable/41260352 (accessed on 16 May 2025).
Hardee, K., Kumar, J., Newman, K., Bakamjian, L., Harris, S., Rodríguez, M., Brown, W. (2014): Voluntary, human rights-based family planning: a conceptual framework. Studies in Family Planning 45(1), 1–18, https://doi.org/10.1111/j.1728-4465.2014.00373.x, PMID: 24615572.
Hay, P., Garcia, M. (2007): Poor Countries Losing Out on Family Planning Benefits – New Bank Report. Available online: https://documents1.worldbank.org/curated/en/099231506032220520/pdf/IDU0acf256660866f046d70a46c0454505d5ed98.pdf (accessed on 26 April 2025).
Chaurasia, A. R. (2014): Contraceptive Use in India: A Data Mining Approach. International Journal of Population Research 2014: 821436, https://doi.org/10.1155/2014/821436.
Chaurasia, A. R. (2020): Contraceptive Method Skew in India 1992–2016. Preprint, medRxiv 2020.07.14.20154013, https://doi.org/10.1101/2020.07.14.20154013.
Jahan, R. (2007): Securing Maternal Health Through Comprehensive Reproductive Health Services: Lessons from Bangladesh. American Journal of Public Health 97(7), 1186–1190, https://doi.org/10.2105/AJPH.2005.081737.
Jeffery, P., Jeffery, R. (1997): Population, gender and politics: demographic change in rural North India. Contemporary South Asia 3. Cambridge University Press, http://catdir.loc.gov/catdir/toc/cam022/96043848.html.
Jha, A., Wheeler, R. J., Comas, S., Gribble, J., Urgaz, J., Gonzalez-Pier, E. (2023): Getting ready to reduce donor dependency: The co-financing of family planning commodities. Health Policy and Planning 39(1), 87–93, https://doi.org/10.1093/heapol/czad106.
Kantorová, V., Bongaarts, J. (2024): Contraceptive Change and Fertility Transition. Population and Development Review 50(S2), 459–485, https://doi.org/10.1111/padr.12689.
Khan, M. N., Harris, M., Loxton, D. (2020): Modern contraceptive use following an unplanned birth in Bangladesh: An analysis of national survey data. International Perspectives on Sexual and Reproductive Health 46, 77–87, https://www.jstor.org/stable/10.1363/46e8820.
Kongawad, P., Boodeppa, G. K. (2014): National family planning programme – During the five-year plans of India. Journal of Evolution of Medical and Dental Sciences 3(19), 5172–5178, https://doi.org/10.14260/jemds/2014/2569.
Ledbetter, R. (1984): Thirty Years of Family Planning in India. Asian Survey 24(7), 736–758, https://doi.org/10.2307/2644186.
Memon, Z. A., Tahmeena, Fazal, S. A., Reale, S., Spencer, R., Bhutta, Z., Soltani, H. (2024): Effective strategies for increasing the uptake of modern methods of family planning in South Asia: a systematic review and meta-analysis. BMC Women's Health 24: 13, https://doi.org/10.1186/s12905-023-02859-2.
Mercer, A., Khan, M. H., Daulatuzzaman, M., Reid, J. (2004): Effectiveness of an NGO primary health care programme in rural Bangladesh: evidence from the management information system. Health Policy and Planning 19(4), 187–198, https://doi.org/10.1093/heapol/czh024.
Ministry of Health and Family Welfare: Family Planning. National Health Mission, Government of India. Available online: https://nhm.gov.in/index1.php?lang=1&level=2&lid=222&sublinkid=821 (accessed on 20 May 2025).
Ministry of National Health Services, Regulations & Coordination (2021): FP 2030 National Commitments. Available online: https://pakistan.unfpa.org/en/publications/fp-2030-national-commitment (accessed on 19 March 2025).
Nishtar, S. (2010): Choked Pipes: Reforming Pakistan's Mixed Health System. Karachi: Oxford University Press.
Oxford Policy Management (2019): Lady Health Worker Programme, Pakistan. Performance Evaluation. Available online: https://www.unicef.org/pakistan/reports/performance-evaluation-report-lady-health-workers-programme-pakistan (accessed on 14 April 2025).
Perera, S., Ramani, S., Joarder, T., Shukla, R. S., Zaidi, S., Wellappuli, N., Ahmed, S. M., Neupane, D., Prinja, S., Amatya, A., Rao, K. D. (2024): Reorienting health systems towards Primary Health Care in South Asia. The Lancet Regional Health-Southeast Asia 28: 100466, https://doi.org/10.1016/j.lansea.2024.100466.
Phillips, J. F., Hossain, M. B., Arends-Kuenning, M. (1996): The long-term demographic role of community-based family planning in rural Bangladesh. Studies in Family

Planning 27(4), 204–219, https://www.jstor.org/stable/2137954.
Phillips, J. F., Stinson, W. S., Bhatia, S., Rahman, M., Chakraborty, J. (1982): The Demographic Impact of the Family Planning – Health Services Project in Matlab, Bangladesh. Studies in Family Planning 13(5), 131–140, https://www.jstor.org/stable/1966283.
Phillips, J. F., Stinson, W. S., Bhatia, S., Rahman, M., Chakraborty, J. (1988): Determinants of Reproductive Change in a Traditional Society: Evidence from Matlab, Bangladesh. Studies in Family Planning 19(6), 313–334, https://doi.org/10.2307/1966627.
Phillips, J. F., Hossain, M. B. (2003): The Impact of Household Delivery of Family Planning Services on Women's Status in Bangladesh. International Family Planning Perspectives 29(3), 138–145. Available online: https://www.guttmacher.org/sites/default/files/article_files/2913803.pdf (accessed on 20 May 2025).
Population Services International Pakistan (2021): The use of media in family planning programming. Available online: https://psipakistan.org/wp-content/uploads/2021/09/Media-Engagement-Literature-Review-14-121.pdf (accessed on 5 May 2025).
Pradhan, M. R., Dwivedi, L. K. (2019): Changes in contraceptive use and method mix in India: 1992–92 to 2015–16. Sexual & Reproductive Healthcare 19, 56–63, https://doi.org/10.1016/j.srhc.2018.12.006.
Pritchett, L. (1994): Desired Fertility and the Impact of Population Policies. Population and Development Review 20(1), 1–55, https://www.jstor.org/stable/2137629.
Ram, U., Ram, F. (2021): Demographic Transition in India: Insights into Population Growth, Composition, and Its Major Drivers. Oxford Research Encyclopedia of Global Public Health, https://doi.org/10.1093/acrefore/9780190632366.013.223.
Robinson, W. C., Ross, J. A. (2007): The global family planning revolution: Three decades of population policies and programs. The World Bank. Available online: https://documents1.worldbank.org/curated/en/259541468177850467/pdf/405350Familiy0101OFFICIAL0USE0ONLY1.pdf (accessed on 15 April 2025).
Robinson, W. C. (2023): Common Beginnings but Different Outcomes: The Family Planning Programmes of Pakistan and Bangladesh. In Sathar, Z. A., Phillips, J. F. (Eds.), Transition in South Asia. Oxford University Press, https://doi.org/10.1093/oso/9780199241859.003.0017.
Sathar, Z. A. (2007): Stagnation in fertility levels in Pakistan. Asia-Pacific Population Journal 22(2), 113–131, https://doi.org/10.18356/8302f8fc-en.
Sathar, Z. A., Zaidi, S. (2010): Fertility prospects in Pakistan. Expert Paper No. 2011/07, United Nations, Department of Economic and Social Affairs, Population Division, New York. Available online: https://www.un.org/development/desa/pd/content/fertility-prospects-pakistan (accessed on 15 February 2025).
Sathar, Z. A., Kazi, S. (2000): Women's autonomy in the context of rural Pakistan. The Pakistan Development Review 39: 2, 89–110. Available online: https://file.pide.org.pk/pdf/PDR/200/Volume2/89-110.pdf (accessed on 15 May 2025).
Sathar, Z., Royan, R. (2013): Capturing the Demographic Dividend in Pakistan. Population Council, Islamabad: Population Council, https://doi.org/10.31899/pgy5.1012.
Sathar, Z. A., Zaidi, B., Riaz, A. (2013): Maternal Health and Family Planning in Pakistan: A Historical Overview. In Women's Health in Pakistan: Progress and Challenges. Population Council, Islamabad, 15–38.
Seltzer, J. R. (2002): The Origins and Evolution of Family Planning Programs in Developing Countries. RAND Corporation, https://www.jstor.org/stable/10.7249/mr1276wfhf-dlpf-rf.
Shabuz, Z. R., Haque, M. E., Islam, M. K., Bari, W. (2022): Link between unmet need and economic status in Bangladesh: Gap in urban and rural areas. BMC Women's Health 22: 176, https://doi.org/10.1186/s12905-022-01752-8.
Sinding, S. W. (2000): The great population debates: How relevant are they for the 21st century? American Journal of Public Health 90(12), 1841–1845, https://doi.org/10.2105/AJPH.90.12.1841.
Snyder, H. (2019): Literature review as a research methodology: An overview and guidelines. Journal of Business Research 104, 333–339, https://doi.org/10.1016/j.jbusres.2019.07.039.
Sultan, M., Cleland, J. G., Ali, M. M. (2002): Assessment of a new approach to family planning services in rural Pakistan. American Journal of Public Health 92(7), 1168–1172, https://doi.org/10.2105/ajph.92.7.116.
United Nations (2005): Executive Board of the United Nations Development Programme, the United Nations Population Fund. Available online: https://docs.un.org/en/dp/2005/29 (accessed on 15 May 2025).
United Nations Development Programme (2011): Executive Board – Annual Session 2011: Advance documentation. New York: United Nations Development Programme. Available online: https://www.undp.org/executive-board-annual-session-2011 (accessed on 15 May 2025).
United Nations (2016): Executive Board of the United Nations Development Programme, the United Nations Population Fund and the United Nations Office for Project Services: Report of the Executive Board on its work during 2016 – Official Records, 2016, Supplement No. 15 (E/2016/35). New York: United Nations. Available online: https://www.undp.org/sites/g/files/zskgke326/files/2021-07/E201635e.pdf (accessed on 15 May 2025).
United Nations Development Programme (2024): UNFPA Bangladesh Annual Report 2023. New York: United Nations Development Programme. Available online: https://reliefweb.int/report/bangladesh/unfpa-bangladesh-annual-report-2023 (accessed on 15 May 2025).
United Nations, Department of Economic and Social Affairs, Population Division (2024): World Population Prospects 2024: Summary of Results. Available online: https://population.un.org/wpp/assets/Files/WPP2024_Summary-of-Results.pdf (accessed on 15 May 2025).
United Nations (1994): Programme of Action of the International Conference on Population and Development. Cairo: United Nations Population Fund (UNFPA), https://www.unfpa.org/publications/international-conference-population-and-development-programme-action (accessed on 15 May 2025).
United Nations (2022): World Family Planning 2022: Meeting the Changing Needs for Family Planning

– Contraceptive Use by Age and Method. Available online: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2023/Feb/undesa_pd_2022_world-family-planning.pdf (accessed on 15 May 2025).
United Nations (2000): United Nations Millennium Declaration. New York: United Nations General Assembly. Available online: https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_55_2.pdf (accessed on 15 May 2025).
United Nations Population Fund (UNFPA) (2023): State of World Population 2023: 8 Billion Lives, Infinite Possibilities. New York: UNFPA. Available online: https://www.unfpa.org/publications/state-world-population-2023-8-billion-lives-infinite-possibilities (accessed on 15 May 2025).
United Nations Population Fund (n.d.): UNFPA India – Transparency Portal. UNFPA Data Portal. Available online: https://www.unfpa.org/data/transparency-portal/unfpa-india (accessed on 18 March 2025).
Visaria, L., Ved, R. R. (2016): India's Family Planning Programme: Policies, Practices and Challenges. London: Routledge, https://doi.org/10.4324/9781315652498.
Wazir, M. S., Shaikh, B. T., Ahmed, A. (2013): National program for family planning and primary health care Pakistan: A SWOT analysis. Reproductive Health 10(1), 60, https://doi.org/10.1186/1742-4755-10-60.
World Bank (1993): Effective Family Planning Programs. Washington, DC: The World Bank.
World Bank (2018): World's population will continue to grow and will reach nearly 10 billion by 2050. Available online: https://datatopics.worldbank.org/world-development-indicators/stories/world-population-will-continue-to-grow.html (accessed on 15 February 2025).
World Bank (2023): Contraceptive prevalence, any modern method (% of married women ages 15–49) [dataset]. World Development Indicators. Available online: https://data.worldbank.org/indicator/SP.DYN.CONM.ZS (accessed on 15 February 2025).
World Health Organization (2004): Family Planning in South-East Asia: Factsheets. Available online: https://iris.who.int/bitstream/handle/10665/279771/fp-fs-eng.pdf (accessed on 15 February 2025).
World Health Organization (2015): Bangladesh Health System Review. Health Systems in Transition 5(3). Geneva: WHO.
World Health Organization (2018): Bangladesh and Family Planning: An Overview. Available online: https://apps.who.int/iris/bitstream/handle/10665/279771/fp-fs-eng.pdf (accessed on 15 February 2025).
World Health Organization (n.d.): Health Service Delivery: Lady Health Workers. Cairo: WHO Regional Office for the Eastern Mediterranean. Available online: https://www.emro.who.int/pak/programmes/service-delivery-lady-health-workers.html (accessed on 15 February 2025).
Zaidi, S., Bigdeli, M., Langlois, E. V., Riaz, A. (2019): Health systems changes after decentralisation: progress, pitfalls and prospects in Pakistan. BMJ Global Health 4(1): e001013, https://doi.org/10.1136/bmjgh-2018-001013.

Comparison of the success rates of candidates for geography studies in solving different types of tasks in 2016 and 2024

Tomáš Matějček[1,2,*], Jakub Jelen[1]

[1] Charles University, Faculty of Science, Department of Social Geography and Regional Development, Czechia
[2] Jan Evangelista Purkyně University in Ústí nad Labem, Faculty of Science, Department of Geography, Czechia
* Corresponding author: tomas.matejcek@natur.cuni.cz

ABSTRACT

The aim of this study is to identify changes in geographic literacy over time by comparing the success rates of geography applicants when solving identical tasks used in the geography entrance exam test in 2016 and 2024 at the Faculty of Science of Charles University.

We summarize the results of a comparison of the success rates in selected test tasks and we look at the shift in geographic literacy between two groups of respondents with similar characteristics (age, interest in studying geography at the same university) in different time periods. Through our study, we aim to open up the possibility of using geography entrance exam tests as one of the possible sources for studying the evolution of geographic literacy over time. Longitudinal studies focusing on changes in geographic literacy are still very rare, which we consider to be a research gap. The tasks used in the entrance tests in 2016 and 2024 were compared in order to determine which types of test tasks experienced the greatest change in success rate. The comparison included answers to a total of 25 test tasks, which were intentionally set identically in both years to make such a comparison possible. Answers were available from 269 respondents in 2016 and from 132 respondents in 2024. When evaluating the results, the tasks were divided according to various criteria (thematic focus, category of educational objectives, use of mathematical skills, inclusion of a visual element, etc.). The results indicate a relatively high rate of change in success in solving certain types and groups of test tasks. The results also show changes in the level of geographical literacy of students who come to university from secondary schools. This information could be helpful not only for universities themselves (who will get better information about changes in the level of applicants from secondary school level), but also for secondary school educators and experts engaged in curriculum development (who will get feedback on secondary education results). The results underline the importance of systematically monitoring changes in geographical literacy and call for further research on a larger dataset and across more time points.

KEYWORDS

item analysis; entrance exam; geographical education; test task; geographic literacy

Received: 5 May 2025
Accepted: 12 November 2025
Published online: 26 November 2025

Matějček, T., Jelen, J. (2025): Comparison of the success rates of candidates for geography studies in solving different types of tasks in 2016 and 2024. AUC Geographica 60(2), 285–295
https://doi.org/10.14712/23361980.2025.23

1. Introduction

The goal of geographical education can be considered, in the most general sense, to be the development of geographic literacy, which we understand as the ability of a person to understand the world from a geographic perspective, to orient themselves in spatial contexts and to understand the relationships between people, places and the environment (for more details, e.g. Binimelis Sebastián et al. 2024; Dikmenli 2014; Soleh et al. 2022; Řezníčková 2003). Geographic literacy thus represents a complex set of knowledge, skills, values, attitudes and competences that enable individuals to understand spatial phenomena, processes and relationships at the local, regional and global levels. It includes the ability to identify and analyze spatial patterns and contexts, interpret geographic information (including maps, data and visualizations) and critically approach their sources and meanings. Geographic literacy therefore includes spatial orientation (knowledge of basic geographical concepts, the ability to read a map, understand where something is and why), understanding natural and social processes, relationship to place (the ability to perceive one's own place in space, to have an overview of one's locality and global context) and the ability to make decisions (to use geographic information in responsible decision-making). The achieved level of geographic literacy at individual levels of school education is commonly verified and assessed in various forms, such as oral exams, essays or written achievement tests), the advantage of which is a high degree of objectivity in evaluating results. The results obtained then indicate the current achieved level in particular aspects of geographic literacy of the group of pupils or students in question. However, research studies that focus on changes in the achieved level of geographical education, or geographic literacy over time, are very rare. Although there are certain partial studies most often focused on specific areas of geographic literacy, longitudinal studies dealing with the change in geographic literacy over time are particularly lacking. We perceive the lack of studies devoted to the development of geographic literacy over time as a research gap. With our study, we would like to open up the possibility of using geography entrance exam tests as one of the possible sources for studying the development of geographic literacy over time.

The aim of this study is to identify changes in geographic literacy over time by comparing the success rates of geography applicants when solving identical tasks used in the geography entrance exam test in 2016 and 2024 at the Faculty of Science of Charles University.

Through the research conducted, we will try to answer the following research question:

How did the success rate of solving different types of test tasks with different thematic focuses change when comparing candidates from different years?

Our research should be understood as a pilot study, using only a limited amount of data. We see its importance mainly in testing the possibilities of evaluating year-on-year changes in the success rate in solving tasks based on various criteria. Nevertheless, we believe that the results of our analysis may be an interesting contribution to the discussion on the development of geographic literacy, which is not yet systematically monitored. The results of our research point to changes in the level of geographic literacy (or more precisely the part of it that can be assessed through entrance tests) over time, for a specific age group, namely secondary school graduates who aspire to study geography at university.

2. Geographic literacy and its changes over time

Bendl et al. (2024) state in their study that there are changes in the concept of geographical thinking, a shift from an emphasis on factual knowledge to a focus on geographical thinking and geographical competences. Along with the development of geography, there should also be a change in geographic literacy, but this will certainly not be a parallel process.

The study by Binimelis Sebastián et al. (2024) highlights the low level of geographic literacy of Spanish high school students and states that rote learning (mechanical learning of facts) does not lead to permanent geographic literacy. At the same time, it emphasizes the need for new teaching methods – active, based on working with maps, and developing spatial competences – to increase the geographic literacy of the younger generation.

The importance of strengthening geographic literacy is also mentioned by McFarlane (2024), for example, who emphasizes that geographic literacy is a necessity that has the potential to affect the economy, security, sustainability and cultural understanding.

There are other studies that focus on changes in a specific aspect of geographic literacy. In the USA, for example, research was conducted to map the development of the performance of eighth-grade students in geography between 1994 and 2018 (Solem et al. 2021). This research is based on the National Assessment of Educational Progress (NAEP), a federal program in the USA, sometimes referred to as the "Nation's Report Card". This is a standardized test that monitors the knowledge and skills of American students in various areas – including geography. The results show that American children generally have low levels of geographic literacy, which decreased between 2014 and 2018 (National Center for Education Statistics 2020).

In the framework of scientific literacy research, some geography questions (particularly focused on physical geography) are only occasionally included at a more general level and at the category of pupils in the second stage of primary school (i.e. four or more years younger than the respondents we surveyed) in the regular international surveys TIMSS (Trends in International Mathematics and Science study) and PISA (Programme for International Student Assessment). These surveys deal with the analysis and assessment of pupils' knowledge and skills, and take place at regular intervals. The TIMSS places emphasis on the curriculum and teaching content as well as monitoring the development of knowledge over time (Shi et al. 2016). The PISA survey tests three basic areas: reading literacy, mathematical literacy and science literacy and assesses not only what students know, but also how they are able to work with their knowledge, i.e. the emphasis is on skills, understanding and practical application (Wu 2010).

More detailed analyses of TIMSS and PISA results on science literacy are summarized, for example, by Teig et al. (2022), who identified 82 studies that analyzed data from TIMSS or PISA and focused on different aspects of science teaching and learning. Zhang et al. (2023) developed and validated an instrument for assessing students' science literacy. Kjærnsli and Lie (2004) focused on science literacy in the Nordic countries, analyzing PISA items according to whether they required conceptual understanding or intellectual process skills.

3. Admission tests and their evaluation

The admissions process is the gateway to higher education. It is a process during which applicants are assessed according to predetermined criteria and then a decision is made on their admission or non-admission to study. Admissions processes at universities in Czechia are not centrally managed and it is therefore entirely up to each individual university to decide how to organize them. In general, it can be said that the conditions of the admissions process depend primarily on the characteristics of the field of study and the traditions of each university.

If a university uses the option of selecting applicants for study on the basis of the results of an entrance examination, there are several options for how to implement it. Of the various forms (essays, written works, portfolios or oral interviews) tests are the most widespread, usually with automated evaluation using automatic optical scanners (optical mark recognition). Such tests are usually composed of closed questions with one or more correct answers (Pérez-Benedito et al. 2014).

A candidate's performance can be assessed either in absolute terms (reaching a certain threshold required for admission) or on the basis of relative performance (percentile score). In this case, the candidate is compared to other candidates, and the result is expressed relative to other test takers (Ørberg 2018; Frey and Detterman 2004).

Each of the aforementioned admissions options has its advantages and disadvantages, and there are studies devoted to choosing the appropriate tool to determine the qualities of candidates. Researchers also agree that while standardized admissions test scores usually measure cognitive abilities, successful study requires a number of cognitive and non-cognitive attributes that are not verified by admissions tests (Camara and Kimmel 2005; Krumrei-Mancuso et al. 2013; Robbins et al. 2004; Weissberg and Owen 2005).

Silva et al. (2020) states that the best way to select suitable students for university studies is to compare their results in a specific test, which provides objective data, ensuring greater transparency and objectivity of the admission process. On the other hand, he does not consider it appropriate to decide on the admission of students based on their secondary school grades, due to the diversity of schools and the difficulty of comparing academic results of applicants. Other authors oppose the suitability of a combination of different approaches. Bartáková et al. (2018) clearly show that the use of both entrance tests and secondary school grades is justified in university admissions processes for predicting whether a student will be successful in their studies.

Achievement tests are commonly used in entrance exams for the study of geography in Czechia and elsewhere. Tests that can be considered high-quality should meet basic parameters, specifically they should be reasonably difficult and sufficiently sensitive as well as valid, reliable and objective. Objectivity should be ensured by the test being administered to all participants in the same way and by having clear and precisely defined criteria for evaluating the results. The evaluation is therefore independent of the evaluator.

Test reliability indicates how reliable the test is, or rather how stable the results are. An important factor is the consistency of the results in repeated measurements. If a test has high reliability, it will show similar results in different repetitions and under different conditions. Low reliability can be caused by the properties of the test itself, e.g. unclear formulation of the task or inadequate level of tested knowledge related to the tested group (Schindler et al. 2006).

The validity of a test denotes whether the test detects and measures what it is supposed to detect, and that it verifies the tested knowledge at the appropriate level. Validity can be ensured primarily by the relevance of the test's content, i.e. that the test includes all essential parts of the curriculum that the student should know. Pilot testing in a smaller group is also appropriate before implementing the test in practice (Průcha 2009; Štuka and Vejražka 2022; Viktorová and Charvát 2014).

Sensitivity is the ability of the test to correctly identify individuals who have the desired characteristic or feature, e.g. students who have the required skills or knowledge. In the context of achievement tests, sensitivity refers to the ability of the test to distinguish between different levels of performance in students. If the test is highly sensitive, it will be able to distinguish high-performing students from low-performing ones. If students who have better overall knowledge solve a task with great success, while students with worse overall knowledge achieve poor results, this task has high sensitivity (Chráska 1999). A sensitive task, like a sensitive achievement test, is intended to favor students with better knowledge or skills.

There are several ways to determine sensitivity. It is possible to analyze the success rates of individual tasks in a test to determine how well each test task discriminates between high and low-performing students. This is done using the so-called discrimination index, while for a deeper analysis of the sensitivity of a test, the Rasch model, for example, is used (Schindler et al. 2006; Kalhous and Obst 2009; Průcha 2009 and others).

Research on university achievement tests was conducted by, for example, Brožová and Rydval (2014), who analyzed exam tests in mathematics over a period of 13 years. The authors examined the tendencies towards worsening results, the difficulty of the tests and the suitability of the assessment system. The work assessed the quality of the tests using the difficulty index, discriminative ability and reliability.

Entrance exam tests have been analyzed in some previous studies. An analysis of entrance exam tests in mathematics was conducted by Zhang (2023), the properties of tasks in chemistry were analyzed by Šrámek and Teplá (2021); Šrámek and Teplá (2022).

4. Materials and methods

For the analysis of test tasks of entrance exams for geography study programs of the Faculty of Science of Charles University, entrance exams from 2016 and 2024 were selected. In 2016, a total of 269 candidates took the test. In 2024, 132 candidates took it. Both entrance exams contained a total of 40 multiple choice test tasks with four answer options, only one answer being correct. Candidates had 60 minutes to complete the tests. Answers were recorded on a separate recording sheet, which was automatically processed and evaluated using computer software.

For both tests, reliability was determined as part of their evaluation by calculating Cronbach's alpha value, which is an internal consistency coefficient that expresses the extent to which individual items of a test or questionnaire measure the same construct and therefore how reliable the test is overall. For the 2016 test, the Cronbach's alpha value was determined to be 0.775 and for the 2024 test it was 0.757. Both of these values indicate good internal consistency of the test, generally considered acceptable for research purposes (Cortina 1993; Tavakol and Dennick 2011). The content validity of the test was assessed by two independent researchers. The main criteria were the thematic focus of the questions and the relevant choice of possible answers.

When selecting test tasks suitable for repetition in 2024, tasks with non-functional distractors, tasks that were too easy (success rate higher than 90%) and tasks that used data that was already outdated in 2024, were eliminated. In this way, a total of 25 test tasks were selected that were used in the same form in the tests in both 2016 and 2024.

These test tasks were subjected to a deeper analysis, which mainly included the calculation of the sensitivity expressed by the discrimination index (Schindler et al. 2006). If a test task has a high discrimination index value, it was solved mainly by students who were successful in the entire test, i.e. these are tasks that favor successful students. Low discrimination index values can also be achieved by tasks that are formulated in a complicated way or tasks for which there are different solution strategies. More successful students may thus try to apply complex solutions, while less successful students only try to guess the correct answer. This discrimination index is calculated using the ULI (upper-lower index) method.

It was used, for example, by Logayah et al. (2024) when evaluating the results of the Geographic Olympiad in Indonesia. An example of its use in other fields is the research of Lucky et al. (2025). In our analysis, we calculated the ULI for individual tasks as a check that there were no test tasks among the analyzed sample that could be ambiguous or incorrectly formulated.

The formula for calculating the discrimination index according to Schindler et al. (2006) is:

$$D = \frac{n_L - n_H}{0.5 \times N},$$

where n_L is number of students from the better half who solved the given task correctly,

n_H is number of students from the worse half who solved the given task correctly,

N is total number of students who solved the task.

For the purposes of further evaluation, individual test tasks were classified based on thematic focus, on the educational objectives they verify, as well as on the use of visual elements and the use of mathematical skills.

In terms of thematic focus, the test tasks were divided into five categories: planetary geography, physical geography, social geography (including demography), regional geography (including topography) and cartography. This classification is based on the traditional dichotomy between physical and human (social) geography, as presented by, for example, Cloke et al. (2005) or Hampl (2000).

The evaluation of the educational objectives was carried out using a simplified version of the revised

Bloom's taxonomy, hereinafter referred to as RBT (Amer 2006; Anderson and Krathwohl 2001; Byčkovský and Kotásek 2004). This classification represents the most commonly used classification of educational objectives and is widely used both in theoretical studies and in educational practice. The cognitive process dimension was classified into three categories: Remember, Understand and Apply, with the first two categories corresponding to RBT, and the Apply category also including test tasks that fulfill educational objectives classified in RBT as Analyze and Evaluate. We combined the three categories of Apply, Analyze and Evaluate because the boundary between these them is not always sharp and distinguishing educational objectives between these categories is often difficult (e.g. Amer 2006). The Create category was not included at all, because multiple choice test tasks do not allow the use of this category in practice. Our resulting division of cognitive process dimensions largely (in rough outline) corresponds to the categories proposed by Nimierko (1979). We adopted the assessment of the knowledge dimension from the RBT without modification, thus distinguishing the categories of factual, conceptual, procedural and metacognitive knowledge.

We applied the use of visual elements and use of mathematical skills as sorting aspects because these task characteristics can represent a potential obstacle for a certain type of candidate (especially use of mathematical skills) or, conversely, make solving the task easier for some (use of visual elements). A similar division of tasks was used by Bláha et al. (2024).

The classification of the test tasks (division into categories of thematic focus, prevailing goal according to RBT, knowledge dimension according to RBT, use of visual elements, use of mathematical skills) was done independently by two researchers. The intercoder reliability was 99.2% (that is, 124 out of 125 total items). Only for one task did the evaluators suggest a different classification, which was subsequently discussed.

Tab. 1 shows the resulting classification of individual test tasks. Based on this classification, the tasks

Tab. 1 Classification of test tasks based on various criteria and their sorting into groups.

Test task	Thematic focus	Prevailing goal according to RBT	Knowledge dimension according to RBT	Use of visual element	Use of mathematical skills
1	regional geography	remember	factual	no	no
2	cartography	apply	procedural	no	yes
3	regional geography	remember	factual	no	no
4	regional geography	remember	factual	no	no
5	planetary geography	apply	conceptual	no	no
6	physical geography	remember	factual	no	no
7	cartography	apply	conceptual	map	no
8	regional geography	remember	factual	map	no
9	regional geography	apply	factual	map and elevation profile	no
10	cartography	apply	procedural	map	yes
11	cartography	apply	conceptual	map and elevation profile	no
12	cartography	apply	conceptual	map	no
13	regional geography	remember	factual	no	no
14	planetary geography	apply	procedural	no	yes
15	planetary geography	apply	procedural	no	yes
16	physical geography	remember	factual	no	no
17	physical geography	apply	conceptual	climate diagram	no
18	physical geography	understand	factual	no	no
19	physical geography	understand	conceptual	no	no
20	social geography	remember	factual	no	no
21	social geography	understand	conceptual	diagram	no
22	social geography	understand	conceptual	table	no
23	regional geography	remember	factual	no	no
24	cartography	apply	conceptual	no	no
25	physical geography	apply	conceptual	diagram	no

were divided into groups, for which a summary evaluation was subsequently performed.

The table shows that in terms of thematic focus, the test tasks covered all thematic areas as well as all categories of prevailing goals according to RBT distinguished within the analysis (the Apply category was the most represented). The metacognitive category is missing in the knowledge dimension according to the RBT category, which is not too surprising, since the verification of this dimension would be very unusual in an admission test. Visual elements were represented in ten test tasks, and four test tasks required the use of mathematical skills.

Furthermore, the success rate in % for the years 2016 and 2024 and the difference in the success rate of the given test task between these years were calculated for each test task. This difference was expressed in percentage points with a positive (increase in

Tab. 2 Task principle and selected indicators of the analyzed test tasks.

Test task	Task principle	ULI 2016)	Success rate (2016)	Success rate (2024)	Change of the success rate (2016–2024)
1	recognize the relative position of two cities	0.312	56.5%	55.3%	−1.2%
2	identify the distance between two points on a map	0.305	57.3%	61.4%	4.1%
3	classify topographic objects by regions	0.327	46.1%	37.9%	−8.2%
4	identify false information (about Canada)	0.156	87.0%	80.3%	−6.7%
5	identify true information (about Earth's movements)	0.290	57.3%	45.5%	−11.8%
6	identify the true definition of the term (pampas)	0.253	82.9%	75.0%	−7.9%
7	recognize the area of greatest distortion on the map	0.223	82.2%	90.9%	8.8%
8	select the true names of objects displayed on a blank map (of Poland)	0.297	50.9%	43.9%	−7.0%
9	select a transect on a map that corresponds to the displayed elevation profile (North America)	0.253	67.3%	68.9%	1.7%
10	recognize the difference in elevation between two points on a topographic map (imaginary area)	0.409	64.7%	62.1%	−2.6%
11	select a transect on a map that corresponds to the displayed elevation profile (imaginary area)	0.335	59.9%	65.9%	6.1%
12	decide which of the drawn watercourses is real (imaginary area)	0.275	56.9%	56.1%	−0.8%
13	select the characteristic that corresponds to both given countries (Norway and Chile)	0.260	75.8%	62.1%	−13.7%
14	select the local time difference between two locations (defined by geographical coordinates)	0.223	69.5%	78.0%	8.5%
15	select the season during which the Sun is at a given height above the horizon on a given parallel	0.230	48.0%	31.1%	−16.9%
16	identify true information (about exogenous processes)	0.320	59.9%	52.3%	−7.6%
17	identify the factor responsible for the difference in climatic conditions	0.283	49.4%	41.7%	−7.8%
18	decide what the difference is between two terms	0.245	78.1%	75.0%	−3.1%
19	recognize which factor is responsible for the difference in salinity and how salinity is affected by it.	0.223	65.8%	58.3%	−7.5%
20	choose a pair of languages that are most easily understood by each other	0.372	38.7%	34.1%	−4.6%
21	select the process name represented by the diagram (suburbanization)	0.216	75.5%	82.6%	7.1%
22	select a description for the data in the table (imaginary data)	0.208	74.7%	69.7%	−5.0%
23	select the relevant area according to the description (protected landscape area)	0.327	42.8%	40.2%	−2.6%
24	determine the relative position of two places (defined by geographical coordinates)	0.372	67.7%	71.2%	3.6%
25	select a diagram that corresponds to reality (stream in a river)	0.141	71.0%	68.2%	−2.8%
	Average	0.274	63.4%	60.3%	−3.1%

success rate) or negative (decrease in success rate) value. Subsequently, the difference in success rate was calculated for each group of test tasks, which were divided according to the above criteria.

5. Results

The research results show that the overall success rate of candidates in solving tasks was lower by 3.1% after eight years. However, visible differences can be identified not only between individual tasks, but also when dividing these tasks into groups according to thematic focus, educational objectives, use of visual elements or use of mathematical skills. While some groups of tasks recorded a visible decrease, only minimal changes or even improvements were visible in other groups of tasks. Detected changes in the success of candidates in solving various types of tasks can also be understood as changes in the geographic literacy of a selected group of respondents of the tested sample.

5.1 Results for individual tasks

The basic characteristics of the individual test tasks and their brief description are shown in Tab. 2. The ULI values for each task are also listed. The analysis of 25 test tasks that were repeated in 2024 found that the ULI of most of these selected test tasks was higher than 0.2 (the exception was only two test tasks with ULI = 0.156 and 0.141). It can therefore be stated that test tasks with a relatively high value of the ULI discrimination index were selected for comparison, i.e., test tasks of relatively high quality, with a high level of sensitivity.

Based on the success rate values, most tasks can be rated as moderately difficult, average or moderately easy. Only a few tasks had a success rate lower than 40%, in 2016 the lowest success rate was 38.7% (task no. 20), in 2024 the lowest success rate was 31.1% (task no. 15). We can therefore say that the test did not contain extremely difficult tasks. Similarly, very easy tasks, with a success rate higher than 80%, were represented only exceptionally. The highest success rate in 2016 was 82.9% (task no. 6), in 2024 it was 90.2% (task no. 7). The largest decrease in the success rate in the monitored period occurred in task no. 15 (−16.9%), while the largest improvement occurred in task no. 7 (+8.8%).

5.2 Results for specified task groups

As mentioned above, the results of candidates who took the entrance exams in 2024 were 3.1% worse overall than those who took the entrance exams in 2016. However, a summary evaluation of the different test task types shows that some groups experienced a more visible deterioration, while others experienced only slight changes and in one case even improved results. The results of the summary assessment according to various groups of test tasks is shown in Tab. 3.

The greatest decline in success occurred in test tasks thematically focused on physical geography and planetary geography. In physical geography tasks, there was a decrease in success rates for all six test tasks that were included in this category. In the case of planetary geography, while one of the test task saw

Tab. 3 Success rate of different groups of test tasks and its changes between 2016 and 2024.

Group of test tasks		Number of tasks in the group	Average success rate (2016)	Average success rate (2024)	Average change of the success rate (2016–2024)
thematic focus	planetary geography	n = 3	58.2%	51.5%	−6.7%
	physical geography	n = 6	67.8%	61.7%	−6.1%
	social geography	n = 3	62.9%	62.1%	−0.8%
	regional geography	n = 7	60.9%	55.5%	−5.4%
	cartography	n = 6	64.7%	67.9%	3.2%
prevailing goal according to RBT	remember	n = 9	60.1%	53.5%	−6.6%
	understand	n = 4	73.5%	71.4%	−2.1%
	apply	n = 12	63.1%	61.8%	−1.3%
knowledge dimension according to RBT	factual	n = 11	62.4%	56.8%	−5.5%
	conceptual	n = 10	66.0%	65.0%	−1.0%
	procedural	n = 5	59.9%	58.1%	−1.7%
other characteristics	with visual element	n = 10	65.2%	65.0%	−0.2%
	without visual element	n = 15	62.2%	57.2%	−5.0%
	with mathematical skills	n = 4	59.9%	58.1%	−1.7%
	without mathematical skills	n = 21	64.1%	60.7%	−3.4%

an increase in success rates (a task focused on calculations related to time zones), two tasks saw a visible deterioration in in results – for task no. 15 it was 16.9%, which was the largest decline of all the monitored tasks (this question focused on determining the height of the Sun above the horizon).

Conversely, the only category in which there was an increase in success was the test tasks on cartography. In this category, there was an improvement in results in five out of six test tasks evaluated.

In terms of the predominant objectives according to the RBT, the greatest decrease in success occurred in test tasks that follow objectives in the Remember category and in tasks that verify the factual knowledge dimension. The decrease in success was relatively small in tasks that follow objectives in the Understand and Apply categories and in tasks that verify the conceptual and procedural knowledge dimensions.

Visible differences were noted when test tasks were divided according to whether they contained visual elements and whether mathematical skills were required to solve them. Tasks that did not contain visual elements showed a noticeable decrease in success, while tasks that contained visual elements showed only a minimal decrease in success. Less considerable differences were seen when test tasks were divided according to whether their solution required mathematical skills, but even here differences were noticeable, with a smaller decrease in success for tasks that required some mathematical skills.

6. Discussion

This research provides some results worthy of discussion. We consider an important finding to be a visible decrease in success for tasks focused on physical geography and planetary geography and, conversely, the relative increase in success for tasks focused on cartography.

However, it is important to note that thematic focus is not independent of the knowledge dimension according to RBT: in our dataset, tasks classified as physical or planetary geography are predominantly located within the factual knowledge dimension, whereas cartographic tasks more frequently involve procedural or conceptual knowledge. This overlap reflects the broader curricular shift described by Bendl et al. (2024), who point out that the concept of geographical thinking is shifting from the dominance of factual knowledge towards the development of geographical competences, which also requires a gradual redefinition of geographic literacy. Therefore, the differences in success rates across thematic areas may partly reflect these underlying differences in cognitive demands rather than purely thematic characteristics.

In the context of the ongoing reform of the Czech education system, which, among other things, aims to increase the importance of higher cognitive goals compared to simply memorizing facts (for more details, e.g. Trahorsch and Korvasová 2023), the findings of our analysis in the area of educational goals according to RBT can be seen as encouraging, which shows that tasks verifying higher cognitive goals according to RBT showed a smaller decrease in success compared to tasks focused on memorization and the factual knowledge dimension. Similar findings also result from research conducted by Bláha et al. (2024), who analyzed the exam results of university students in the subject of cartography and found that students achieve the worst results in tasks testing factual knowledge, while achieving the best results in procedural knowledge.

On an international scale, Binimelis Sebastián et al. (2024) document that mechanical learning in Spanish schools does not lead to sustainable geographic literacy, and call for approaches that are more active based on working with maps and strengthening spatial competencies, which also corresponds to the intention of the Czech reform.

One may speculate whether the minimal decrease in success rates for tasks with visual elements compared to the more substantial decrease in success rates for tasks without visual elements is related to the higher level of orientation of the younger generation towards audiovisual elements (for more details, e.g. Ateiku et al. 2023; Setyani et al. 2021). Similarly, the causes of the lower decrease in success rates for tasks requiring mathematical skills compared to tasks that do not require mathematical skills may be various, and drawing more serious conclusions would require the analysis of a larger number of tasks over a longer time period. Bláha et al. (2024) found somewhat different results in their research in this case – according to their findings, students achieved worse results in tasks requiring more complex mathematical operations. They also found that the inclusion of visual elements had minimal impact on the success of tasks, but in oral exams, students preferred questions where visualization can be applied or where visual elements are present.

A more detailed analysis would also be needed to examine the potential connection between the increase in success on tasks focused on cartography and the relatively lower decrease in success on tasks with visual elements and on tasks requiring mathematical skills. Cartography tasks usually contained a visual element and solving some of them also required the use of mathematical skills. However, to examine whether these characteristics are related, a larger research sample should be analyzed.

The importance of such analyses is demonstrated by longitudinal studies abroad, such as the NAEP research in the USA (Solem et al. 2021; National Center for Education Statistics 2020). These findings underline that the challenge of strengthening geographic literacy is not unique to the Czech context but represents a broader international trend.

6.1 Limitations

We are aware that our study has many limitations and its results should be understood as such. The main limitation of the research is the low number of evaluated test tasks (n = 25). For this reason, the presented study should be understood as a pilot study, the purpose of which was mainly to examine the possibilities of evaluating year-on-year changes in success according to various criteria. The results cannot therefore be considered statistically significant and no fundamental conclusions can be drawn from them. Nevertheless, it can be stated that even the results of the analysis of such a limited sample of examined tasks are stimulating.

Another limitation of the research results from the fact that in the monitored period there was a certain expansion of the group of applicants who were exempted from the entrance exam due to participation in higher levels of subject knowledge competitions (e.g. participation in the national level of the Geography Olympiad, the Secondary School Research Project Competition and other competitive activities). Another reason is the possibility of waiving the entrance exam based on the grade average in selected subjects in secondary school. The percentage of students admitted in this way varies annually, but the value is around 10%.

Certain limitations may also result from the characteristics of the test used. Due to their limited number, the test tasks do not cover the entire breadth of the given disciplines and are not thematically balanced (the proportional representation according to thematic focus is not the same). The same applies to the representation of different types of questions according to educational objectives. There is also a lack of variability in the different types of questions: all questions are multiple choice. However, this results from the nature of the test, because the admission test must be clearly specified and clearly evaluable.

The establishment of categories for classifying test tasks according to educational objectives and the actual classification of tasks into these categories could also be discussed. We are aware that the classification of tasks into individual categories can be subjective, despite the above-described procedure, the aim of which was to minimize the degree of subjectivity. It would also be useful to try out other taxonomies of educational objectives (e. g. the very detailed classification of Tollingerová 1971) for the evaluation of multiple-choice test tasks and to determine the most suitable ones for this purpose. However, this would require a more comprehensive analysis, which is beyond the scope of our study. We consider this to be an interesting challenge and topic for further research.

7. Conclusions

The results of our study showed that between 2016 and 2024, when the same test tasks were administered to candidates applying to study geography at the Faculty of Science at Charles University, there was a certain decrease in success rates. However, this decrease was manifested to varying degrees in different types of test tasks, specifically when classifying tasks into different categories according to thematic focus, the educational objectives pursued according to the RBT, as well as other criteria.

While there was a considerable decrease in success rates for test tasks focused on physical geography and planetary geography, there was a relative increase in success rates for tasks focused on cartography. There was a greater decrease in success rates for tasks that tested rote memorization and the factual knowledge dimension, compared to tasks that tested the achievement of higher goals according to the RBT. This finding may indicate a certain shift in the structure of geographic literacy, where tasks based on rote memorization of facts are becoming more difficult for current applicants, while tasks requiring conceptual understanding and application skills are relatively less affected. This trend corresponds to the broader discussion on the transition from factual knowledge towards geographical competences. There was a relatively lower decrease in success rates for tasks that contained visual elements and for tasks whose solution required mathematical skills than for tasks that did not contain visual elements and whose solution did not require the application of mathematical skills. This suggests that candidates may be more accustomed to working with visual and quantitative representations, possibly reflecting their broader experience with audiovisual materials, digital technologies and data-based school assignments.

The results presented here should be viewed with some caution, as they are based on a comparison of only two years of candidates. The uniqueness of our study, however, lies in the fact that the test tasks assessed were identical in both years. Nevertheless, the analysis shows that monitoring year-on-year changes in success rates can provide valuable feedback not only for the refinement of entrance tests (e.g. balancing task types, cognitive levels or thematic coverage) but also for secondary school education or curriculum development.

This pilot study showed that year-on-year comparison of candidates' success rates has considerable potential, and the analysis of a larger volume of data could in the future provide inspiring insights for various readers. Future research should therefore aim to include a larger sample of tasks and more time points, which would allow for more reliable identification of long-term trends in the evolution of geographic literacy among Czech students.

References

Amer, A. (2006): Reflections on Bloom's Revised Taxonomy. Electronic Journal of Research in Educational Psychology 4(1), 213–230.

Anderson, L. W., Krathwohl, D. R. (Eds.) (2001): A Taxonomy for Learning, Teaching and Assessing. New York: Longman.

Ateiku, J. N., Segbefia, S. K., Gyampoh, A. O. (2023): Effect of Audio-visual Aids on Students Understanding: A Comparative Analysis of JHS Students in the Ho Municipality. Integrated Journal for Research in Arts and Humanities 3(5), https://doi.org/10.55544/ijrah.3.5.4.

Bartáková, I., Chvál, M., Martinková, P. (2018): Predikční validita zkoušek zakončujících středoškolské vzdělávání a přijímacích testů na vysoké školy. Pedagogika 68(1), 66–85, https://doi.org/10.14712/23362189.2017.1016.

Bendl, T., Krajňáková, L., Marada, M., Řezníčková, D. (2024): Geographical thinking in geography education: a systematic review. International Research in Geographical and Environmental Education, 1–19, https://doi.org/10.1080/10382046.2024.2354097.

Binimelis Sebastián, J., Gómez Gonçalves, A., Gómez Trigueros, I. M., Muntaner Guasp, J. J. (2024): Geographic literacy in third-year compulsory secondary education students. Journal of Baltic Science Education 23(6), 1119–1133, https://doi.org/10.33225/jbse/24.23.1119.

Bláha, J. D., Trahorsch, P., Bartůněk, M., Hladík, P. (2024): Critical issues in undergraduate cartographic education: Analysis of final tests and oral examinations. AUC Geographica 60(1), 61–74, https://doi.org/10.14712/23361980.2024.19.

Brožová, H., Rydval, J. (2014): Analysis of exam results of the Subject. Applied Mathematics for IT. Journal on Efficiency and Responsibility in Education and Science 7(3–4), 59–65, https://doi.org/10.7160/eriesj.2014.070303.

Byčkovský, P., Kotásek, J. (2004): Nová teorie klasifikování kognitivních cílů ve vzdělávání: revize Bloomovy taxonomie. Pedagogika 54(3), 227–242 (in Czech).

Camara, W., Kimmel, E. W. (Eds.) (2005): Choosing Students: Higher Education Admissions Tools for the 21st Century (1st ed.). Routledge, https://doi.org/10.4324/9781410612533.

Cortina, J. M. (1993): What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology 78(1), 98–104, https://doi.org/10.1037/0021-9010.78.1.98.

Cloke, P., Crang, P., Goodwin, M. (2005): Introducing Human Geographies. London: Hodder Education, https://doi.org/10.4135/9781446221235.

Dikmenli, Y. (2014): Geographic literacy perception scale (GLPS) validity and reliability study. Mevlana International Journal of Education 4(1), 1–15, http://dx.doi.org/10.13054/mije.13.43.4.1.

Frey, M. C., Detterman, D. K. (2004): Scholastic Assessment or g? The Relationship Between the Scholastic Assessment Test and General Cognitive Ability. Psychological Science 15(6), 373–378, https://doi.org/10.1111/j.0956-7976.2004.00687.x.

Hampl, M. (2000): Sociální geografie: hledání předmětu studia. Geografický časopis 52(1), 33–40 (in Czech).

Chráska, M. (1999): Didaktické testy: příručka pro učitele a studenty učitelství. Brno: Paido, ISBN 80-85931-68-0 (in Czech).

Kalhous, Z., Obst, O. (2009): Školní didaktika. Praha: Kosmas, ISBN 978-80-7367-571-4 (in Czech).

Kjærnsli, M., Lie, S. (2004): PISA and scientific literacy: Similarities and differences between the Nordic countries. Scandinavian Journal of Educational Research 48(3), 271–286, https://doi.org/10.1080/00313830410001695736.

Krumrei-Mancuso, E. J., Newton, F. B., Kim, E., Wilcox, D. (2013): Psychosocial factors predicting first-year college student success. Journal of College Student Development 54(3), 247–266, https://doi.org/10.1353/csd.2013.0034.

Logayah, D. S., Ruhimat, M., Arrasyid, R., Islamy, M. R. F. (2024): Item analysis of National Geography Olympiad multiplechoice questions (MCQs) in Indonesia. Cogent Social Sciences 10(1): 2354971, https://doi.org/10.1080/23311886.2024.2354971.

Lucky, A., Janštová, V., Novotný, P., Mourek, J. (2025): Quantifying ento-literacy: development and validation of an international insect-focused attitude and knowledge survey instrument. International Journal of STEM Education 12: 11, https://doi.org/10.1186/s40594-025-00532-8.

McFarlane, D. A. (2024): Critical Review of Geographic Illiteracy in the United States: A Multifaceted Analysis. Journal of Geography, Environment and Earth Science International 28(1), 11–25, https://doi.org/10.9734/jgeesi/2024/v28i1741.

National Center for Education Statistics (2020): The Nation's Report Card: 2018 U.S. History, Geography, and Civics at Grade 8 – Highlights Reports (NCES 2020017). U.S. Department of Education. Available online: https://nces.ed.gov/use-work/resource-library/report/statistical-analysis-report/nations-report-card-2018-u-s-history-geography-and-civics-grade-8-highlights-reports/ (accessed on 25 March 2025).

Niemierko, B. (1979): Taksonomie celów wychowania. Kwartalnik pedagogiczny 24(2), 66–67 (in Polish).

Ørberg, J. W. (2018): Uncomfortable encounters between elite and "shadow education" in India – Indian Institutes of Technology and the Joint Entrance Examination coaching industry. Higher Education 76, 129–144, https://doi.org/10.1007/s10734-017-0202-5.

Pérez-Benedito, J. L., Aragón, E. Q., Alriols, J. A., Medic, L. (2014): Optical Mark Recognition in Student Continuous Assessment. IEEE Revista Iberoamericana de Tecnologías del Aprendizaje 9(4), 133–138, https://doi.org/10.1109/RITA.2014.2363005.

Průcha, J. (2009): Pedagogická encyklopedie. Praha: Portál, EAN 9788073675462 (in Czech).

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., Carlstrom, A. (2004): Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis. Psychological Bulletin 130(2), 261–288, https://doi.org/10.1037/0033-2909.130.2.261.

Řezníčková, D. (2003): Geografické dovednosti, jejich specifikace a kategorizace. Geografie 108(2), 146–163, https://doi.org/10.37040/geografie2003108020146 (in Czech).

Schindler, R., Brož, F., Hulešová, M., Chvál, M., Kovařík, Š., Matoušek, J., Mičienka, M., Smetáčková, I. (2006):

Rukověť autora testových úloh. Praha: Centrum pro zjišťování výsledků vzdělávání, ISBN 80-239-7111-5 (in Czech).
Setyani, W. A., Jumadi, J., Darmawan, A. S. (2021): The Implementation of Audio Visual Media in Problem Based Learning Model to Improve the Problem Solving Skills.
Proceedings of the 6th International Seminar on Science Education (ISSE 2020). Advances in Social Science, Education and Humanities Research 541, 563–568, https://doi.org/10.2991/assehr.k.210326.081.
Shi, W.-Z., He, X., Wang, Y., Fan, Z. G., Guo, L. (2016): PISA and TIMSS Science Score, Which Clock is More Accurate to Indicate National Science and Technology Competitiveness? Eurasia Journal of Mathematics, Science and Technology Education 12(4), 965–974, https://doi.org/10.12973/eurasia.2016.1239a.
Soleh, T. A., Ruhimat, M., Yani, A. (2022): Factor analysis: geographic literacy level of high school students in North Bandung area. International Journal of Research and Review 9(12), 519–528, https://doi.org/10.52403/ijrr.20221257.
Solem, M., Vaughan, P., Savage, C., De Nadai, A. S. (2021): Student- and School-Level Predictors of Geography Achievement in the United States, 1994–2018. Journal of Geography 120(6), 201–211, https://doi.org/10.1080/00221341.2021.2000009.
Silva, P., Nunes, L. C., Seabra, C., Balcao Reis, A., Alves, M. (2020): Student selection and performance in higher education: admission exams vs. high school scores. Education Economics, 28(5), 437–454, https://doi.org/10.1080/09645292.2020.1782846.
Šrámek, M., Teplá, M. (2021): The Qualities of Tasks in General Chemistry Based on the Analysis of Entrance Procedure at the Faculty of Science, Charles University. Scientia in Educatione 12(2), 32–40, https://doi.org/10.14712/18047106.2068.
Šrámek, M., Teplá, M. (2022): The qualities of tasks in organic chemistry and biochemistry based on the analysis of entrance procedure at the Faculty of Science, Charles University. Scientia in Educatione 13(2), 51–63, https://doi.org/10.14712/18047106.2219.
Štuka, Č., Vejražka, M. (2022): Testování a hodnocení studentů na VŠ. Praha: Univerzita Karlova, https://doi.org/10.14712/9788024651088 (in Czech).
Tavakol, M., Dennick, R. (2011): Making sense of Cronbach's alpha. International Journal of Medical Education 2, 53–55, https://doi.org/10.5116/ijme.4dfb.8dfd.
Teig, N., Scherer, R., Olsen, R. V. (2022): A systematic review of studies investigating science teaching and learning: over two decades of TIMSS and PISA. International Journal of Science Education 44(12), 2035–2058, https://doi.org/10.1080/09500693.2022.2109075.
Tollingerová, D. (1971): Úvod do teorie a praxe programované výuky a výcviku. Odborná výchova 21(5), 143–146 (in Czech).
Trahorsch, P., Korvasová, V. (2023): Jaké je postavení geografie v probíhající kurikulární reformě? Diskuze klíčových témat geografického vzdělávání v rámci revize Rámcového vzdělávacího programu pro základní vzdělávání. Informace ČGS 42(1), 1–11 (in Czech).
Vašková, M. (2013): Přijímací řízení na vysoké školy. Diplomová práce. Masarykova univerzita, Právnická fakulta, Brno. Available online: https://theses.cz/id/fsffjy/ (accessed on 10 March 2025, in Czech).
Viktorová, L., Charvát, M. (2014): K problematice validity testů studijních předpokladů. Testfórum 3, 3–15, https://doi.org/10.5817/TF2014-3-13 (in Czech).
Weissberg, N. C., Owen, D. R. (2005): Do Psychosocial and Study Skill Factors Predict College Outcomes? Comment on Robbins et al. (2004). Psychological Bulletin 131(3), 407–409, https://doi.org/10.1037/0033-2909.131.3.407.
Wu, M. (2010): Comparing the Similarities and Differences of PISA 2003 and TIMSS. OECD Education Working Papers 32. Paris: OECD Publishing, https://doi.org/10.1787/5km4psnm13nx-en.
Zhang, H. (2023): A Comparative Analysis of the 2018–2021 College Entrance Examination Mathematics National Paper (Science) questions. International Journal of Education and Humanities 8(1), 122–130, https://doi.org/10.54097/ijeh.v8i1.7244.
Zhang, L., Liu, X., Feng, H. (2023): Development and validation of an instrument for assessing scientific literacy from junior to senior high school. Disciplinary and Interdisciplinary Science Education Research 5: 21, https://doi.org/10.1186/s43031-023-00093-2.

Machine learning model for stage-discharge curve calculation

Jakub Langhammer*, Miroslav Šobr, Doudou Ba

Charles University, Faculty of Science, Department of Physical Geography and Geoecology, Czechia
* Corresponding author: jakub.langhammer@natur.cuni.cz

ABSTRACT

Stage-discharge relationships (rating curves) are fundamental in hydrology but remain challenging to establish in experimental catchments, where observations are sparse, irregular, and uncertain. Conventional regression models provide simple and interpretable solutions, yet often fail to capture nonlinearities in hydraulically complex environments. Purely data-driven machine learning (ML) models offer flexibility, but their performance deteriorates under data scarcity and they often produce physically implausible results. We present a hybrid physics-informed machine learning (PIML) framework that integrates a log-log regression baseline with residual corrections from Support Vector Regression (SVR) and Multilayer Perceptron (MLP) models. By embedding hydrological constraints such as monotonicity, non-negativity, and continuity, the framework ensures physically consistent rating curves while leveraging ML to capture nonlinear deviations. The approach was developed in four contrasting catchments and validated across 20 independent evaluation sites. Results show that both hybrid models outperform conventional regression, with the Hybrid MLP consistently providing the most accurate and generalizable predictions (median R^2 and NSE > 0.98) even when calibrated with as few as 8–15 discharge measurements. The framework is particularly effective in irregular or hydraulically complex basins, while differences to conventional regression are minimal in stable profiles. These findings demonstrate that PIML enables systematic, transferable, and reproducible rating curve development under sparse and uncertain data conditions. The framework offers a practical alternative to subjective or ad hoc methods, advancing discharge estimation in experimental hydrology and supporting applications in data-limited and hydraulically complex environments.

KEYWORDS

rating curve; discharge; water level; machine learning; physics informed model

Received: 9 September 2025
Accepted: 11 November 2025
Published online: 10 December 2025

Langhammer, J., Šobr, M., Ba, D. (2025): Machine learning model for stage-discharge curve calculation.
AUC Geographica 60(2), 296–315
https://doi.org/10.14712/23361980.2025.25

1. Introduction

Stage-discharge relationships, often referred to as rating curves, are a cornerstone of open channel hydrology. These curves provide a means to estimate river discharge based on measurements of water stage, forming the basis for a wide array of hydrological analyses, including flood forecasting, water resource management, and the design of hydraulic infrastructure. Discharge ratings for gauging stations are usually determined empirically by means of discharge measurements made in the field (Kennedy 1984). Common practice is to measure the discharge of the stream periodically, usually by current meter or Acoustic Doppler Current Profiler (ADCP) system in case of flood events, and to note the concurrent stage. Measured discharge is then plotted against a concurrent stage on graph paper to define the rating curve. At a new station many discharge measurements are needed to define the stage discharge relation throughout the entire range of stage (WMO 2020).

The long-standing development of stage-discharge estimation has led to a variety of methodological approaches, from empirical regression models to physically-based hydraulic simulations. Traditional techniques for constructing rating curves have predominantly relied on simple statistical relationships, such as linear and polynomial regression. These models are favored for their computational simplicity and ease of implementation.

The development of robust stage-discharge modeling techniques is particularly important in experimental catchments. Unlike conventional gauging stations, which are typically maintained by water agencies under standardized and long-term monitoring protocols, experimental profiles are often constrained by numerous limitations. Nevertheless, they represent a critical source of information for hydrological research and water management, particularly in basins where no other data are available.

Experimental catchments are typically characterized by small drainage areas, steep gradients, dynamic flow regimes, and irregular channel morphologies, but also by challenges in data collection. Measurements are often sparse due to the remote locations, and acquired irregularly during episodic field campaigns, in different years, using various instruments and by different surveyors. All these factors contribute to observational noise and increase the uncertainty of the resulting data.

Although such data are often far from optimal quality and consistency when compared to standard hydrometric records, they are typically the only available observations for the catchments. Their scientific and practical value, particularly in understanding hydrological processes in understudied environments, remains substantial. Therefore, there is a need to identify and develop modeling approaches that are both accurate and robust, capable of coping with irregularities, limited sample sizes, and inherent measurement uncertainties. Such methods must ensure physical plausibility while allowing generalization across diverse hydrological conditions, making them suitable for applications in data-scarce, experimental settings.

Conventional regression models, such as power-law or polynomial fits, often fail to capture nonlinear behavior in irregular channels, leading practitioners to rely on ad hoc corrections (e.g., piecewise regression, correction factors, localized hydraulic modeling (Di Baldassarre and Montanari 2009; Dobrovolski et al. 2022; Kiang et al. 2018).

This niche creates an opportunity for the application of machine learning models, which are capable of learning complex, nonlinear relationships directly from data and adapting to site-specific conditions without requiring explicit physical formulations (Bhasme et al. 2022; Nearing et al. 2021; W. Xu et al. 2024). While ML models offer strong predictive power and flexibility, they face substantial challenges:

(i) ML model performance under data-sparse conditions. A significant limitation of conventional ML models is their reliance on large volumes of high-quality training data, which are typically unavailable in stage-discharge monitoring, especially in experimental catchments. For example, montane basins or experimental watersheds in remote areas frequently lack the infrastructure for long-term gauging. To address this, recent studies have explored regional modeling and data synthesis techniques (Poulinakis et al. 2023).

(ii) Lack of integration of physical laws into ML models. A persistent issue with black-box ML models is their disregard for elementary hydrological principles, such as the monotonic increase of discharge with stage or conservation of mass. Traditional models impose these physically meaningful constraints explicitly, whereas many ML models often fail to enforce them due to their purely statistical nature. This drawback has driven the development of PIML modeling approaches, which embed hydrological knowledge into learning processes (Bhasme et al. 2022; W. Xu et al. 2024).

(iii) Poor transferability across catchments. ML models are typically site-specific, limiting their applicability across regions with differing hydrological, topographic, or climatic characteristics. Although some attempts have been made to model across regions using large-sample learning (Kratzert et al. 2019b), cross-basin generalization remains an unsolved challenge.

To address these challenges, we propose a PIML framework for stage-discharge curve estimation, with the following objectives:

(i) To develop a ML-based model that incorporates physical constraints relevant to hydrological processes,

(ii) To test and validate the model across four hydrologically and topographically diverse catchments,
(iii) To benchmark the performance of the proposed model against conventional regression-based approaches.

In recent years, the emergence of PIML models has begun to reshape hydrological modeling by embedding physical principles into data-driven frameworks (Feng et al. 2023; Kratzert et al. 2019b). These approaches address key criticisms of purely statistical models by constraining predictions to obey mass conservation, monotonicity, and other hydrological laws. Applications of PIML in hydrology have so far focused primarily on streamflow prediction in large-sample datasets and ungauged basins, demonstrating improved robustness and physical plausibility compared to conventional machine learning (Bhasme et al. 2022; Esmaeilzadeh and Amirzadeh 2024). However, their use with sparse data, such as for rating curve development in experimental catchments, remains unexplored.

To address the aforementioned challenges, we developed and tested a physics-informed machine learning (PIML) model for stage-discharge estimation. The proposed framework integrates the structure of conventional rating curve models, used here as a baseline, with machine learning techniques that enhance adaptability and robustness to uncertainty. By explicitly enforcing key hydrological constraints such as monotonicity and non-negativity, the model ensures physically reliable behavior even in irregular or data-scarce conditions. We propose that physics-informed machine learning, despite its conventional requirement for large datasets, can provide a more systematic, objective, and reproducible framework as an alternative approach for extracting maximum information from sparse but high-quality measurements.

To evaluate the model performance and generalizability, we applied the model across four hydrologically and geomorphologically distinct test catchments: a steep high-gradient alpine stream in the Tien Shan Mountains, a mid-latitude montane basin in Central Europe, a low-gradient small stream in rural landscape, and a larger lowland river. These diverse test cases were selected to reflect the broad range of environments in which conventional methods often fail, providing a validation of model performance. Consequently, the model was applied to a set of 20 independent assessment sites in the mid-latitude experimental catchments to verify the model‘s applicability to real-world conditions.

2. Materials and methods

2.1 Rating curve modeling

Rating curve development is a central task in hydrometry, providing the functional relationship between river stage and discharge. This relationship may be derived empirically, from hydraulic theory, or by combining both approaches (Herschy 2019; ISO 2021).

Reliable hydrometric measurements form the basis of rating curve construction, with erroneous or inconsistent observations removed during data screening to ensure internal consistency (WMO 2010). The stage-discharge relationship is typically represented using polynomial, exponential, or piecewise functions, and dividing the curve into height zones is often necessary to capture regime shifts across flow stages (Herschy 2019; Lane 1998).

Empirical methods rely on plotting stage-discharge pairs and fitting a smooth curve, but this procedure is inherently subjective and sensitive to sparse or noisy datasets. Power-law formulations remain the standard in operational hydrometry, with parameters estimated through nonlinear regression (Braca 2008; Hrafnkelsson et al. 2022). Hydraulic methods, in contrast, derive the relationship from physical principles, for example using Manning's equation with cross-sectional geometry, slope, and roughness (ISO 2021). Statistical regression approaches, often based on log-log transformations, reduce subjectivity and provide diagnostic measures of model fit (WMO 2010).

In practice, rating curve modeling reflects a trade-off between empirical fitting and physical reasoning. Manual and regression-based approaches can reproduce observed patterns but often lack robustness under sparse or uncertain data conditions. Hydraulic approaches ensure physical consistency but depend on detailed channel information that is not always available. Traditional regression models may also struggle in irregular natural channels, where observed discharges deviate markedly from theoretical assumptions (Ali and Maghrebi 2023; Di Baldassarre and Montanari 2009).

Recent advances in machine learning (ML) provide alternative strategies for rating curve estimation. ML algorithms are effective at capturing nonlinear relationships and handling large, complex datasets (Liu et al. 2022). However, their application in stage-discharge modeling is constrained by the data-driven nature of these methods. Direct discharge measurements are often sparse and uncertain, especially in experimental catchments, which increases the risk of overfitting and can lead to physically implausible results (Ali and Maghrebi 2023; Feng et al. 2023; Roelofs et al. 2019; Ying 2019).

2.2 Proposed hybrid physics-informed ML model

2.2.1 Model principles

To address the limitations of both traditional regression-based rating curves and purely data-driven ML approaches, we developed a hybrid PIML framework. This approach integrates a theoretical baseline model with ML-based residual corrections, thereby preserving physical consistency while capturing

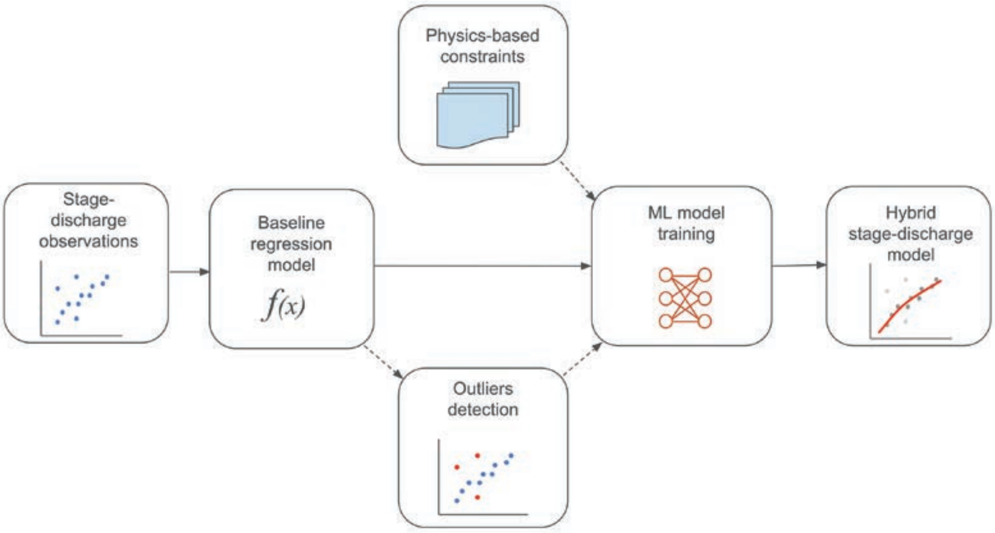

Fig 1 Principal steps in the design of the hybrid rating curve model.

nonlinear deviations (Feng et al. 2023; Raissi et al. 2019).

The framework consists of two components: (i) a log-log regression baseline that represents the primary stage-discharge relationship, and (ii) an ML residual model trained on deviations between observed and baseline predictions. By embedding physical constraints – non-negativity, monotonicity, continuity, and bounded parameter ranges – the approach ensures hydrological realism while allowing ML to refine predictions. Final discharges are obtained by combining baseline predictions with corrected residuals, resulting in smooth rating curves suitable for operational use (Fig. 1).

ML algorithm is then used for fitting the model to the real-world distribution, while being trained on the residuals between the baseline model and the observed values. The ML residual model is trained on the deviations between observed discharges and baseline predictions. To ensure physically consistent behavior, three constraints are enforced: (i) discharge remains non-negative, (ii) discharge increases monotonically with stage, and (iii) rating curves are continuous and site-specific, and (iv) basic shape of the stage-discharge curve is defined by a theoretical distribution. These preserve hydrological realism while allowing the ML model to capture nonlinear deviations. Final predictions are generated by combining the baseline log-log regression model outputs with the machine learning residual predictions.

2.2.2 Baseline model

For the selection of baseline regression models, we considered approaches commonly used in stage-discharge curve reconstruction, namely power law, polynomial regression, and log-log regression (Herschy 2019; WMO 2010). The log-log regression form has been shown to be comparatively more robust under sparse and noisy data conditions, owing to its stabilizing transformation of both variables (Kiang et al. 2018; Le Coz et al. 2014). Baseline model was thus established through a log-log regression model that captures the primary relationship between water levels and discharge measurements. This baseline model follows the form:

$$Q = exp\ exp\ (b) \times (H - H_0)^a,$$

where Q is discharge, H stage, H_0 a reference stage, and a, b, fitted parameters.

Parameters are optimized using non-linear least squares with bounds $a \in [0.1, 5.0]$ and the log coefficient bounded $b \in [-10, 10]$, and H_0 set to 90% of the minimum observed stage. These constraints prevent unrealistic scaling while ensuring positive effective depths. The baseline captures fundamental hydraulic behavior and provides the physical structure for ML refinement.

This baseline model captures the primary hydraulic behavior and serves as the physical constraint for subsequent machine learning corrections, ensuring that any data-driven improvements operate within a physically consistent framework rather than attempting to learn fundamental hydraulic principles solely from observations.

To extend the single-segment log-log baseline, a piecewise power-law model was implemented to represent gradual changes in the slope of the stage-discharge relationship.

The model divides the relationship into several monotonic segments, each described by its own power-law parameters, with continuity enforced at the breakpoints. Model parameters and breakpoint locations were estimated by nonlinear least-squares fitting in log-transformed space. A configuration of three connected segments provided adequate flexibility while avoiding overfitting, and the model served as both an independent benchmark and a reference baseline for the hybrid machine-learning models.

2.2.3 Physics constraints

Physical realism of the model is maintained by enforcing key constraints.

First, monotonicity is enforced to guarantee that the stage-discharge relationship is non-decreasing. This is achieved via sorting and validation of the input data, followed by a post-processing correction with a tolerance of 0.01 m^3/s, leading to curve smoothness. Non-negativity is ensured by removing non-positive inputs and constraining predictions to a minimum discharge of 0.001 m^3/s. Continuity is achieved by interpolating 200 equally spaced points across the observed stage range, preventing discontinuities in curve generation.

The energy consistency conservation principle is preserved by embedding the log-log power-law formulation, which reflects energy conservation in open-channel flow and is consistent with Manning's equation and open channel hydraulic principles.

2.2.4 Residual modeling with machine learning

Eight ML algorithms were initially tested for residual modeling, representing a range of different methodological families used in hydrological applications (Kratzert 2019a; Mosavi et al. 2018; T. Xu and Liang 2021). Specifically, the Artificial Neural Networks (ANN), Multi-Layer Perceptron (MLP), Support Vector Regression (SVR), Random Forest (RF), K-Nearest Neighbors (KNN), Gaussian Process (GP), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost) were tested.

Performance was evaluated using R^2 and RMSE, complemented by qualitative assessment of physical plausibility, since some algorithms may produce non-monotonic curves under limited training data (Nearing et al. 2021; Shortridge et al. 2016).

The evaluation was based on observed stage-discharge data from primary test catchments, ensuring that algorithm performance was tested under data-poor conditions typical of experimental basins with variable physiographies. This approach enabled us to assess both statistical performance and physical plausibility, which guided the selection of algorithms for the PIML framework. As a result, two machine learning models, Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP) were selected for implementation in the hybrid model.

To address the systematic deviations from the baseline predictions, ML models were trained on the residuals between observed values and baseline predictions. To ensure physical consistency and non-negative discharge predictions, the residuals were log-transformed during the training process. Outlier removal was performed using the Interquartile Range (IQR) method to enhance model stability.

Two complementary machine learning approaches, selected from ML model testing, were implemented to provide robust evaluation of the physics-informed framework: Support Vector Regression (SVR) with Radial Basis Function kernel and Multi-Layer Perceptron (MLP) neural networks. This dual approach enables comprehensive assessment of different algorithmic families, allowing the data to determine optimal performance rather than assuming superiority of any particular method. SVR represents kernel-based learning with strong theoretical foundations for small datasets, while MLP represents neural network approaches with proven capability for complex nonlinear relationships.

2.2.5 Data preprocessing and model integration

Preprocessing removed missing or non-physical values, ensured correct ordering of stage-discharge pairs, and identified outliers using the interquartile range method. Residuals were log-transformed with offsets to handle negative values, then back-transformed for final predictions.

Hybrid predictions were obtained as

$$Q_{\mathrm{PIML}} = Q_{\mathrm{baseline}} + \mathrm{Residuals}_{ML},$$

ensuring that baseline hydraulic structure was preserved while allowing ML to capture systematic deviations.

Continuous rating curves were generated using 200 equally spaced stage values from the minimum observed stage to 110% of the maximum, with monotonicity corrections applied selectively to maintain smoothness. Post-processing includes gentle monotonicity enforcement that selectively corrects only when $Q(i) < Q(I - 1) - 0.01$, preserving natural curve smoothness while maintaining hydraulic principles. Final predictions are converted back through exponential transformation and offset removal, ensuring that the original residual scale is preserved while maintaining numerical stability throughout the machine learning process.

2.2.6 Model performance evaluation

Model performance was assessed using Coefficient of Determination (R^2), Nash-Sutcliffe Efficiency (NSE), and Index of Agreement (IoA). Comparisons between baseline and hybrid models quantified predictive improvement, while cross-validation across catchments demonstrated robustness under varying physiographic and data conditions.

The computational workflow processes each monitoring station individually through sequential data loading and validation, model fitting, prediction generation, and performance assessment. Error handling safeguarded against insufficient datasets (<10 points), non-convergence, and numerical instability, ensuring reliable model performance and consistent results across diverse sites.

Model evaluation was restricted to the observed stage range; no extrapolation beyond the highest measured stage was performed. The predictive uncertainty of the hybrid rating-curve models was quantified using a residual-based, non-parametric method adapted from Kiang et al. (2018). Residuals in log-discharge space were evaluated along the stage axis, and for each stage, the k nearest observations were used to derive 5th–95th percentile bounds of the residual distribution.

These local quantiles were smoothed and applied to the fitted curves to form 95% prediction intervals, common in hydrometry (ISO 2020), and providing a transparent, data-driven estimate of discharge uncertainty for models without native uncertainty propagation.

2.3 Study sites and data

The modeling framework was first developed and evaluated in four contrasting test catchments, representing diverse physiographic settings and hydrological regimes: a high mountain basin, a montane headwater catchment, a hilly agricultural stream, and a lowland river section (Fig. 2). To assess robustness and transferability, the approach was subsequently applied to 20 additional evaluation catchments, independent from model development, located in the

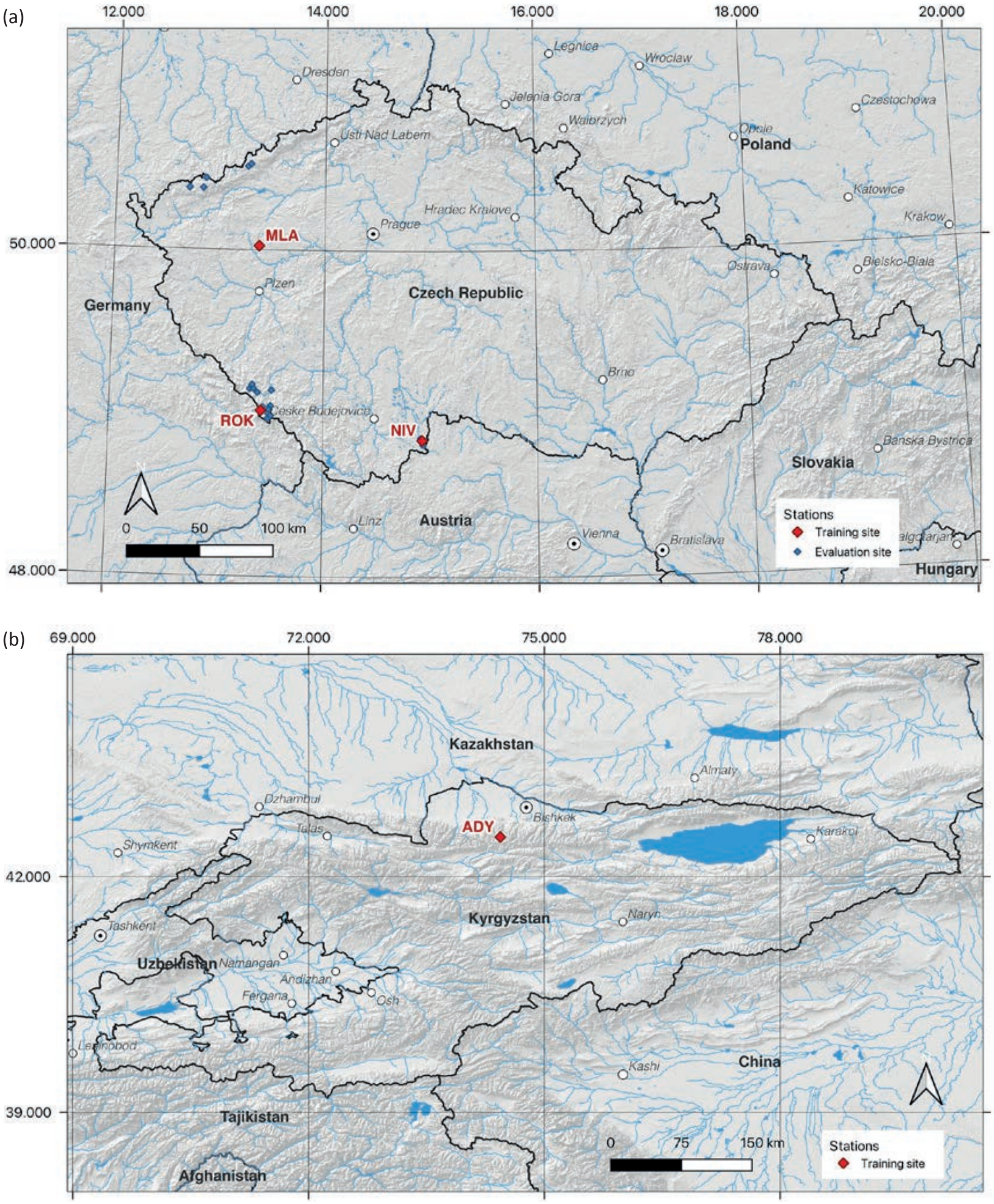

Fig. 2 Location of study sites. (a) Stations in the Czech Republic and (b) Kyrgyzstan.

Šumava, Krušné hory, and Krkonoše mountain ranges as well as in rural midland and lowland basins of Czechia.

The primary test sites encompass substantial variability in natural conditions and hydrological dynamics (Tab. 1). The Adygine catchment (ADY), situated in the Tien Shan Mountains, Kyrgyzstan, represents a high-alpine setting with a gauging station at the outlet of Lake Adygine (Fig. 3a). Its glacial regime produces pronounced daily flow fluctuations. The Rokytka catchment (ROK), located in the Šumava Mountains, Czechia, is characterized by a rain-snow runoff regime with peak flows in spring. The gauging station is positioned downstream of a tunnel through the dam of a former reservoir (Fig. 3b). The Lužnice River site (LUZ), in the Třeboň basin lowlands of Czechia, differs from the others by scale: the gauging station is situated on the middle river reach within a wide floodplain, with a deep channel and very low velocities even at high flows (Fig. 3c). The Mladotický Creek site (MLA), in the hilly agricultural landscape of Western Bohemia, Czechia, is characterized by a rain-snow runoff regime and an intensively farmed basin. The station is installed on a stable concrete bridge (Fig. 3d).

Stage-discharge rating curves at these sites were constructed from direct hydrometric measurements. Data were collected using a SonTek RiverSurveyor ADCP, a SonTek FlowTracker velocimeter, and Ott C2 and C31 propeller meters, following ISO 748 guidelines for velocity measurements. The resulting discharge data have an estimated uncertainty of 2–5%. All gauging stations are located at stable river sections, with repeated measurements confirming rating curve stability over time.

Between 8 and 32 discharge measurements were available per station (Tab. 1). The relatively small sample sizes reflect typical conditions in experimental catchments, while also highlighting the challenges of reliably reconstructing rating curves from limited datasets.

2.4 Hybrid model implementation

Initial testing of baseline regression models indicated that the log-log formulation provided the most suitable foundation for the hybrid approach. Outliers were identified using the interquartile range method and excluded prior to model fitting. Eight machine learning algorithms were evaluated for residual modeling, of which Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP) demonstrated the best performance. Both were trained on the residuals between observed discharges and baseline predictions, thereby capturing systematic deviations from the theoretical relationship.

The SVR implementation employed a radial basis function kernel with regularization parameter C = 10, gamma set to "scale" for automatic variance-based scaling, and epsilon = 0.5 to control tolerance around the support vectors. These settings were chosen to balance generalization and curve smoothness, a requirement for hydraulic applications. The MLP architecture comprised two hidden layers with 50 and 25 neurons, respectively, and used hyperbolic tangent activation functions to produce smooth, bounded outputs. Optimization was performed with the Limited-memory BFGS solver, supported by L2 regularization ($\alpha = 0.01$) to reduce overfitting. The

Fig. 3 Variability of physical properties of gauging stations. (a) Outlet from the Adygine lake.Water drains away via a stable bedrock ridge (riegel), (b) Gauging station of the Rokytka River at the time of maximum water level during the flood situation on December 1, 2015, (c) Lužnice gauging station site at the time of maximum measured water level, (d) Gauging station of the Mladotický Creek during period of increased water level.

Tab. 1 Basic characteristics of stations in experimental catchments used in the study. Data: Charles University.

Station code	Stream	Station name / location	Elevation [m a.s.l.]	Catchment area [sq. km]	Station type*	Number of observations
ADY	Adygine	Adygine	3400	2.8	T	8
ANT	Antýgl	Antýgl	930	1.6	E	9
BRE	Březnický p.	Březník	1140	3.4	E	14
BYS	Bystřice	Abertamy	855	11.1	E	15
CER	Černý	Nová Hůrka	910	1.5	E	13
CIK	Cikánský p.	*catchment outlet*	1055	2.2	E	13
CHH	Chomutovka	Hora	806	8.7	E	17
CHT	Chomutovka	Tišina	650	21.9	E	11
JAV	Javoří p.	*catchment outlet*	1035	14.2	E	20
LOS	Losenice	Rejštejn	570	53.9	E	13
MLA	Mladotický p.	Přehořov	415	34.2	T	10
MOD	Modravský p.	Modrava	991	42.1	E	14
NIV	Lužnice	Niva	450	935.0	T	15
POP	Lužnice	Popelnice	720	1.8	E	8
PTA	Ptačí p.	Ptačí nádrž	1130	5.5	E	30
RKL	Roklanský p.	Modrava	990	47.6	E	32
ROK	Rokytka	Rokytecká nádrž	1090	3.9	T	15
SCH	Lužnice	Suchdol	454	955.0	E	15
SLA	Slatinný p.	*catchment outlet*	850	27.7	E	9
SLP	Slatinný p.	Nové Hamry	755	17.8	E	11
VES	Lužnice	Nová Ves	475	917.0	E	17
ZLA	Zlatý p.	Zlatý kopec	770	5.9	E	12

* Station types: T – Primary test site, E – Independent evaluation site.

maximum number of iterations was set to 1000, and a fixed random state (42) ensured reproducibility.

Model performance was evaluated using multiple criteria, including Mean Squared Error (MSE), Nash-Sutcliffe Efficiency (NSE), coefficient of determination (R^2), and Index of Agreement (IoA), computed for both the baseline and hybrid models.

The hybrid model implementation was developed in Python, relying primarily on the Scikit-learn library for ML modeling. Statistical procedures and baseline regression fitting were carried out with SciPy, including non-linear least squares optimization and continuous curve generation through interpolation routines. Additional preprocessing and smoothing were supported by the scipy.ndimage and scipy.signal modules. Performance evaluation was conducted with HydroEval, while visualization was performed using Matplotlib and Seaborn.

3. Results

3.1 Baseline regression models and ML algorithms

3.1.1 Baseline regression models

Based on theoretical assumptions, three regression models were compared to test the suitability for rating curve determination: log-log, polynomial, and power law (Fig. 4).

All three models exhibited a very high level of fit, with coefficients of determination (R^2) exceeding 0.995. However, despite these strong statistical metrics, some models produced physically implausible behavior. In particular, the polynomial regression curve displayed unrealistic bending in the low-flow region (Fig. 4c), which may yield ambiguous discharge estimates. This highlighted a key limitation: performance metrics alone may mask physical inconsistencies and should therefore not serve as the sole criterion for model selection (McMillan and Westerberg 2015).

The log-log model demonstrated greater robustness, largely due to the logarithmic transformation of both variables. This transformation stabilizes the regression across the full range of observed flows, enhancing reliability even in small, irregularly shaped mountain streams where low-flow conditions are often subject to high measurement uncertainty due to channel morphology. Nevertheless, log-log models also have limitations in steep or morphologically complex channels, where step-pool sequences, backwater effects, or local controls may disrupt the theoretical monotonic relationship (Herschy 2019; Kiang et al. 2018).

3.1.2 Machine learning models

The ML algorithms tested for potential integration into the PIML model revealed two key findings. First, most models, including Artificial Neural Network (ANN), Multi-Layer Perceptron (MLP), Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting (GB), K-Nearest Neighbors (KNN), Gaussian Process (GP), and XGBoost (XGB), achieved high values of statistical performance metrics. However, as the ML models are trained on a limited set of observations, most of them resulted in physically unrealistic rating curves. The high fit achieved by the ML models, with most algorithms reaching R^2 values above 0.96 and low RMSE values, was largely the result of mechanistic fitting to the data rather than capturing the underlying principles governing the stage-discharge relationship. This led to overfitting and unrealistic model behavior, including sharp discontinuities, non-monotonic segments, and excessive sensitivity to sparse data points, all of which are inconsistent with physically plausible rating curves (Mosavi et al. 2018).

This behavior is apparent in the example from the ROK profile, which contains observations spanning the full range of stage values, with a high density of measurements in the mid-range, a distribution typical for experimental basins (Fig. 5). GB, XGB, and GP yielded near-perfect metrics; however, the exceptionally high fits with R^2 values approaching 1.0 are more indicative of overfitting than of true predictive skill (Nearing et al. 2021; Shortridge et al. 2016). RF (R^2 = 0.940) produced the characteristic stepwise predictions typical of ensemble tree-based approaches. KNN, with a similar stepwise fit, was the weakest performer (R^2 = 0.598), failing to reproduce even the basic functional form of the stage-discharge relationship. ANN achieved a balanced performance (R^2 = 0.989) with a realistic curve shape. Based on the combination of performance metrics and smooth, physically relevant curves, MLP (R^2 = 0.997) and SVR (R^2 = 0.995) were among the best-performing models. Importantly, the smooth fits achieved by ANN, MLP, and SVR despite the low number of observations were

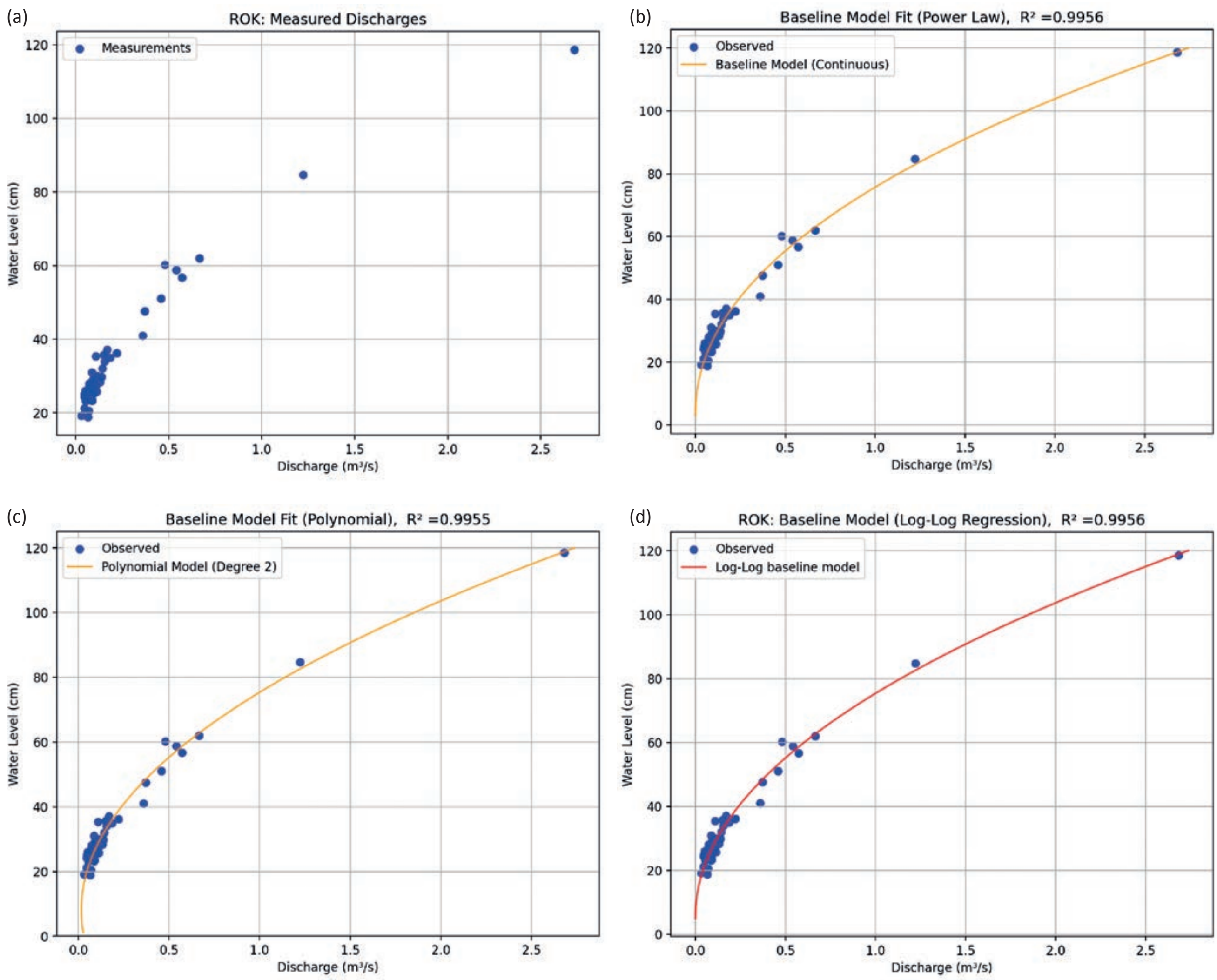

Fig. 4 Stage-discharge reconstruction using conventional regression models for ROK. a) Measured discharges, b) Power Law fit, c) Polynomial fit, d) Log-log fit.

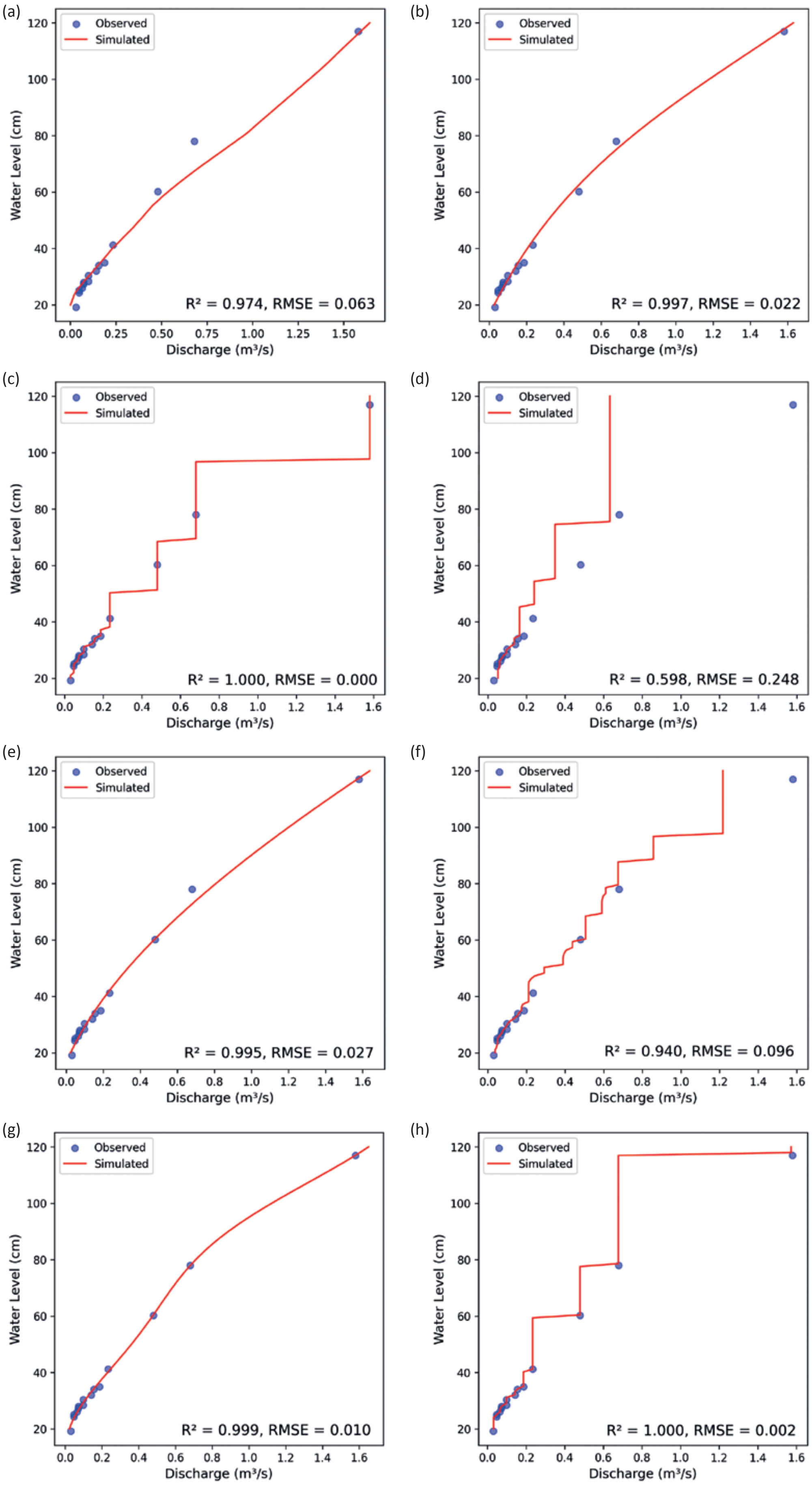

Fig. 5 Stage-discharge reconstruction using ML models for ROK basin. a) Artificial Neural Network, b) Multi-Layer Perceptron, c) Gradient Boosting, d) KNN, e) Support Vector Regression, f) Random Forest, g) Gaussian Process, h) XGBoost.

possible only after careful optimization of model settings and hyperparameters.

MLP and SVR models, representing neural network and kernel-based approaches, were therefore selected for integration into a hybrid modeling framework as representative robust ML methods, combining empirical flexibility with physical plausibility.

3.2 Hybrid model performance in test catchments

The hybrid models demonstrated close agreement with observed stage-discharge relationships, particularly at lower flows where the conventional log-log transformation often underperforms (Fig. 6). Hybrid MLP model was the most consistent performer, producing rating curves closely aligned with observations and maintaining physical plausibility (Fig. 6). Hybrid SVR model achieved similarly strong results but was slightly less stable in extrapolation ranges. Both substantially improved on baseline regression, particularly at low and high flows.

In the ADY basin (n = 8, Fig. 6a), hybrid models outperformed the baseline at higher discharges, with MLP achieving the closest fit (R^2 = 0.983 vs. baseline R^2 = 0.976). In ROK (n = 14) and MLA (n = 10, Fig. 6b–c), all models performed nearly identically, reflecting hydraulically stable conditions (R^2 > 0.996). The NIV basin (n = 12, Fig. 6d) showed the largest improvement from hybridization: the baseline underestimated medium and high flows (R^2 = 0.897), while both hybrid models markedly reduced this deviation, with MLP performing best (R^2 = 0.996).

Model testing in contrasting environments showed that the hybrid framework improved the ability to capture nonlinearities in the rating curves, particularly in the transition from low to high flow regimes. Both SVR and MLP improved predictive accuracy, with MLP consistently achieving the highest R^2 values. The effect of the hybrid models was minimal in hydraulically simple channels (ROK, MLA), but pronounced in complex environments represented by a highly dynamic high mountain stream (ADY) or a large lowland basin (NIV).

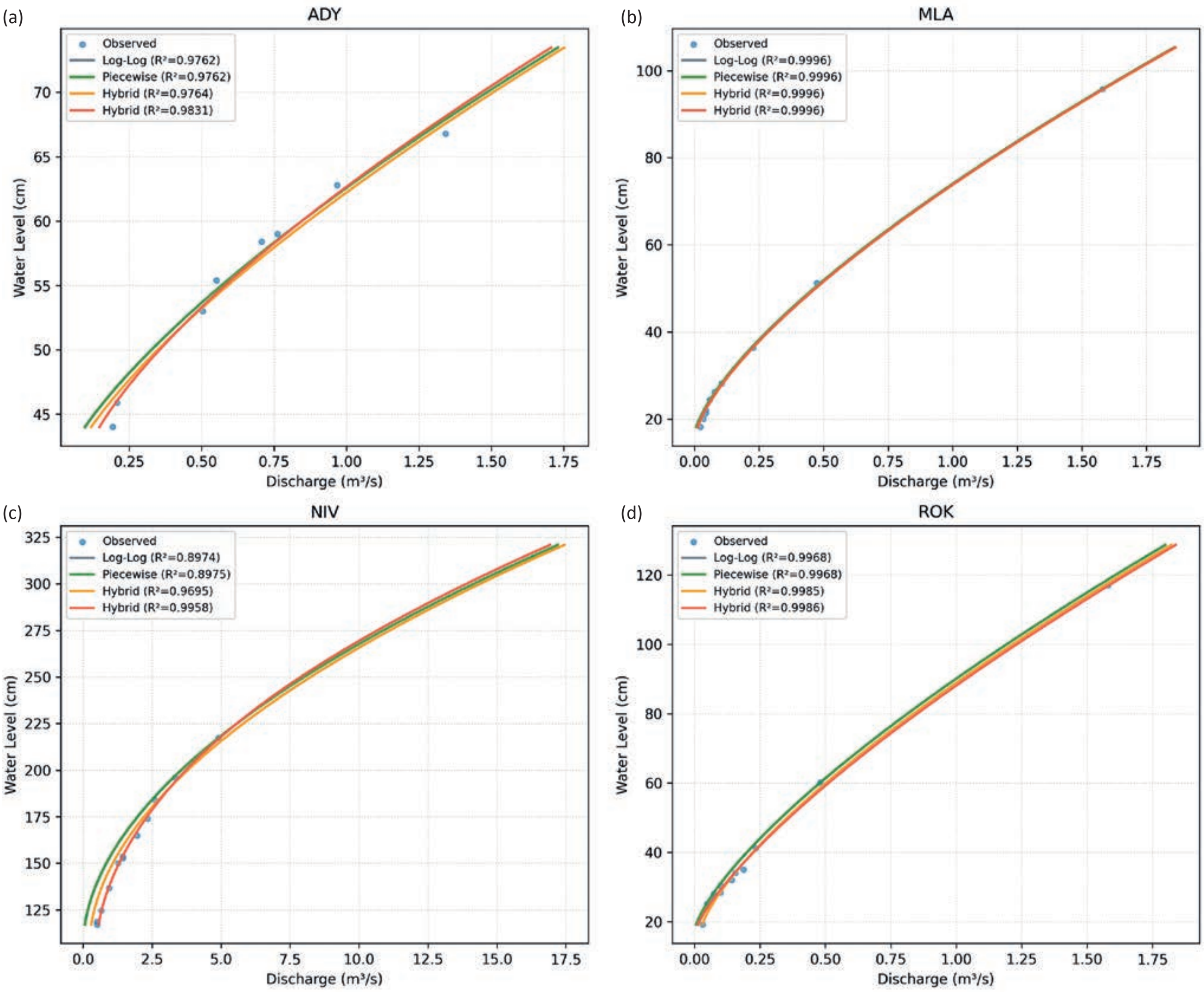

Fig. 6 Observed stage-discharge relationships and model predictions from the baseline log-log regression, hybrid SVR, and hybrid MLP models across the four experimental basins: (a) ADY (n = 8), (b) ROK (n = 14), (c) MLA (n = 10), and (d) NIV (n = 12).

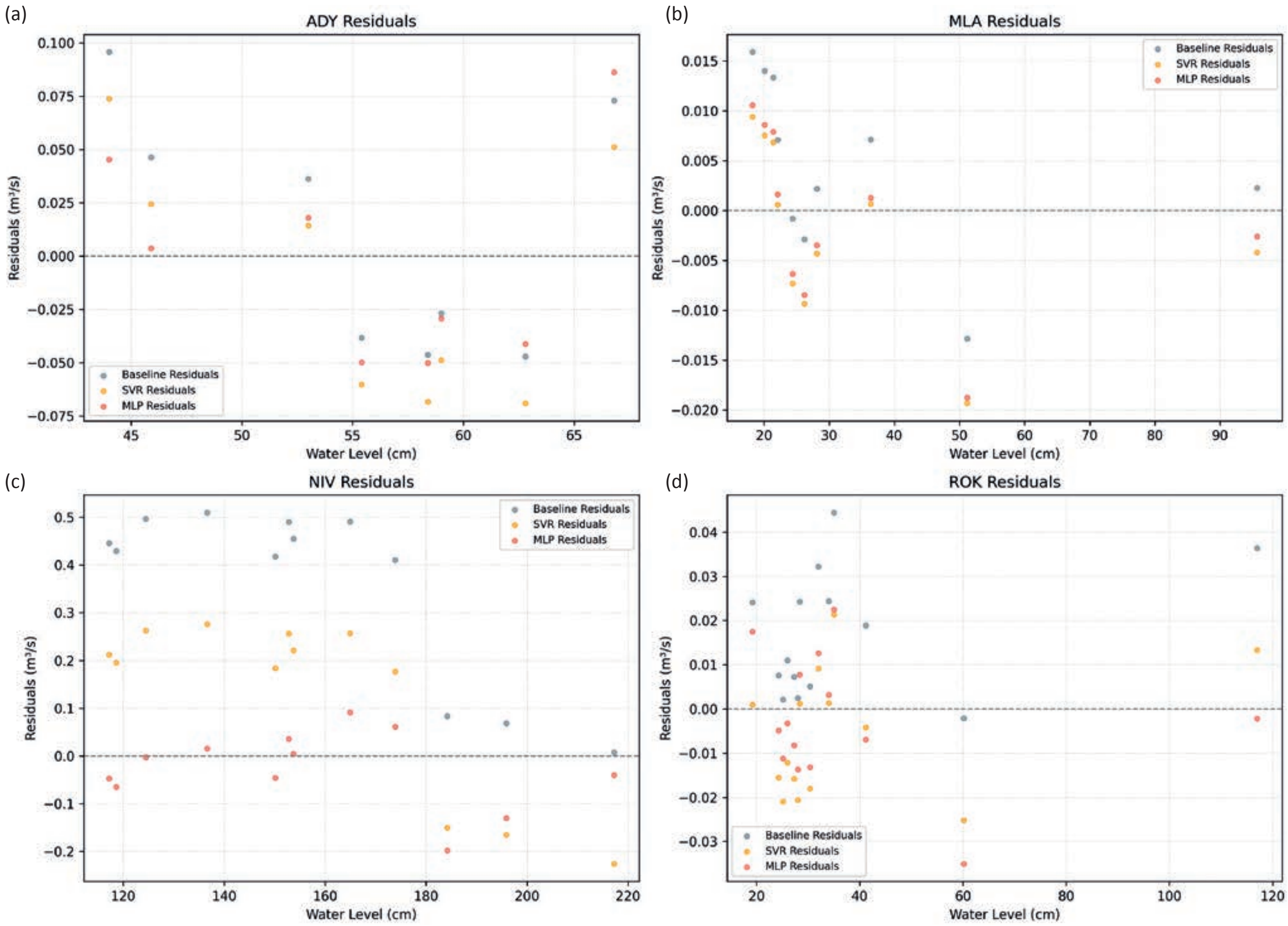

Fig. 7 Residuals from the baseline, SVR and MLP models (a) ROK, (b) JAV, (c) CIK, (d) PTA.

Model performance metrics (R^2, NSE, IoA; Table 2) confirm the effect of hybrid models on the improvements of the fit. MLP achieved the highest predictive skill (mean R^2 = 0.9853, mean IoA = 0.9986), followed closely by SVR (mean R^2 = 0.9825, mean IoA = 0.9968), while the baseline log-log model showed lower accuracy (mean R^2 = 0.9564, mean IoA = 0.9913). Basin-specific patterns are consistent: hybrid models were nearly indistinguishable in MLA and ROK but produced the largest gains in NIV, and moderately improved performance in ADY. These results demonstrate that hybrid models are particularly effective in basins where the stage-discharge relationship departs from simple log-log behavior, while in hydraulically stable basins the differences between models are less pronounced.

Residual plots (Fig. 7) reveal patterns consistent with these findings. In ADY, the baseline model exhibited systematic positive residuals at low flows and negative residuals at high flows, which were mitigated by both hybrid models. ROK and MLA showed compact residual distributions with minor differences among models, while NIV exhibited the greatest residual spread; hybrid models, particularly MLP, substantially reduced residual magnitudes at high flows.

The residual plots (Fig. 7) indicate some degree of heteroscedasticity across all basins, with residual variance appearing to vary with water level. This pattern is most evident in the NIV basin, where residual scatter increases at higher water levels, though the hybrid models help mitigate this effect.

Tab. 2 Model performance metrics by basin and model type, calculated using R2, NSE, and IoA metrics.

Basin	ADY			ROK			MLA			NIV		
Model	log-log	SVR	MLP	log-log	SVR	MLP	log-log	SVR	MLP	log-log	SVR	MLP
R^2	0.9762	0.9764	0.9831	0.9968	0.9985	0.9986	0.9996	0.9996	0.9997	0.8974	0.9695	0.9958
NSE	0.9762	0.9764	0.9831	0.9968	0.9985	0.9986	0.9996	0.9996	0.9997	0.8974	0.9695	0.9958
IoA	0.9943	0.9943	0.9957	0.9992	0.9996	0.9997	0.9999	0.9999	0.9999	0.9777	0.9933	0.9990

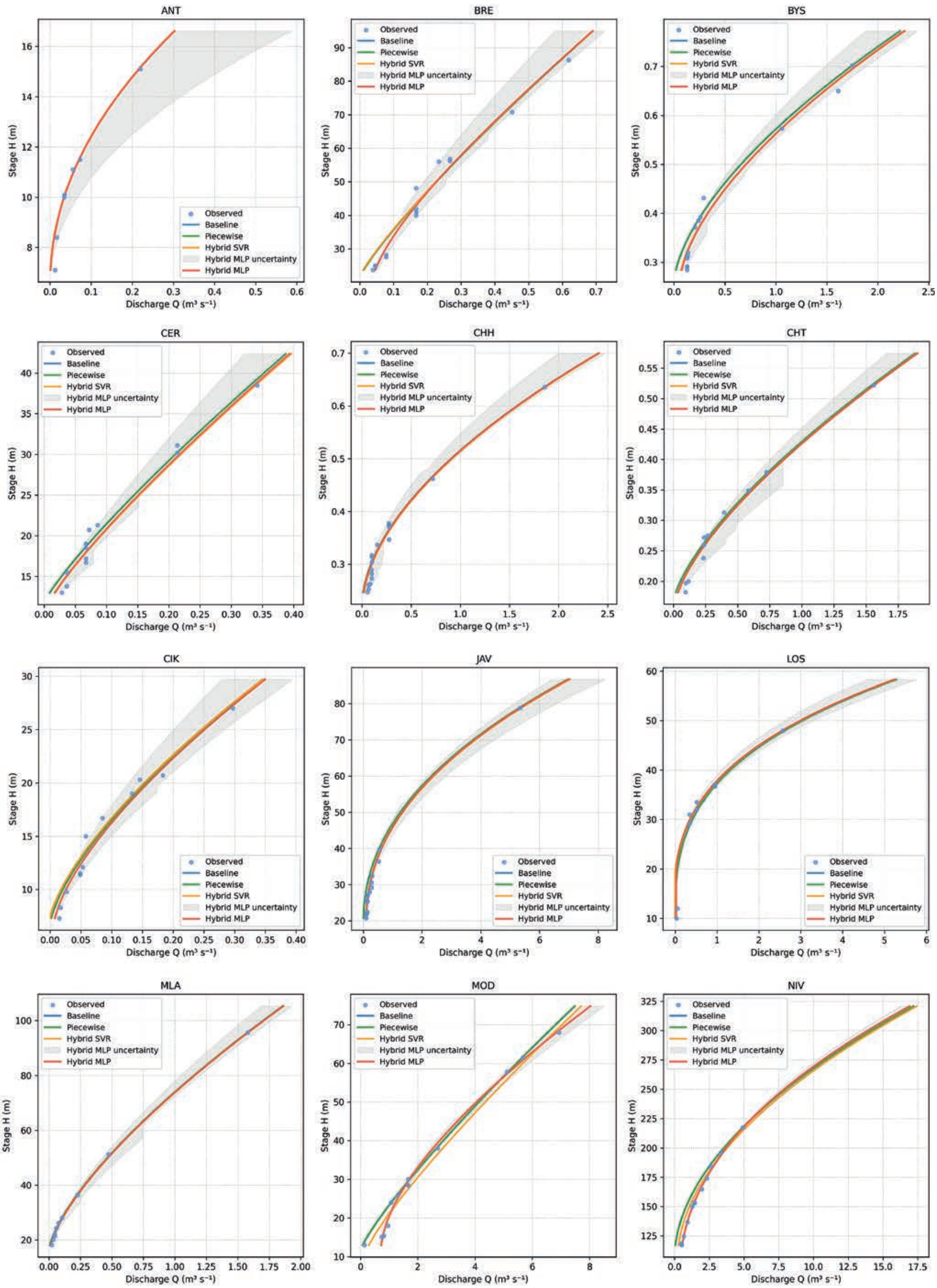

Fig. 8 Observed stage-discharge data and rating curves derived from the log-log baseline, Hybrid SVR, and Hybrid MLP models for 20 independent evaluation catchments.

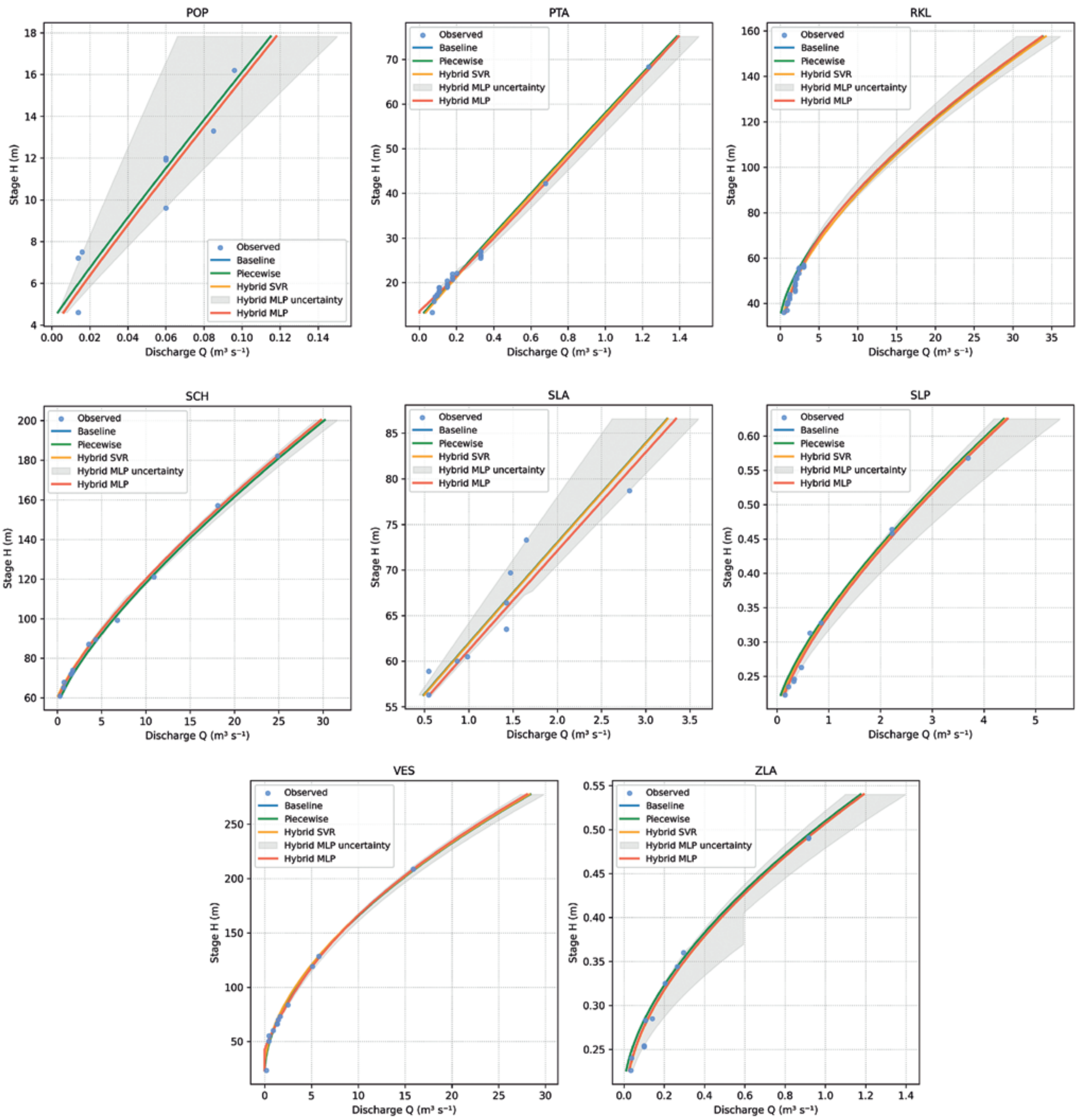

Fig. 8 Observed stage-discharge data and rating curves derived from the log-log baseline, Hybrid SVR, and Hybrid MLP models for 20 independent evaluation catchments.

3.3 Hybrid model performance in independent catchments

The proposed approach was evaluated on 20 independent catchments to assess its applicability and reliability (Fig. 8). Rating curves derived from the log-log baseline, hybrid SVR, and hybrid MLP models were compared with observed stage-discharge measurements.

Across most sites, the Hybrid MLP closely matched observations, capturing both slope and curvature, even under sparse measurement conditions. The Hybrid SVR generally performed well but showed larger deviations at high or low discharges, while the log-log baseline reproduced overall trends but tended to underfit nonlinear relationships in sites with wide discharge ranges. These results demonstrate the capacity of the Hybrid MLP to generalize the stage-discharge relationship across diverse physiographic and hydrological settings.

In addition to model performance, the predictive uncertainty of the hybrid MLP curves was assessed using residual-based quantile envelopes (Fig. 8). The shaded envelopes indicate 95% prediction intervals, typically narrowing in the central stage-discharge

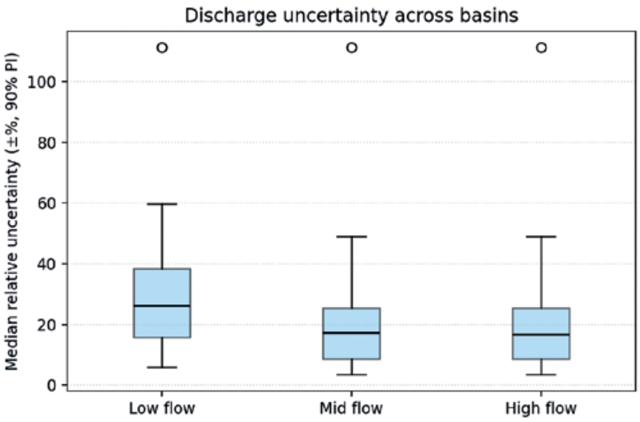

Fig. 9 Distribution of relative discharge uncertainty (±%, 90% prediction interval) across all catchments, grouped by flow terciles.

domain and widening toward the high-flow range. The hybrid MLP reproduces both slope and curvature of the stage-discharge relationship across diverse hydrological conditions, while maintaining narrow uncertainty envelopes for most of the observed range. The widening of prediction intervals toward high stages reflects the limited availability of extreme-flow measurements and the increased extrapolation uncertainty typical of rating-curve applications.

The corresponding summary of relative uncertainty across flow categories (Fig. 9) shows that median uncertainty decreases from low- to high-flow conditions, with median values typically below ±25% for mid- and high-flow regimes, indicating consistent predictive stability of the hybrid MLP model across varying flow conditions. The systematically higher uncertainty at low flows highlights the influence of measurement noise and channel-bed variability, emphasizing the importance of representative low-stage gauging when calibrating machine-learning rating curves.

Distribution of model performance metrics by station (Fig. 10) indicate high goodness of fit for all models, with median R^2 and NSE above 0.99. Nevertheless, the Hybrid MLP showed the narrowest variability and the highest consistency, while Hybrid SVR exhibited wider variation, and the log-log baseline displayed the lowest median performance and broader spread, particularly at low-performance tails (Fig. 10).

All models achieved high median values ($R^2 \approx 0.99$, NSE ≈ 0.98, IoA ≈ 0.99), their variability revealed differences in robustness (Tab. 3). The log-log baseline

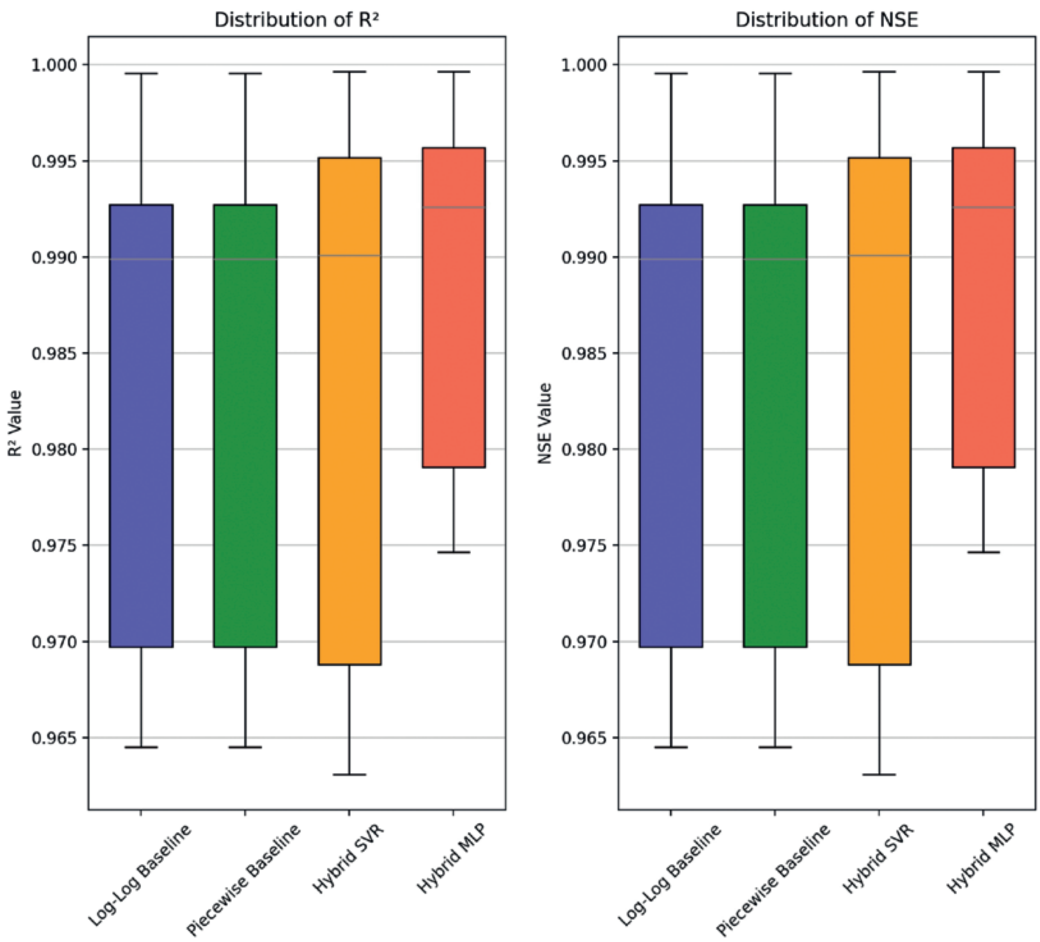

Fig. 10 Distribution of (a) R^2 and (b) NSE model performance values for the log-log and piecewise baseline, Hybrid SVR, and Hybrid MLP models across 20 independent evaluation catchments.

Tab. 3 Performance metrics for the baseline, hybrid MLP, and hybrid SVR models in 20 independent assessment catchments.

Metrics	R2			NSE			IoA			Standard deviation		
	median	max	min	median	max	min	median	max	min	R2	NSE	IoA
Hybrid MLP	0.9926	0.9997	0.8812	0.9926	0.9997	0.8812	0.9982	0.9999	0.9690	0.0343	0.0343	0.009
Model Hybrid SVR	0.9901	0.9996	0.8929	0.9901	0.9996	0.8929	0.9975	0.9999	0.9718	0.0360	0.0360	0.009
log-log Baseline	0.9899	0.9996	0.4669	0.9899	0.9996	0.4669	0.9975	0.9999	0.8929	0.1182	0.1182	0.024
Piecewise Baseline	0.9899	0.9996	0.4673	0.9899	0.9996	0.4673	0.9975	0.9999	0.8930	0.1182	0.1182	0.024

showed the largest spread (std R^2 = 0.1182) and lowest minimum performance (min R^2 = 0.467), indicating underfitting in catchments with nonlinear behavior. This comparison also revealed virtually identical results between the single-segment and three-segment log-log models (mean ΔNSE < 0.001, $\Delta R^2 < 0.0001$), indicating that the additional breakpoints did not improve the fit within the observed range (Tab. 3). Both hybrid models reduced performance variability (std $R^2 \approx 0.034–0.036$), with the MLP achieving the highest median values and narrowest spread, highlighting its generalizability. Maximum metrics were similar across all models, showing accurate fits in well-behaved catchments, while differences emerged primarily in complex or extreme stage-discharge conditions.

Analysis of best-performing models per station (Fig. 11), based on R^2 showed that the Hybrid MLP achieved the highest performance in 14 out of 20 catchments. The Hybrid SVR model achieved the highest performance in four catchments, while the baseline log-log and piecewise models achieved the highest performance in one catchment each.

4. Discussion

4.1 Effect of sample size on model performance

Our results show that hybrid PIML models perform reliably even with sparse calibration data, challenging the assumption that large datasets are essential. The Hybrid MLP, in particular, consistently outperformed both SVR and the log-log baseline across most stations. This highlights the framework's potential for experimental catchments where only 8–15 discharge measurements are typically available.

Specifically, hybrid MLP performed best at 15 out of 20 stations, while hybrid SVR performed best at 4 stations, and the log-log baseline model performed best at only 1 station.

These results demonstrate that ML-based approaches, particularly the Hybrid MLP, can provide accurate rating curve estimates and high NSE and R^2 values even with sparse calibration data (Fig. 12). The Hybrid MLP showed consistently strong performance regardless of the number of calibration points, indicating robustness under data-scarce conditions.

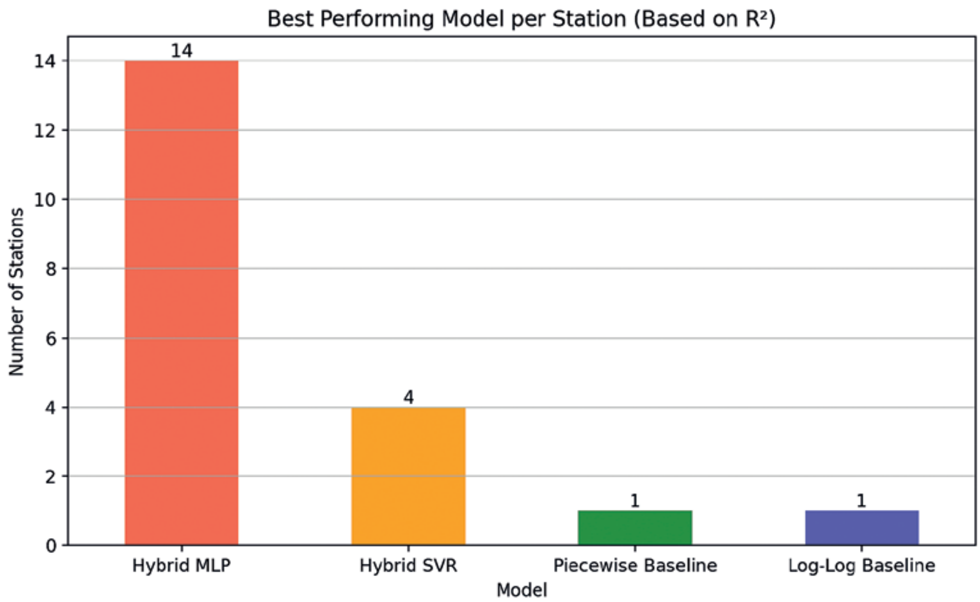

Fig. 11 Number of stations where each model achieved the highest R^2.

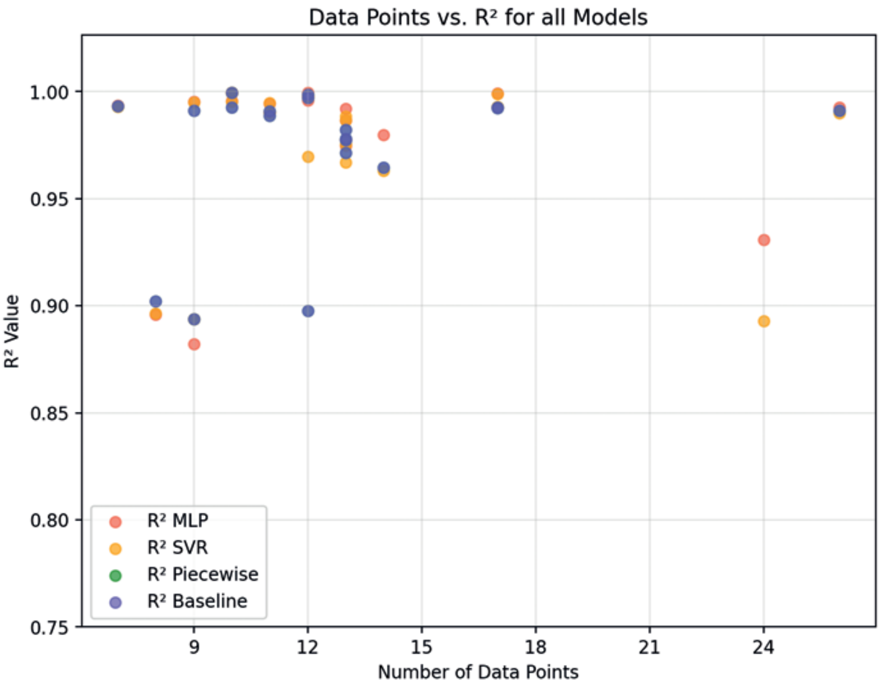

Fig. 12 Model performance metrics according to the number of points, used for rating curve modeling by the model type.

In contrast, the Hybrid SVR and the log-log baseline exhibited greater variability, especially when fewer than 12 observations were available.

While the number of data points influences performance, the accuracy of the underlying hydrometric observations exerts a stronger control. Poor-quality data characterized by noise or systematic errors cannot be compensated for by algorithmic sophistication.

The MLP outperformed SVR despite the latter's theoretical advantages for small datasets. This highlights the importance of empirical evaluation in physics-informed modeling. While SVR's margin-based learning and kernel methods are often recommended for limited data scenarios, the smooth and continuous relationships required by hydrological rating curves, appear better suited to the MLP architecture. This emphasizes that algorithm selection in physics-informed ML should prioritize alignment with the physical characteristics of the system rather than general assumptions about small-data performance.

4.2 Applications of the hybrid model

Hybrid machine learning (ML) models show particular promise for rating curve development in situations where conventional approaches fail or require extensive manual postprocessing. ML component offers robustness and capability to capture complex, nonlinear relationships that traditional methods may struggle to represent (Liang et al. 2023). This can be beneficial in cases where stage-discharge relationships vary due to differences in channel morphology, bed roughness, or flow conditions (Feng et al. 2023).

Our results demonstrate that systematic, physics-informed methods can extract meaningful relationships even from datasets that would be considered inadequate for traditional ML applications. While hybrid approaches cannot compensate for deficiencies in observation quality or coverage, they contribute in several important ways:

(i) Objective methodology: By replacing subjective manual curve fitting with a standardized procedure, hybrid models provide reproducible results that reduce practitioner bias.

(ii) Maximum information extraction: Combining physical constraints with flexible learning algorithms allows greater use of limited data than purely empirical or purely physical models.

(iii) Transferability: The consistent methodology supports comparative hydrology by enabling application across diverse catchments.

These characteristics make hybrid models attractive for several hydrometric contexts. Hybrid models, particularly those using nonlinear algorithms such as multi-layer perceptrons (MLPs) or support vector regression (SVR), can better capture complex stage-discharge relationships (Mosavi et al. 2018). PIML models are likewise useful in rivers with high flow variability, turbulence, or backwater effects, where strong nonlinearities reduce the performance of traditional rating curves (Di Baldassarre and Montanari 2009). Hybrid models can also integrate data from different instruments with varying levels of precision that would otherwise degrade predictive accuracy.

Applications extend to long-term records, where channel morphology may evolve over time. Finally, hybrid models hold potential for reconstructing historical records from sparse or fragmented observations. By combining physically plausible baseline functions with ML corrections, they provide more robust reconstructions than either method alone (Belitz and Stackelberg 2021).

4.3 Limitations and future directions

Despite its advantages, this study remains a proof-of-concept rather than a definitive framework for physics-informed ML in rating curve modeling. A key limitation is the black-box nature of the ML component: unlike physically based or empirical models, its internal decision processes are opaque, which reduces interpretability and complicates transferability in operational contexts (Beven 2019).

Model performance also depends strongly on data quality and quantity. Sparse or noisy observations, typical of experimental catchments, increase the risk of overfitting and poor generalization (Huntingford et al. 2019). The small datasets in this study (8–30 measurements per site) highlight these constraints, limiting statistical inference about algorithm superiority. Nonetheless, within such conditions, the hybrid framework offers a systematic and reproducible alternative to subjective manual fitting while embedding essential physical constraints.To mitigate these challenges, recent research advocates physics-informed ML (Raissi et al. 2019), embedding physical laws such as conservation principles or monotonicity directly into model structures.

Future research should advance physics-informed ML by developing uncertainty quantification methods tailored to small-sample hydrology, testing performance on controlled synthetic datasets, and defining evidence-based guidelines for minimum data requirements. Integrating auxiliary information, such as hydraulic modeling or remote sensing, could further enhance model robustness under data scarcity.

Hybrid rating curve models therefore represent a promising alternative to conventional approaches, particularly in basins with irregular morphology, dynamic flow regimes, heterogeneous measurements, or evolving channels. Their strength lies in combining physical principles with flexible learning, but reliable application will require continued attention to data quality, constraint design, and rigorous evaluation. Beyond the data-driven focus of this study, physically based Bayesian frameworks such as BaRatin (Le Coz et al. 2014; Kiang et al. 2018) demonstrate how hydraulic principles and uncertainty quantification can be jointly incorporated in rating-curve modeling. Implementing such models requires detailed cross-sectional and control-type information that was beyond the scope of our dataset, but their concepts are highly complementary to the machine-learning approaches tested here. Future work will therefore explore hybrid Bayesian-ML strategies that combine the interpretability and uncertainty propagation in tools such as BaRatin with the flexibility and generalization capacity of machine-learning models.

5. Conclusions

This study demonstrates that physics-informed machine learning (PIML) provides a systematic framework for developing reliable stage-discharge relationships under the data constraints typical of experimental catchments. By combining a physically based log-log baseline with machine learning residual corrections, the approach ensures monotonicity, non-negativity, and continuity while enhancing predictive skill. While acknowledging the constraints imposed by small sample sizes, the results show that the hybrid model consistently improves on conventional regression approaches, particularly in basins with irregular morphology or high measurement uncertainty.

Among the tested algorithms, the Hybrid MLP proved the most robust and generalizable across diverse hydrological settings, performing best at the majority of sites. The Hybrid SVR also delivered strong results and showed comparative advantages in highly irregular environments, suggesting that algorithm choice may be guided by site-specific conditions. In contrast, the traditional log-log regression performed best only in hydraulically simple profiles, highlighting the benefits of hybridization where nonlinearities dominate.

The strength of the proposed framework lies not in maximizing statistical fit with limited samples, but in providing a reproducible and physically constrained methodology that reduces the subjectivity of manual curve fitting. This makes it particularly relevant for experimental hydrology, where discharge estimation is often based on sparse, heterogeneous, and uncertain observations.

The MLP model demonstrated best overall performance (median R^2 and NSE > 0.98), achieving best results at 15 out of 20 evaluation sites, while the SVR model performed best at 4 sites, and the log-log baseline at only one site. Statistical analysis confirms MLP‘s consistency, with lower variability in performance metrics and higher median values across all metrics.

Limitations remain, notably the dependence on the quality of available hydrometric data and the limited interpretability of the machine learning components. Nevertheless, the approach establishes a proof-of-concept for integrating physical principles with data-driven learning in rating curve development. Future work should focus on refining uncertainty quantification, testing transferability under changing channel conditions, and integrating auxiliary data sources such as hydraulic modeling or remote sensing.

By bridging physics-based principles and machine learning techniques, this framework offers experimental hydrology a practical tool for objective and transferable rating curve development, extending the scope of reliable discharge estimation in data-scarce and hydraulically complex environments.

Acknowledgments

This research was supported by the Czech Science Foundation project 22-12837S: Hydrological and hydrochemical responses of montane peat bogs to climate change, and by the Technology Agency of the Czech Republic project SS02030040.

References

Adarsh, S., John, A. P., Anagha, R. N., Abraham, A., Afiya, M. P., Arathi, K. K., Azeem, A. (2018): Developing stage-discharge relationships using multivariate empirical mode decomposition-based hybrid modeling. Applied Water Science 8: 230, https://doi.org/10.1007/s13201-018-0874-8.

Ali, G., Maghrebi, M. F. (2023): A robust approach for the derivation of rating curves using minimum gauging data. Journal of Hydrology 623: 129609, https://doi.org/10.1016/j.jhydrol.2023.129609.

Belitz, K., Stackelberg, P. E. (2021): Evaluation of six methods for correcting bias in estimates from ensemble tree machine learning regression models. Environmental Modelling & Software 139: 105006, https://doi.org/10.1016/j.envsoft.2021.105006.

Beven, K. (2019): How to make advances in hydrological modelling. Hydrology Research, 50(6), 1481–1494, https://doi.org/10.2166/nh.2019.134.

Bhasme, P., Vagadiya, J., Bhatia, U. (2022): Enhancing predictive skills in physically-consistent way: Physics Informed Machine Learning for hydrological processes. Journal of Hydrology 615: 128618, https://doi.org/10.1016/j.jhydrol.2022.128618.

Braca, G. (2008): Stage-discharge relationships in open channels: practices and problems, https://api.semanticscholar.org/CorpusID:53496646.

Di Baldassarre, G., Montanari, A. (2009): Uncertainty in river discharge observations: a quantitative analysis. Hydrology and Earth System Sciences 13(6), 913–921, https://doi.org/10.5194/hess-13-913-2009.

Dobrovolski, S. G., Yushkov, V. P., Vyruchalkina, T. Y., Sokolova, O. V. (2022): Are There Fundamental Laws in Hydrology? Pure and Applied Geophysics 179, 1475–1484, https://doi.org/10.1007/s00024-022-03003-1.

Esmaeilzadeh, M., Amirzadeh, M. (2024): Replication Study: Enhancing Hydrological Modeling with Physics-Guided Machine Learning. Computer Science. Machine Learning, arXiv:2402.13911v1, https://doi.org/10.48550/arXiv.2402.13911.

Feng, D., Beck, H., Lawson, K., Shen, C. (2023): The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment. Hydrology and Earth System Sciences 27(12), 2357–2373, https://doi.org/10.5194/hess-27-2357-2023.

Herschy, R. W. (2019): Streamflow measurement (3rd ed.). CRC Press, https://doi.org/10.1201/9781482265880.

Hrafnkelsson, B., Sigurdarson, H., Rögnvaldsson, S., Jansson, A. Ö., Vias, R. D., Gardarsson, S. M. (2022): Generalization of the power-law rating curve using hydrodynamic theory and Bayesian hierarchical modeling. Environmetrics 33(2): e2711, https://doi.org/10.1002/env.2711.

Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., Yang, H. (2019): Machine learning and artificial intelligence to aid climate change research and preparedness. Environmental Research Letters 14(12): 124007, https://doi.org/10.1088/1748-9326/ab4e55.

ISO (2020): Hydrometric uncertainty guidance (HUG), ISO 25377:2020.

ISO (2021): Hydrometry – Measurement of liquid flow in open channels – Velocity area methods using point velocity measurements, EN ISO 748:2021.

Kennedy, E. J. (1984): Discharge ratings at gaging stations. Techniques of Water-Resources Investigations 03A10, https://doi.org/10.3133/twri03A10.

Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., Belleville, A., Sevrez, D., Sikorska, A. E., Petersen-Øverleir, A., Reitan, T., Freer, J., Renard, B., Mansanarez, V., Mason, R. (2018): A Comparison of Methods for Streamflow Uncertainty Estimation. Water Resources Research 54(10), 7149–7176, https://doi.org/10.1029/2018WR022708.

Kratzert, F., Klotz, D., Herrnegger, M. (2019a): Toward improved predictions in ungauged basins: Exploiting the power of machine learning. Water Resources 55(12), 11344–11354, https://doi.org/10.1029/2019WR026065.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G. (2019b): Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrology and Earth System Sciences 23(12), 5089–5110, https://doi.org/10.5194/hess-23-5089-2019.

Lane, S. N. (1998): Hydraulic modelling in hydrology and geomorphology: A review of high resolution approaches. Hydrological Processes 12(8), 1131–1150, https://doi.org/10.1002/(SICI)1099-1085(19980630)12:8<1131::AID-HYP611>3.0.CO;2-K.

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., Le Boursicaud, R. (2014): Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. Journal of Hydrology 509, 573–587, https://doi.org/10.1016/j.jhydrol.2013.11.016.

Liang, W., Chen, Y., Fang, G., Kaldybayev, A. (2023): Machine learning method is an alternative for the hydrological model in an alpine catchment in the Tianshan region, Central Asia. Journal of Hydrology, Regional Studies 49: 101492, https://doi.org/10.1016/j.ejrh.2023.101492.

Liu, X., Lu, D., Zhang, A., Liu, Q., Jiang, G. (2022): Data-Driven Machine Learning in Environmental Pollution: Gains and Problems. Environmental Science & Technology 56(4), 2124–2133, https://doi.org/10.1021/acs.est.1c06157.

McMillan, H. K., Westerberg, I. K. (2015): Rating curve estimation under epistemic uncertainty. Hydrological

Processes 29(7), 1873–1882, https://doi.org/10.1002/hyp.10419.
Mosavi, A., Ozturk, P., Chau, K.-W. (2018): Flood Prediction Using Machine Learning Models: Literature Review. Water 10(11): 1536, https://doi.org/10.3390/w10111536.
Nearing, G. S., Kratzert, F., Sampson, A. K. (2021): What role does hydrological science play in the age of machine learning? Water Resources Research 57(3): e2020WR028091, https://doi.org/10.1029/2020WR028091.
Poulinakis, K., Drikakis, D., Kokkinakis, I. W., Spottswood, S. M. (2023): Machine-Learning Methods on Noisy and Sparse Data. Mathematics 11(1): 236, https://doi.org/10.3390/math11010236.
Raissi, M., Perdikaris, P., Karniadakis, G. E. (2019): Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics 378, 686–707, https://doi.org/10.1016/j.jcp.2018.10.045.
Roelofs, R., Shankar, V., Recht, B., Fridovich-Keil, S., Hardt, M., Miller, J., Schmidt, L. (2019): A Meta-Analysis of Overfitting in Machine Learning. Advances in Neural Information Processing Systems 32. Available online: https://dl.acm.org/doi/pdf/10.5555/3454287.3455110 (accessed on 10 May 2025).
Shortridge, J. E., Guikema, S. D., Zaitchik, B. F. (2016): Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrology and Earth System Sciences 20(7), 2611–2628, https://doi.org/10.5194/hess-20-2611-2016.
WMO (2010): Manual on stream gauging. Volume I – Fieldwork. WMO-No. 1044, World Meteorological Organization. Available online: https://library.wmo.int/records/item/35848-manual-on-stream-gauging-vol-i-fieldwork (accessed on 10 May 2025).
WMO (2020): Guide to Hydrological Practices I. Hydrology – From Measurement to Hydrological Information. WMO-No. 168, World Meteorological Organization. Available online: https://library.wmo.int/records/item/35804-guide-to-hydrological-practices-volume-i?offset=6 (accessed on 10 May 2025).
Xu, T., Liang, F. (2021): Machine learning for hydrologic sciences: An introductory overview. WIREs Water 8(5), https://doi.org/10.1002/wat2.1533.
Xu, W., Chen, J., Corzo, G., Xu, C.-Y., Zhang, X. J., Xiong, L., Liu, D., Xia, J. (2024): Coupling deep learning and physically based hydrological models for monthly streamflow predictions. Water Resources Research 60(2): e2023WR035618, https://doi.org/10.1029/2023wr035618.
Ying, X. (2019): An Overview of Overfitting and its Solutions. Journal of Physics, Conference Series 1168(2): 022022, https://doi.org/10.1088/1742-6596/1168/2/022022.