ACTA UNIVERSITATIS CAROLINAE

AUC **GEOGRAPHICA**

53
2/2018

CHARLES UNIVERSITY • KAROLINUM PRESS • 2018

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

The renaming of streets in post-revolutionary Ukraine: regional strategies to construct a new national identity

Oleksiy Gnatiuk

Taras Shevchenko National University of Kyiv, Ukraine
* Corresponding author: alexgnat22@ukr.net

ABSTRACT

After the 2014 revolution, a massive renaming of toponyms related to the communist ideology took place in Ukraine. The results of this renaming help understand the essentiality of Ukrainian delayed post-socialist and post-colonial transition and, in particular, national and regional identities that Ukrainians are going to build, and ideology that local and national authorities are going to impose. This study covers the 36 largest cities in Ukraine with a population of more than 100,000 and focuses on the new street names that appeared as the result of de-communization. It was found that in south-eastern Ukraine, the renaming strategy was targeted to avoid the commemorative names, especially those related to military-political events and personalities, as well as to depoliticize urban toponyms, by using non-commemorative categories of street names, like topographical or poetic ones, or restoring the historical toponyms. On the contrary, new toponyms in western and central Ukraine reflect the legacy of the national liberation movements of the 20th century. However, urban toponyms in the most eastern regions, including Donbas, continue to retain close links with the Soviet period. The memory of the Cossack era and Ukrainian Peoples Republic seems to be the well-perceived and common-shared strata of national identity. At the same time, the Ukrainian Insurgent Army and the Soviet legacy remain topics dividing the Ukrainian society. The street renaming process in post-revolutionary Ukraine shares some common characteristics with other post-socialist European countries, such as the appeal to the pre-socialist period in order to reinforce the national identity, replacement of political and military place-names with those related to local and national culture and heritage, and the increased importance of local and regional toponyms.

KEYWORDS

urban toponymy; renaming of streets; identity; ideology; Ukraine

Received: 29 August 2017
Accepted: 20 March 2018
Published online: 18 June 2018

Gnatiuk, O. (2018): The renaming of streets in post-revolutionary Ukraine: regional strategies to construct a new national identity. AUC Geographica 53(2), 119–136
https://doi.org/10.14712/23361980.2018.13

1. Introduction

After the Revolution of Dignity, Ukraine officially set a political course for decommunization. This process includes renaming of urban toponymy, related to the communist regime. In fact, significant share of urban place-names were renamed even earlier, since Ukrainian independence in 1991. However, this process was non-systemic and quite uneven in regional dimension: while in Western Ukraine nearly all Soviet-originated place-names have been already renamed, in Eastern and Southern Ukraine the Soviet toponymy remained practically intact. Decision about this Soviet legacy was totally in hands of the local authorities. In contrast, todays renaming of urban place-names is non-discretionary and represents a part of the state decommunization policy. Therefore, all cities are legally obliged to go through decommunization. Clearly identified list of urban toponymy must perish from the maps.

Therefore, local communities must pass through decommunization even if they wish to avoid this process. However, the official legislation provides no strict instructions about new place-names which should appear instead. In this respect, local communities have a large range of discretion. Local authorities can choose different renaming strategies, based on local cultural traditions together with electoral and political situation. The ideology of erased toponyms may be replaced by a new national ideology, but the content of this ideology may be quite different depending on the region: emphasis can be placed on different values, different historical periods, various distinguished figures. There is also a scenario in which the new national ideology does not manifest itself; instead, preservation of the Soviet ideology in a hidden form or addressing to a purely regional or local identity may take place.

Therefore, the results of street renaming in the framework of decommunization give abundant food for reflection on a series of questions:

1. What different strategies do local authorities in Ukraine use while implementing decommunization in view of high level of cultural interregional diversity?
2. What are more or less hidden geopolitical and cultural fault lines existing in contemporary post-Revolution and war-affected Ukraine?
3. In what way local authorities in certain Ukrainian regions try to construct new national and/or regional identity (including how they cope with discrepancies between already existing local identity and official state ideology)?
4. What elements (strata) of national identity integrate or disintegrate the country?

The answers to these questions are important to understand the essentiality of Ukrainian delayed post-socialist (and post-colonial) transition, in particular, what state Ukrainians are going to build, and on what grounds? The findings may clarify the real state of interregional differences, indicate well-integrated and badly-integrated areas, and point at inclusive values for contemporary nation-building as well as at disjunctive ones that must be avoided now in interregional dialog. Certainly, is it very luckily to have now a wide array of renamed place-names exactly in Eastern and Southern Ukraine, where national identity may (or may not) undergo crucial changes.

2. Scientific background

Toponyms are not merely abstract names in the spatial structure of cities, but also represent the construct of social and power relations, through which the identity of the city and society is being formed (Berg and Voulteeenaho 2009). Reuben Rose-Redwood, Derek Alderman and Maoz Azaryahu argue that the study of place naming has recently undergone a critical reformulation as scholars have moved beyond the traditional focus on etymology and taxonomy by examining the politics of place-naming practices (Rose-Redwood et al. 2010). On the one hand, names are changing because society is doing so; however, place names are often transformed to propose or bring about evolution in society (Guyot and Seethal 2007). Place names, in line with this approach, are considered dually as: 1. a mirror of existing, actual identities, accepted by spatially-localized community; 2. an instrument used by political regimes to transform or erase the existing identity and create new one instead. The official factor must be stressed, since identities articulated by political elites are not necessarily the same as in the general population; however, with time, new names may become generally accepted: the political elite decides by whom or by which institution or social events streets will be named, but there is not always a compromise between the elites and the population of the city when naming certain streets. However, with time, new street names may be accepted by population (Stiperski et al. 2015; Bucher et al. 2013).

Changes of street names constitute an important part of the process of creating new public iconographic landscapes in harmony with the values of the new regimes, so analysing such changes may provide important insights into the ways in which post-communist countries alter the contours of national identities and national pasts (Stiperski et al. 2015). Graham et al. (2000) emphasize that political elites, if they are aware of the symbolic power denoted by space, erase symbols of previous regimes and implant their own ones; new regimes determine new versions of history and new world-views, accentuating persons, events and historical facts that can be useful to their rule, and trying to erase from the population's memory others that might be detrimental to them. The toponymic reworkings imposed by the community's succession of political leaders reflect the goals, tactics, and, indeed, the ethos of each new national government (Gonzalez Faraco and Murphy 1997). Kadmon (2004) even uses the term of "toponymic warfare" to describe the extent to which marginalized nationalities and linguistic cultures within countries have appropriated and rewritten place names on maps as part of their campaigns of resistance.

According to Azaryahu (2011), (re)naming street names is well-tried and effective way of introducing and disseminating authorized version of history that political elite tries to implement into ordinary setting of "common" people's everyday life, with the main intention of creating new collective memory. Both the urban toponymy and the process of its renaming reflect certain needs of political elites and their ability of manipulation with the symbols of common cultural and historical heritage. Commemorative street names are instrumental in the symbolic construction of national identity, mainly in terms of historical heritage (Azaryahu and Kook 2002), they represent an important vehicle for bringing the past into the present (Alderman 2003). Gonzalez Faraco and Murphy (1997) claim that toponymic reworkings reflect the goals, tactics, and the ethos of each new national government, and reveal distinctive mix of local, regional, and national orientations.

Almost every regime change in XIX–XX centuries was marked by the following widespread renaming of urban toponymy. The most prominent examples enlightened in the literature are: nascent and descent of right authoritarian political regimes in Europe (Italy, Germany, Spain, Portugal etc.); fall of the socialist system in the countries of the Central and Eastern Europe; decolonization of Asian and African states; inner ethnical conflicts all over the world (Ireland, Bosnia and Herzegovina, Israel etc.). Consequentially, three main directions of research may be distinguished: 1. studies that rely on the interests regarding the fall of the great colonial system; 2. studies that deal with the ethnical disputes in ethnically divided areas, which also emphasize the importance of (re)naming streets as the mechanism of symbolic demarcation of ethnic groups, as well as the significance of hodonyms in building national and local identities; 3. studies that address the issues about changes of political regimes in the Middle European, Eastern European and Southeastern European countries (Stanić et al. 2009).

The Ukrainian case belongs to the third group of studies, however, also with some of post-colonial (in fact, Ukraine, like other Soviet republics, was a Russian colony) and ethnic (despite the annexation of Crimea and Donbas conflict, Ukraine remains a multiethnic state) discourses. In this view, the results of toponymy studies from the other post-Socialist and, especially, post-Soviet countries are of large importance. In general, different authors seek to demonstrate how, through the control of state and local government, the political elite can promote its ideology in relation to the nomenclature of the street network.

A book of Sänger (1996) represents thorough analysis of the official memory policy in relation to street names in former GDR. Light (2002) studied the changes in nomenclature of the street network of Bucharest from 1948 to 1965 and disclosed a wide range of symbols and characters incorporated to the urban space in order to legitimise and institutionalise the ideology of revolutionary socialism. His further research of the mass renaming of streets in Bucharest in the 1990s (Light 2004) revealed the occurrence of hodonyms named after notable persons, events and figures that evoked the memories on the time before socialism. Thus, these renamings were expressions/manifestations of a reconfiguration or restructuring of space and history, constituting a vital and integral element of the post-communist transformation. Similar processes were noticed in the Russian capital city of Moscow (Gill 2005).

The studies of urban place names revealed the general idea of the necessity of awakening national consciousness and identity among the citizens of new-born independent state (Crljenko 2012). Comparative analysis of place names in Croatian cities indicated that regional identity was more strongly expressed in Istria than in the region of Kvarner; analysis of toponyms in the Croatian town of Senj proved that its historical-geographic development could be to a great degree reconstructed from these toponyms (Crljenko 2006). It is worth to emphasize a good state of toponymy studies in Croatia: the researchers tried to investigate the main ideas, followed by the decision makers in the processes of renamings, in numerous cities of Zagreb (Rihtman-Augustin 2000; Marjanović 2007; Stanić et al. 2008), Vukovar (Crlenjak 2005), Zadar (Begonja 2006), Pula (Bertoša 2008; Krizmanić 2008), Rijeka (Malešić 2011).

Studying changes in the toponymy of the capital of Bosnia and Herzegovina, Sarajevo, Robinson et al. (2001) found out that names associated with Serbia, Croatia or respective nations, even those from the Bosnian-Herzegovinian context, were replaced mainly by names associated with Sarajevo and the Bosniak Muslim population and culture. They make a conclusion that this was an attempt to create a Bosnian-Herzegovina's identity based primarily on the Bosniak element, as opposed to the identity in the previous Yugoslav period, which was founded on a mix of Bosniak, Serbian and Croatian elements.

Balode (2012) investigated street-names coined from toponyms in Riga, including the motivation of renaming and Latvians' efforts to conserve their national place-names in urban toponymy. Comparative studies of urban toponymy at both national and international levels were carried out in Slovakia (Bucher et al. 2013) and by the international scientific team (Stiperski et al. 2011). The first study aims to determine the intensity of local, regional and national identity in Slovakia. The authors tried to verify the statement that the names of streets, squares or parks in the historical centre of cities serve as indicators of official views and ideological impressions on the political, social and historical events, in the context of Central Europe. The second one investigates urban place names in international scale and provides interesting conclusions on the more and less typical structure of toponyms in terms of their semantics.

David (2013), studying post-Socialist renaming in Czechia, provides arguments supporting the preservation of the original urbanonyms in Soviet-time neighbourhoods, despite their ideologically motivated origins, and focuses on the role of toponyms as a part of cultural heritage and elements contributing to local identity. In the other paper (2012), the same author describes the structure of commemorative street names and stability of urbanonyms related to different periods of Czech history. The role of urban toponymy in development of local identity in post-Socialist period is described, based on rich empirical material, also in contributions by Odaloš and Majtán (1996), David (2011), Knappová (2013), David and Mácha (2014). Renaming process was not painless: in some countries, it came with heated discussions (Palonen 2008).

Also, there are many helpful studies on toponymy outside the region of post-Socialist Eastern Europe, including scientific papers representing toponymic evidence from Singapore (Yeoh 1992, 1996), Spain (Gonzalez Faraco and Murphy 1997), Israel (Azaryahu and Kook 2002), Unites States (Alderman 2003), South Africa (Guyot and Seethal 2007) etc.

As Azaryahu (2011) suggests, de-colonization of commemorative street names following independence may conform to three main strategies: 1. to erase all "colonial" toponymy to signify a complete break from the colonial past; 2. to leave colonial commemorations in their place; 3. to implement a selective de-commemoration of the colonial past. Although the discourse of Azaryahu refers to the post-colonial renaming, these strategies are applicable to post-Soviet renaming as well. Legacy of different periods may be emphasized to pursue some political goals, including attempts to unite a country in ideology and identity. E.g. renaming the East German past in Potsdam did not entail a return to the Prussian past: notably Prussian names, predominant before being erased by the communist administration in December 1945, were not restored to their former place (Azaryahu 2012). After the death of Francisko Franco in 1975, the socialist government in the town of Almonte, Spain, fashioned a crafty symbolic compromise aimed at ending the onomastic cycle of victors and vanquished (Gonzalez Faraco and Murphy 1997).

3. Ukrainian context for renaming the urban toponymy

The history of Ukraine is characterized by long period without own independent national state as well as the division of the country between neighbouring states (Polish-Lithuanian Commonwealth, the Ottoman Empire, the Russian Empire, the Austro-Hungarian Empire, the Second Polish Republic, Czechoslovakia, Romania, USSR), each of which tried to impose local people own state ideology, including by the renaming of urban toponymy. This process has been especially pronounced in the XX century, when the names that did not match communist ideology were massively replaced by the names related to communist leaders and organizations. Even nowadays, as Diesen and Keane (2017) emphasise, both the West and Russia seek to encourage a particular historical narrative in Ukraine that is compatible with their interests in the region.

At the same time, the complicated history of the country has led to the formation of differences in the regional mentality, which lay from the time of ancient history, but eventually formed during the Soviet rule. Passing over the details, it should be noted that inhabitants of Western Ukraine has the most nationalist (sometimes even radical) and anti-Russian, but simultaneously the most pro-European views. The activities of the Ukrainian Nationalists Organization and the Ukrainian Insurgent Army (OUN-UIA), which during the WW II had cooperated with the Nazis to a certain moment, and then opened the struggle both against the Nazis and the Communists, constitute an integral part of local identity. The population of the central part of Ukraine as a whole shares the same ideology, but it has less radical appearance. During the period of Independence of Ukraine the influence of Soviet propaganda has practically faded away, therefore the activity of the Ukrainian Nationalists Organization and the Ukrainian Insurgent Army is perceived here mostly favourably, but without fanatical admiration. Simultaneously, in the south-east, much more industrialized and with significant share of ethnic Russians, pro-Russian sentiments are still widespread even in the face of a bloody military conflict in Donbas: if in the west the personification of evil is Russia as a state, then in the east and in the south it is the ruling regime and personally "Vladimir Putin who will sooner or later go", while the "fraternal Russian people will remain". In addition, there is still widespread pro-Soviet sentiment and nostalgia for the "great state" that built the industrial power of this land.

Regional differences in geopolitical attitudes have affected the character of the renaming of streets in the period since the collapse of the Soviet Union (1991) to the Revolution of Dignity (2014). During this period, the renaming of streets was absolutely voluntary and depended only on the free will of local governments.

In particular, in the cities of Western Ukraine, especially in Galicia, almost all names related to the communist regime were erased during this period. Commemorative names, which had nothing to do with communist ideology, but associated with the Russian culture and science, were removed as well. In the cities of central Ukraine, including Kyiv, the renaming was limited and related to the streets in the central parts of cities and streets that carried the names of the most odious communist regime representatives. In most cities of southern and eastern Ukraine, with some exceptions, the renaming of streets practically did not occur, and the Soviet toponymy was preserved almost entirely.

After the Revolution of Dignity, a lively discussion broke out in the Ukrainian society and politics about the need to erase the Soviet toponymy and symbolism, as propagating the values of the criminal communist regime, from the faces and maps of Ukrainian cities. Consequently, on April 9, 2015, the Ukrainian Parliament passed the Law "On Conviction of the Communist and National-Socialist (Nazi) Totalitarian Regimes in Ukraine and Prohibition of the Promotion of Their Symbols". In addition to other aspects, this Law regulates the issue of naming and renaming of geographical objects. According to the Law, the list of obligatorily renamed toponyms includes place names related to the persons who held leadership positions in the Communist Party (the position of the secretary of the district committee and above), the higher authorities of the USSR, the Ukrainian SSR, and other Union and Autonomous Soviet republics (except for cases related to the development of Ukrainian science and culture), worked in the Soviet authorities of the state security, the names of the USSR, the Ukrainian SSR, other Soviet republics and derivatives from them, as well as names related to the activities of the Communist Party, the establishment of Soviet power in Ukraine or in its administrative units, prosecution of people struggling for Ukrainian independence in the XX century.

The renaming of streets in Ukrainian settlements is carried out by local self-government bodies within 6 months from the effective date of the Law (until October 9, 2015). If no decision on renaming is taken by the local government within the specified time period, such a decision should have been taken by the village, town, or city mayor within 3 month (until January 9, 2016). If within specified time the mayor is not taking the appropriate decision on renaming, such renaming is carried out by the order of the head of the relevant regional state administration. The head of the regional state administration is obliged to accept such an order in the 3-month period (until April 9, 2016).

Thus, city councils and city mayors played a key role in renaming the urban toponymy in the majority of cities. Special commissions were created at city councils, which received and studied suggestions on renaming that came from the local inhabitants and history experts.

Due to historical circumstances, contemporary Ukraine is quite heterogenous in terms of historical and cultural background. The network of historical regions overlaps with ethnic and language polarization from the west to the south-east. This generates high level of development of regional identities, more or less integrated with the national (or ethnic) one. Nagorna (2008) performed a rather in-depth analysis of the formation and present state of regional identities in Ukraine, with a focus on socio-political manifestations and consequences, concluding that regional identity has turned into a powerful factor affecting the entire sphere of geographic, socio-economic and ethnocultural life. Rewakowicz (2010) examined the representations of four Ukrainian cities in a few selected fictional narratives by four contemporary Ukrainian authors and found out underscored sense of belonging to the local territory, yet the sense of belonging to the nation and the world is not dismissed. A series of regional identities were identified by Melnychuk et al. (2014) using names of private enterprises as identity markers. A number of studies (in particular, comparative) are devoted to the peculiarities of the identity of the population within certain regions (Melnychuk and Gnatiuk 2012; Gnatiuk 2015; Gnatiuk 2016) and cities (Mikheyeva 2008; Komarov 2008; Gnatiuk 2012). Strong regional effect on political values and voting behaviour has been shown in the papers of Birch (2000), Barrington (2002), Barrington and Herron (2004) and Kulyk (2016). Some regional stereotypes are permanently used by leading political parties in mobilizing their electorate (Osipian and Osipian 2012).

Taking into account the foregoing, the study of the new Ukrainian urban toponymy may reveal various strategies used by local authorities in the renaming process, including the desire to politicize or depoliticize the toponymy, making reference to certain historical epoch or deliberately avoiding some strata of national identity. If the historical memory of certain epochs or governmental powers plays consolidating role for a nation, then corresponding toponymy will be encountered throughout the country, even where this historical memory stratum is not represented. Conversely, if these elements of historical memory are unacceptable for people in a certain part of the country, we will see significant disparities in their spatial distribution. Also, based on the spatial distribution of the regional-specific place-names, we may draw conclusions about the spatial limits of the certain regional identities.

4. Data and methods

The study covers 36 largest cities in Ukraine with a population of more than 100,000 within government-controlled territories outside the temporarily occupied Crimea and part of the Donbas, and focused on the new street names that appeared as a result of decommunization. Information on renaming was taken from the official decisions of local governments and administrative decrees of the heads of local state administrations. We traced the renaming of streets, squares, lanes, passages, avenues, embankments and other similar elements of urban infrastructure. Hereinafter, the term “street” is used referring to all of the above mentioned elements of the urban landscape. Unlike many other studies, this one does not take into account the hierarchy of streets and is not limited to the historical centre of cities.

Toponyms that arose as a result of decommunization were classified according to a number of criteria. The employed taxonomy is similar to those used in papers of Bucher et al. (2013) and Stiperski et al. (2015). First, they distinguished toponyms of local, regional, national and international scale; second, the set of street names was classified into basic groups: personalities, geographical names, historical events/institutions, craft and trade, and the other toponyms. Personalities were further divided by the realms of politics, art and culture, religion, science, business and entrepreneurship. However, out taxonomy is different as it was adapted to the country context and the goals of this particular study. First, we distinguished between the following main categories of the decommunized toponyms:

I. Restored historical names: the street was given back one of its previous names.

II. Non-historical names: the street has never borne such a name before.

The non-historical names were divided into three groups:

1. Topographical names, indicating the location of the street relative to other elements of the urban landscape. E.g.: Vokzalna (near the railway station, “vokzal” in Ukrainian); Zakhidna (in the western part of the city, “zakhid” in Ukrainian); Filvarkova (in the historical urban area of Filvarky).
2. Commemorative names. These names immortalize the memory of certain real historic figures, organizations, events, phenomena etc. that may or may not be directly related to the street. E.g.: Hrushevskoho (after the first Ukrainian President in 1917–1918), Nezalezhnosti (in honour of Ukrainian independence, “nezhalezhnist” in Ukrainian), Konotopskoyi Bytvy (in honour of the Battle of Konotop).
3. Poetic (or figurative) names. This group includes toponyms that do not relate to the actual characteristics of the street and provide some emotional load (usually positive): “Soniachna” (Sunny), “Radisna” (Cheery), “Mriylyva” (Wistful), etc.

Restored historical names, topographical names and poetic names taken together may be referred as non-commemorative names, because they were given without a direct goal to memorialize somebody or something; these names rarely have certain ideological burden.

Commemorative names were further classified. First, they were divided between (a) political and military names and (b) other names. The first group unites names related to political and military historic figures, events and concepts. These toponyms were additionally divided into subgroups according to the respective historical context: Kievan Rus; Polish-Lithuanian Commonwealth; Cossack State; Russian and Austro-Hungarian Empires; Ukrainian Struggle for Independence in 1917–1922; Soviet Union; Ukrainian Insurgent Army and related liberation movements; Independent Ukraine (including Revolution of Dignity and Donbas military conflict). The second group includes all other commemorative names, predominantly related to the realms of culture, science, religion etc. These names, according to their denotations, were also additionally divided into pre-Soviet, Soviet-persecuted (prominent people persecuted by the Soviet authorities for several reasons), Soviet-favoured (prominent people having no obvious problems with the Soviet regime), ex-Soviet (people, related to Ukraine but constantly living outside of Ukraine in the Soviet epoch), and post-Soviet ones.

Second, commemorative names were classified according to their relation to the local, regional and national context. If memorialized person, object or phenomenon has some relation to the city in which the street is located, it was classified as local-specific; if it has no relation to the city but has some relation to the surrounding region (administrative oblast), it was classified as region-specific; if it has no relation neither to the city nor to the oblast, but has some relation to Ukraine in whole, it was classified as nation-specific. Finally, if this memorialized person, object or phenomenon has no relation to the city, region and nation, or this relation cannot be strictly defined, the name was treated as non-classified.

Conclusions and suggestions given in this paper are based on the spatial distribution of aforementioned groups and subgroups of street names, as well as their proportions (structures) and mutual correlations in particular cities and regions.

It should also be noted that in the cities of Lviv and Ternopil very few toponyms were renamed after 2014. Therefore, conclusions about these cities should be taken carefully in view of the lack of sufficient statistical data set. However, in order to confirm or refute some of assumptions, previous post-Socialism renaming history of 1991–2013 was reckoned with.

5. Results and discussion

The total quantity of place-names in test cities, renamed since 2014, is 2897. The quantity of such toponyms per city fluctuates between a few streets in the cities of Western Ukraine to several hundred streets in the cities of the southeast.

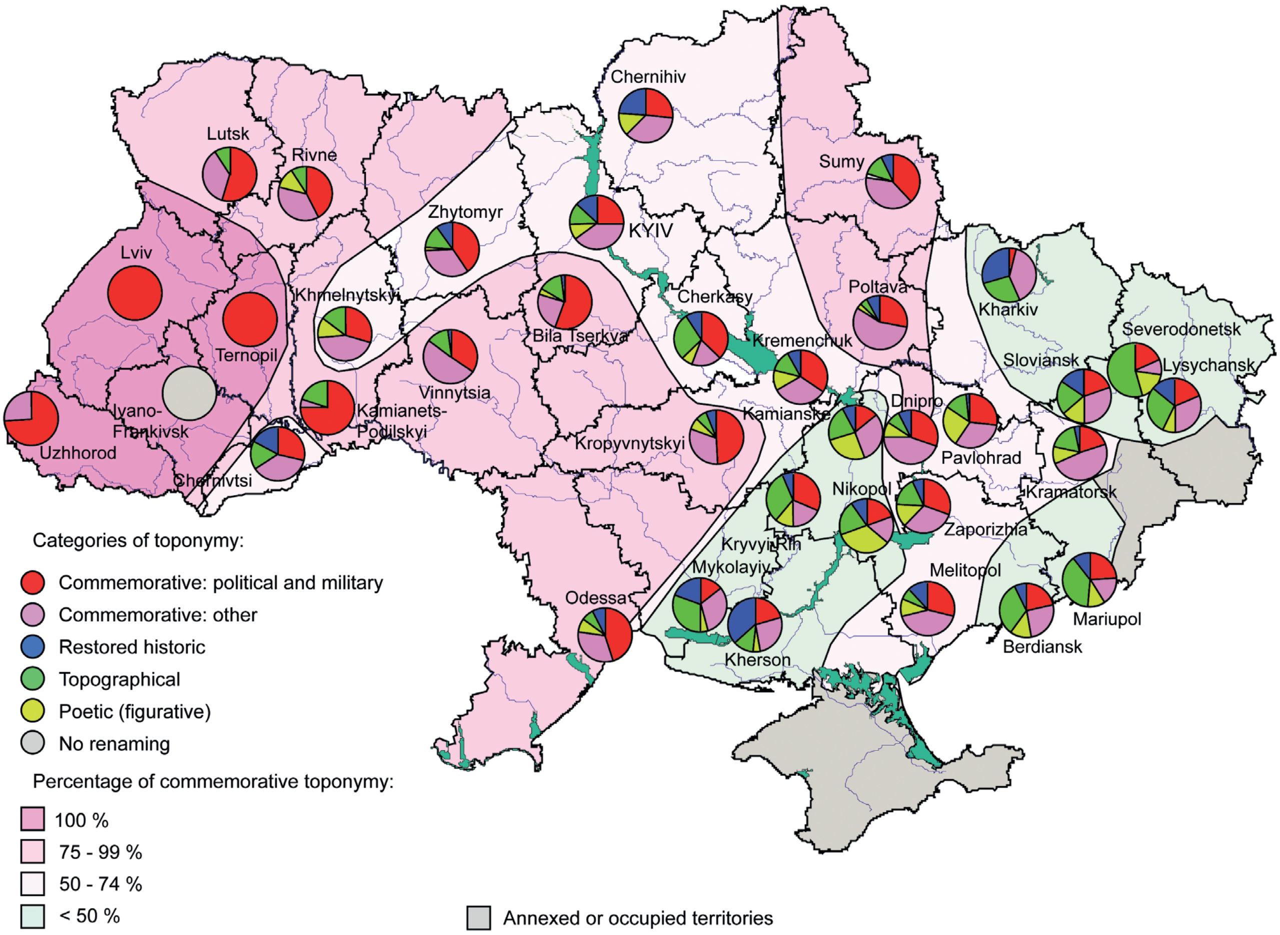

Fig. 1 Main categories of renamed toponymy.

The share of commemorative urban place-names gradually decreases from the west to the south-east (Figure 1), while the total share of non-commemorative toponymy increases in the same direction. The ratio of military-political and other commemorative toponymy shows a great individuality depending on a particular city, but generally also decreases in the south-eastern direction: from the absolute majority of renamed place-names in Galicia to 20–25% in Black Sea Region, Prydniprovya and Donbas.

This suggests that the renaming strategy in most cities in the south-eastern Ukraine was targeted to maximally avoid the commemorative names, especially those related to military-political events and personalities. The motivation for such a strategy could be as follows:

1. The toponymy associated with the Ukrainian national liberation movements of the XX century is unfamiliar and alien to a large part of the population of the southeast, moreover, as a result of prolonged communist propaganda, they are perceived by the public as plainly hostile. During the Soviet era, these events and historic figures were marked as ideologically hostile, bourgeois, and some of them (in particular, figures of Ukrainian Insurgent Army) were treated as criminals and collaborators of the Nazis. Moreover, until the very Revolution of Dignity, these speculations were widely used not only by left-wing political forces, but also by the ruling oligarchic and pro-Russian Party of Regions to mobilize the electorate. Therefore, the city governments, which include also representatives of the above-mentioned political forces, firstly, sought to minimize the possible discontent of the locals with the mass appearance of such names, and secondly, implemented their own position on their inappropriateness or even inadmissibility.

2. Alternative could be the place-names associated with prominent figures of the Soviet era not prohibited to memorialize by the decommunization law. As will be shown below, such toponyms really appeared after renaming and concentrated exactly in southern and eastern Ukraine. However, the city authorities feared to massively approve such names, given that such a step could be considered as a betrayal by patriotically-minded citizens and as a signal of poor loyalty to the Ukrainian national project.

3. The cities of the southeast of Ukraine, where decommunization practically had not started before 2014, had significant reserves for restoring the historical street names. Some cities have really gone this way. As a rule, these are large cities (Kharkiv, Kherson, and Mykolaiv), which at the beginning of the Soviet era already had a significant number of

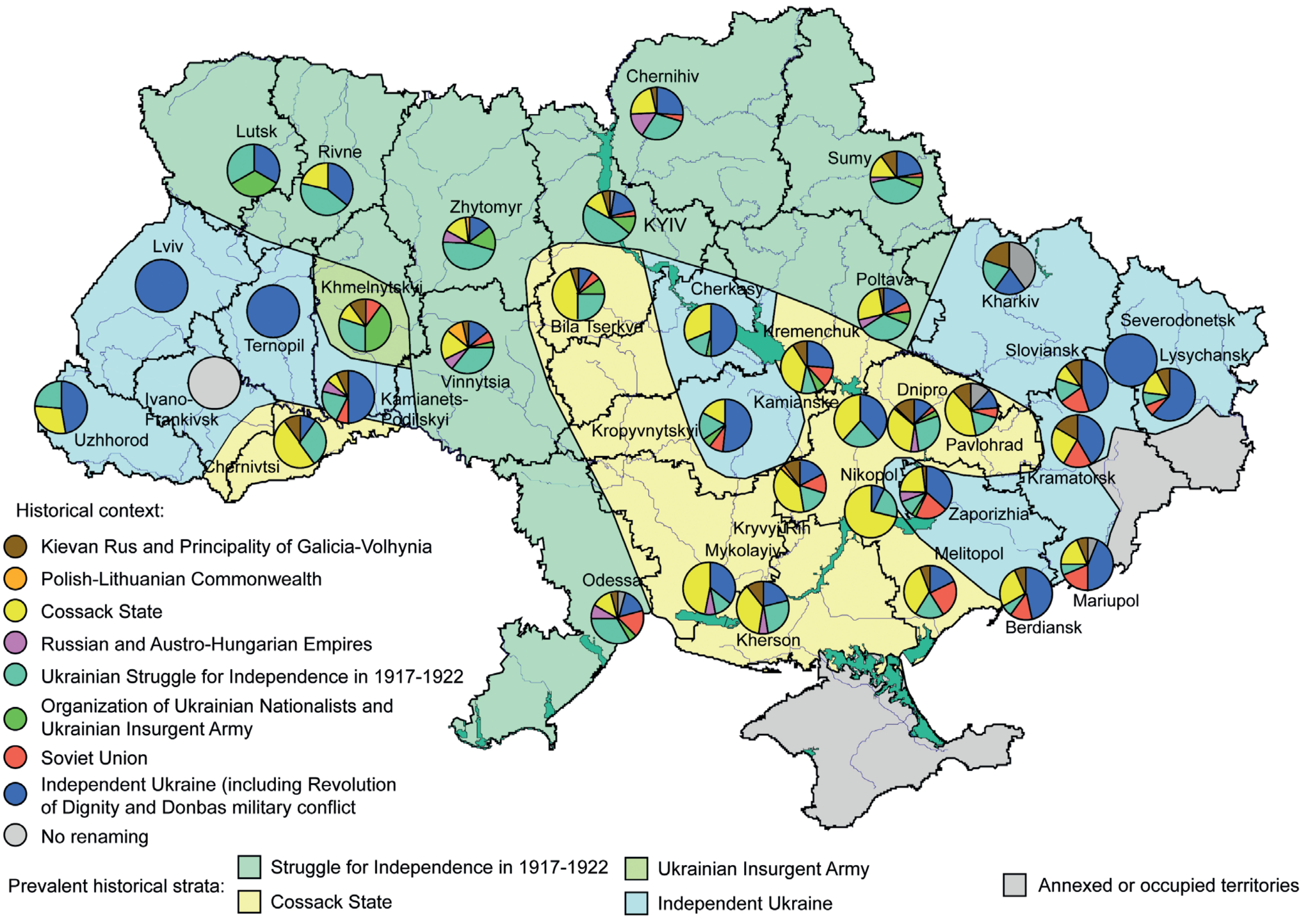

Fig. 2 Politic and military commemorative toponymy: historical cross-sections.

streets today concentrated in their historical centres. It is specific that the proportion of restored historical street names within Donbas is considerably higher in the more ancient Slovyansk, Lysychansk and Mariupol compared with more recently founded Kramatorsk and Severodonetsk. Interestingly, beyond the region of the southeast, significant proportions of the restored historical names are also observed in cities with a long history: Chernihiv, Chernivtsi, Kyiv, and Zhytomyr. At the same time, cities, having no possibility to massively restore historic names (or, e.g. city authorities did not want to restore historical names, since they were too closely related to the era of Russian tsarism), approved a large number of topographical and/or poetic street names.

Such a strategy of avoiding commemorative names was explained to the public by the need of deideologization and depoliticization of urban toponymy. I.e., it is not reasonable to assign the names of new heroes to the streets, if over time the power will change and a new wave of renaming will begin. Neutral names, which will not interfere with any government, must be assigned instead. However, this way of thinking holds two hidden ideas: first, attitude towards the idea of Ukrainian national state as an alien phenomenon, second, the perception of the current Ukrainian political regime as a temporary one.

The brightest exceptions among the cities of the south-eastern Ukraine are Dnipro and Odessa. These cities have a relatively large proportion of both commemorative street names in general and military-political street names in particular, comparable to the cities of the right-bank part of Central Ukraine. To understand the reasons for this situation, one must remember that in Odessa the streets were renamed not by a decision of the city council, but by the order of Mikheil Saakashvili, the head of the Oblast State Administration. Thus, this renaming pattern does not reflect the vision of the city government and the local community; it rather reflects the position of the central government. As for Dnipro, the key factor is the role of this city as a patriotic Ukrainian outpost in the eastern part of the country and an organizational base for logistical and medical support of military actions in Donbas.

As for the cities of the Western Ukraine and, to a lesser extent, the Right Bank of Central Ukraine, the high proportion of commemorative names in these regions is explained, firstly, by the fact that the strategy of local authorities is directed here to immortalize the memory of the characters of the national liberation movement leaders (which is perceived as a normal and, moreover, a necessary phenomenon by both the authorities and the population), and secondly, much smaller possibilities for the apparition of non-commemorative street names, at least because of the fact that most of the Soviet urban toponyms, especially in the central parts of the cities, had been renamed there long before the Revolution of Dignity.

On the map (Figure 2) it is easy to see the three most represented categories of military-political commemorative street names. Firstly, these are names related to the Cossack epoch, which are present in almost all test cities; they make up an especially large share in the south-eastern part of the country, in particular in Prydniprovya. Secondly, these are names related to the struggle for independence in 1917–1922. They are also distributed practically throughout the territory of Ukraine, but the largest share is observed in the cities of Western and Central Ukraine, whereas in the south-east, especially in Donbas, their proportion is significantly smaller. Thirdly, these are toponyms related to the prominent figures and events of the modern independence period, and more than 90% of them concern the commemoration of victims of the Revolution of Dignity and the current military conflict in the Donbas. This category of street names is particularly well represented in Western Ukraine, as well as in the eastern part of the country, especially in Donbas.

Other categories of military-political toponyms are relatively less represented and/or have limited spatial distribution throughout the country. The names related to events and prominent figures of the Soviet era are concentrated in the south-east, especially in the cities of Donetsk and Zaporizhia oblasts, where their share makes up 15–25% of the total number of military-political toponyms. Simultaneously, in the central part of Ukraine, this figure is reduced to 5–10%, while in the western regions this category of toponyms is absent at all. The share of the urban place-names associated with the Kievan Rus legacy is increasing from the west (where there are practically absent) to the south-east (with a maximum of 15–20% in the cities of Slobozhanschyna, Donbas and Prydniprovya). Street names, associated with the Ukrainian Insurgent Army, are very unevenly distributed over country: their share reaches 15–20% in the cities of the west and north of the country and 5–10% in the central part, but in the cities of the southeast, with the exception of Odessa, Dnipro and Zaporizhia, such toponyms are totally absent. The names associated with the Russian Empire (predominantly of local significance) are present in small quantities throughout the country, except for the extreme west (Galicia, Volhynia, Transcarpathia, and Northern Bukovina) and the extreme east (Donbas). Finally, few toponyms, related to the Polish-Lithuanian period, are located in several cities of the central Ukraine (Zhytomyr, Vinnytsia, Kryviy Rih, and Dnipro).

These spatial patterns are affected by the two main factors.

The first factor is: how the development of each particular city is connected with given historical epoch. Exactly this factor can explain the following regularities: (a) Concentration of names associated with the Cossack epoch in Prydniprovya, where the Cossack Sichs (fortified military settlements) were located, and in Central Ukraine, which was under intense and long-term control of the Cossack state; (b) Concentration of names related to the struggle of the Ukrainian people for independence of 1917–1922 in the central and western parts of Ukraine, which were under the most effective and long-term political control of the Ukrainian Peoples Republic and Western Ukrainian Peoples Republic; (c) Concentration of names related to the Ukrainian Insurgent Army in the areas of its highest activity; (d) Concentration of names associated with the Russian Empire and the Polish-Lithuanian Commonwealth in the areas that were actually controlled by these states; (e) Concentration of urban toponyms associated with the modern Ukrainian history, in Western Ukraine, from where the largest number of Maidan victims are descending, in Donbas, where many military men and women lost their lives, and in certain cities through the country, where troops, suffered heavy losses during the conflict, have their headquarters.

However, this factor cannot fully explain all the revealed regularities. E.g., why the names associated with Kievan Rus are not concentrated in the historical core of this state (i.e. around Kyiv, in the Middle Dnieper Region), but have the largest density in the south-east of the country, which at that time was practically unsettled and controlled rather by nomadic tribes of Torks, Cumans and Pechenegs then by Kyivan princes? How to explain the high proportion of names related to the Cossack epoch in the western cities of Chernivtsi, Uzhhorod and Rivne, where the role of Cossacks in local history was minimal? To a lesser extent, this is the very issue of the seaports of Mykolayiv and Kherson, as these cities were founded and intensely developed during the Russian Empire occupation without the direct participation of Cossacks. And, at the same time, why Russian Empire are poorely commemorated in Donbas, whose economic base was laid in precisely that period? Finally, why in Western Ukraine, which for the longest time was under the control of the Polish-Lithuanian Commonwealth, there are no new toponyms associated with the corresponding era?

Therefore, it is advisable to talk about the second factor: the ability of the corresponding historical epoch to be a consolidating element of identity for all Ukrainians, regardless of their ideological preferences and the place of residence.

Thus, aforementioned strange spatial distribution of the names associated with Kievan Rus and spatial sprawl of toponyms associated with the Cossack epoch is explained by the fact that these historic periods are equally well perceived by residents of all regions and representatives of different ideologies from right to left over the spectrum. Implementing the renaming policy, local authorities in the south east gave preference to Kievan Rus and Cossack legacy because such names are equally well acceptable by all strata of Ukrainians, regardless of ideology and electoral preferences, and no one will protests. Simultaneously, it was a convenient way to minimize the quantity of place names, negatively perceived by both pro-Russian and pro-Soviet population (e.g., related to the activities of the Ukrainian Insurgent Army) and pro-Ukrainian and pro-EU patriotic population (e.g., toponyms associated with the Soviet legacy).

The numerical prevalence of the toponymy associated with the Cossack epoch is the consequence of the fact that this period of Ukrainian history is considered to be an era of struggle for national self-determination, unlike the era of Kievan Rus, which is not directly and exclusively identified with the Ukrainian national state.

The share of military-political names, which are clearly not related to any specific period of Ukrainian history, or do not have any relation to Ukraine at all, increases in the south-eastern direction, and this is another manifestation of the strategy for de-politicization.

The struggle of the Ukrainian people for independence in 1917–1922 was anti-Soviet one, therefore it is negatively perceived by the part of Ukrainians, especially in the south-east, where nostalgia for the Soviet Union still persists. However, these street names are presented in all test cities, except for Kramatorsk and Severodonetsk, and in absolute number of names they are only slightly inferior to the "Cossack" toponymy. Consequently, this liberation movement also becomes a consolidating element of Ukrainian national identity, and its values are more or less shared by people from all regions of Ukraine. However, the unifying potential of the Cossack heritage is still more powerful, given the lack of nationalistic connotations. Examples of Chernivtsi and Uzhhorod are quite illustrative here. Corresponding regions of Transcarpathia and Northern Bukovyna have traditionally been characterized by a smaller national patriotic pathos than the rest of the Western Ukraine. Therefore, when renaming the streets, the local authorities of these two cities decided to appeal to the Cossack and Kievan Rus heritage rather than to the XX century liberation movements.

On the contrary, the legacy of Ukrainian Nationalists Organization and Ukrainian Insurgent Army still remains rather disruptive than uniting factor for Ukrainian society, because it is perceived negatively in the south-eastern part of the country. The Soviet legacy is a very similar case: it serves as a "red cloth" for the population of the central, northern and, especially, western regions.

It is also interesting that among the cities of the southeast, the largest share of street names related to the national liberation movements, is observed in Odessa (where the renaming, as noted above, does not reflect the vision of the local self-government and community), Dnipro ("The Eastern Outpost of Ukrainians", "The Rear of the Donbas Front") and Kherson (a city in the agrarian region where the proportion of the pro-Russian population is relatively small comparing to the other south-east regions).

The absence of the "Polish-Lithuanian" toponymy in the regions of Western Ukraine may be explained by the high level of national patriotism and hidden but confrontation with the Poles on the grounds of the Ukrainian-Polish war in 1918–1919 and the Volhyn tragedy of 1943–1944. Instead, in the Central Ukraine, a positive assessment of the Polish-Lithuanian contribution to urban development prevails, in spite of the well-known Ukrainian confrontation with the Poles during the Cossack era. In addition, in Western Ukraine, "Polish-Lithuanian" toponyms could simply not be statistically represented, given their low frequency and the small number of renamed streets. At the same time, such toponyms occur quite often among the streets, renamed before 2014.

As a result, the analysis of military-political names made possible to identify areas where one or another historical stratum of national identity prevails or is currently perceived as the most important (Figure 2).

Since the heritage of the Cossack and the Kievan Rus periods is used to depoliticize urban toponymy, while the legacy of national liberation movements of the XX century, on the contrary, to support the current national political course, and the inheritance of the Soviet period – to emphasize the genetic link with the Soviet Union, interrelations of these three toponymy strata makes it possible to highlight several hidden geopolitical and cultural spatial faults of modern Ukrainian society (Figure 3). The first fault separates areas with a predominance of names associated with the national liberation movements of the XX century and areas with a predominance of "Cossack" toponymy (Figure 3a). The second fault, much more sharp comparing to the previous one, delineates areas with a predominance of names associated with the national liberation movements of the XX century and areas with a predominance of Soviet-era place names (Figure 3b). These two fault lines conditionally divided the country into three parts, which can be called "nationalistic Ukraine", where national identity has clear anti-Soviet and pro-European orientation, "Cossack Ukraine", where the identity of the Cossack epoch plays a role of refined, depoliticized version of national identity, and "post-Soviet Ukraine", which Soviet values are still regenerated, even though without clear ideological wrapper. Interestingly, the limits of "post-Soviet Ukraine" roughly coincide with the area where ethnic Russians constitute significant share of the total population and where the Oppositional Block, successor of the Yanukovych's Party of Regions, won local elections in 2015.

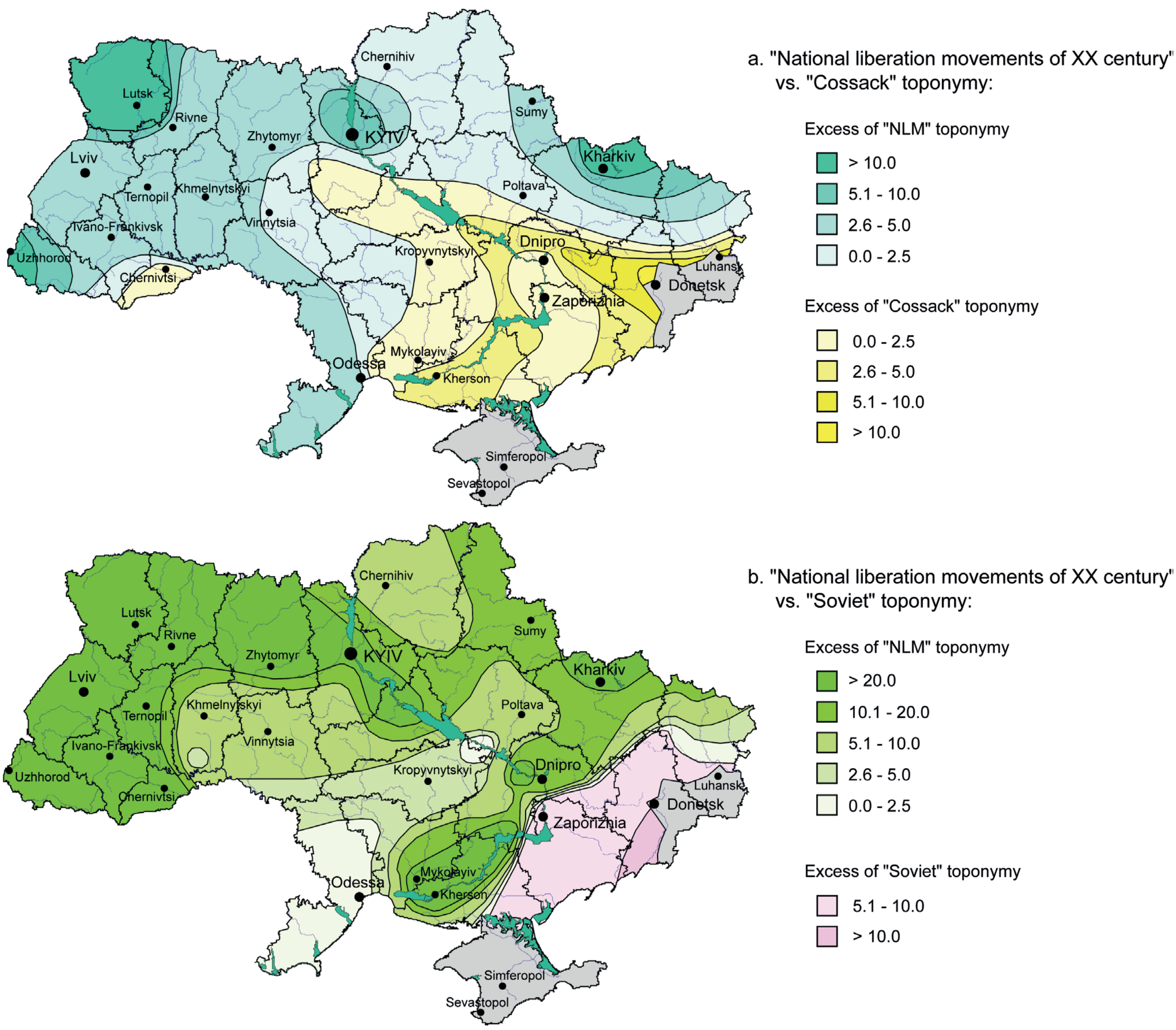

Fig. 3 Relationship between three main strata of politic and military toponymy.

In respect to other commemorative names, which are not directly related to the military-political sphere (Figure 4), the most numerous group of them is associated with the heritage of the pre-Soviet era, mostly the Russian and Austro-Hungarian empires. This phenomenon, like in other post-Socialist countries (Crljenko 2012), marks an accentuated reference to the pre-Soviet history in order to reinforce the national identity. However, toponyms, directly related to the Soviet era (i.e. given in the honour of Soviet-favoured and Soviet-persecuted personalities) account together for more than 50% of all other commemorative names. This leads to interesting considerations, since the military-political names of the Soviet era comprise only 10–20%, even in the most pro-Soviet and pro-Russian regions. Of course, the possibility to introduce Soviet political and military toponyms was substantially limited by the decommunization law, which is not a case of toponyms related to the realms of culture and science. However, it may be supposed the apparent desire to distance from the military-political heritage of the Soviet Union, but a radical departure from the cultural heritage of the Soviet period is not observed, since it continues to be interpreted as a powerful and inalienable layer of Ukrainian culture in whole.

But, perhaps, these prominent figures conflicted with the Soviet authorities, and were memorialized exactly for this? In fact, we observe numerical equality of the street names related to the Soviet-favoured and Soviet-persecuted personalities. Place-names, given in honour of Soviet-favoured people, are concentrated mainly in the south-east, while toponymy related to Soviet-persecuted personalities prevails in in the central and western parts of the country; however, this general rule has many exceptions. Nevertheless, this spatial pattern becomes more accentuated if consider the ratio of the sum of the soviet-persecuted and ex-soviet toponymy to the sum of the soviet-favoured and international toponymy (Figure 4). This implies

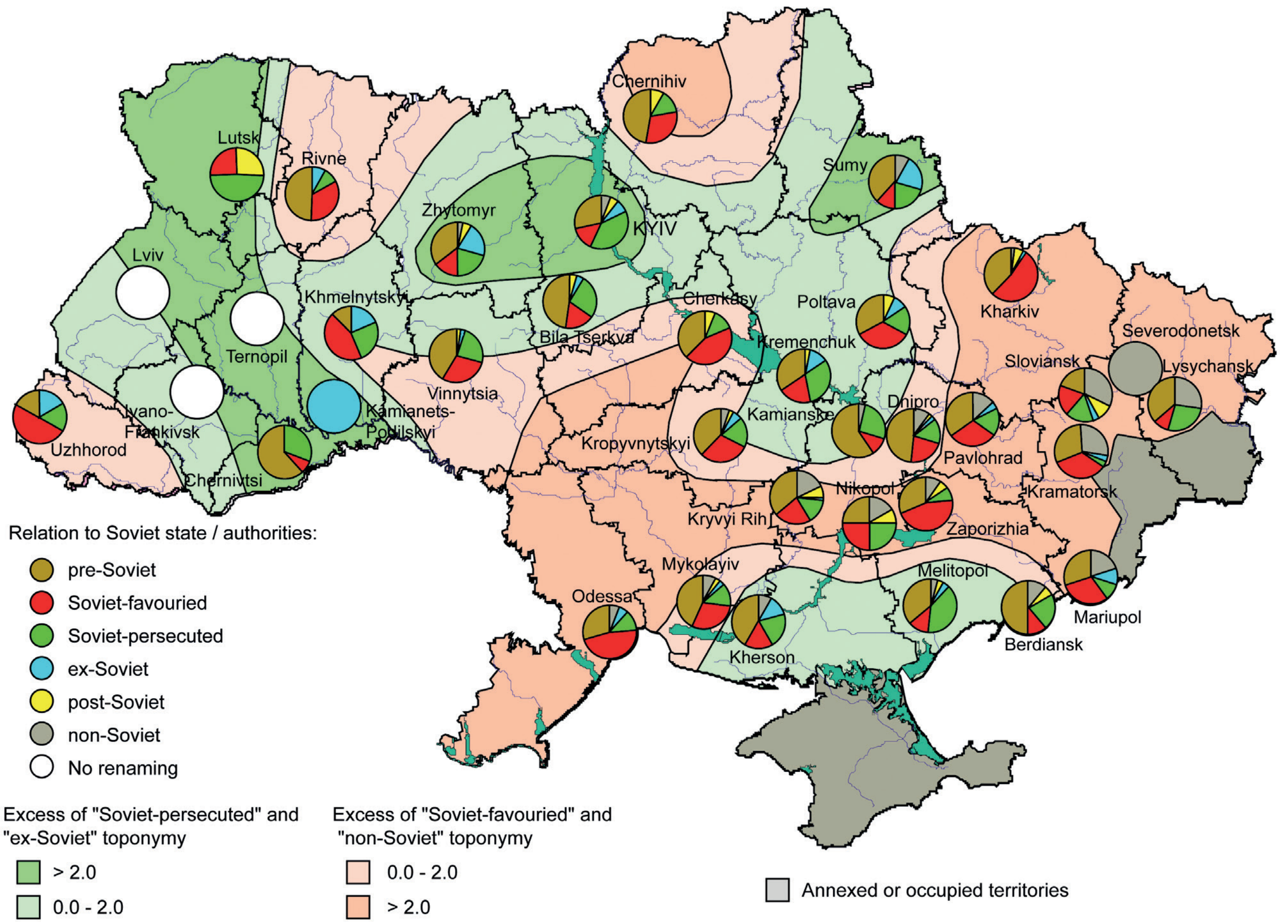

Fig. 4 Other commemorative toponymy: relation to the Soviet authorities/state.

that in western and central parts of the country we observe pronouncedly anti-Soviet commemoration policy, while the south-east once again shows an attempt to depoliticize toponymy and to maximally integrate Soviet heritage into the new identity.

The most unevenly distributed toponyms are those given in honour of the prominent individuals who, due to life circumstances, including persecution by the Soviet authorities, lived and worked outside of Ukraine. These names are concentrated in the western and central parts of the country, which means that exactly in these parts of Ukraine the activities of Ukrainians in emigration is considered as an integral part of national history.

Also, there is a noticeable downfall of the share of post-Soviet other memorial names in comparison with their military-political counterparts: their quantity is simply negligible, and they have no specific spatial distribution. But this does not give grounds for arguing that the modern identity of Ukrainians has only political and military components. The matter is that in the vast majority of cases, like in the whole world, commemorated people are those who have already departed from their lives. In modern Ukraine, these are not representatives of cultural, scientific, or business elites, but the victims of the Maidan and Donbas conflict.

The analysis of the ratio between local-specific, regional-specific and national-specific memorial toponyms shows that decommunization retains the balance between, on the one hand, local and regional, and, on the other hand, national identities. In almost 80% of test cities, the share of national-specific street names is in the range of 25–75%. That is, in only one case of five we have a significant imbalance.

Significant spatial patterns of placement of cities with an emphasis on local or national identity are not observed (Figure 5). In particular, strong preference for local and regional-specific place-names is typical for certain cities from western (Uzhgorod, Lutsk), central (Cherkasy, Vinnytsia, Bila Tserkva, Poltava), southern (Odesa, Mykolaiv) and eastern (Zaporizhia) parts of Ukraine. The only clear regularity is the reduced share of local-specific and, especially, regional-specific names in Donbass. This may be explained, firstly, by the desire of the local authorities to integrate the local toponymy into the national context, and secondly, by the absence of a well-developed regional Ukrainian national narrative in Donbas.

The increased share of names not related to any level of Ukrainian national identity is another characteristic feature of the renamed toponymy in the south-east, especially in Donbas. Typically, such names are associated with outstanding figures of world culture.

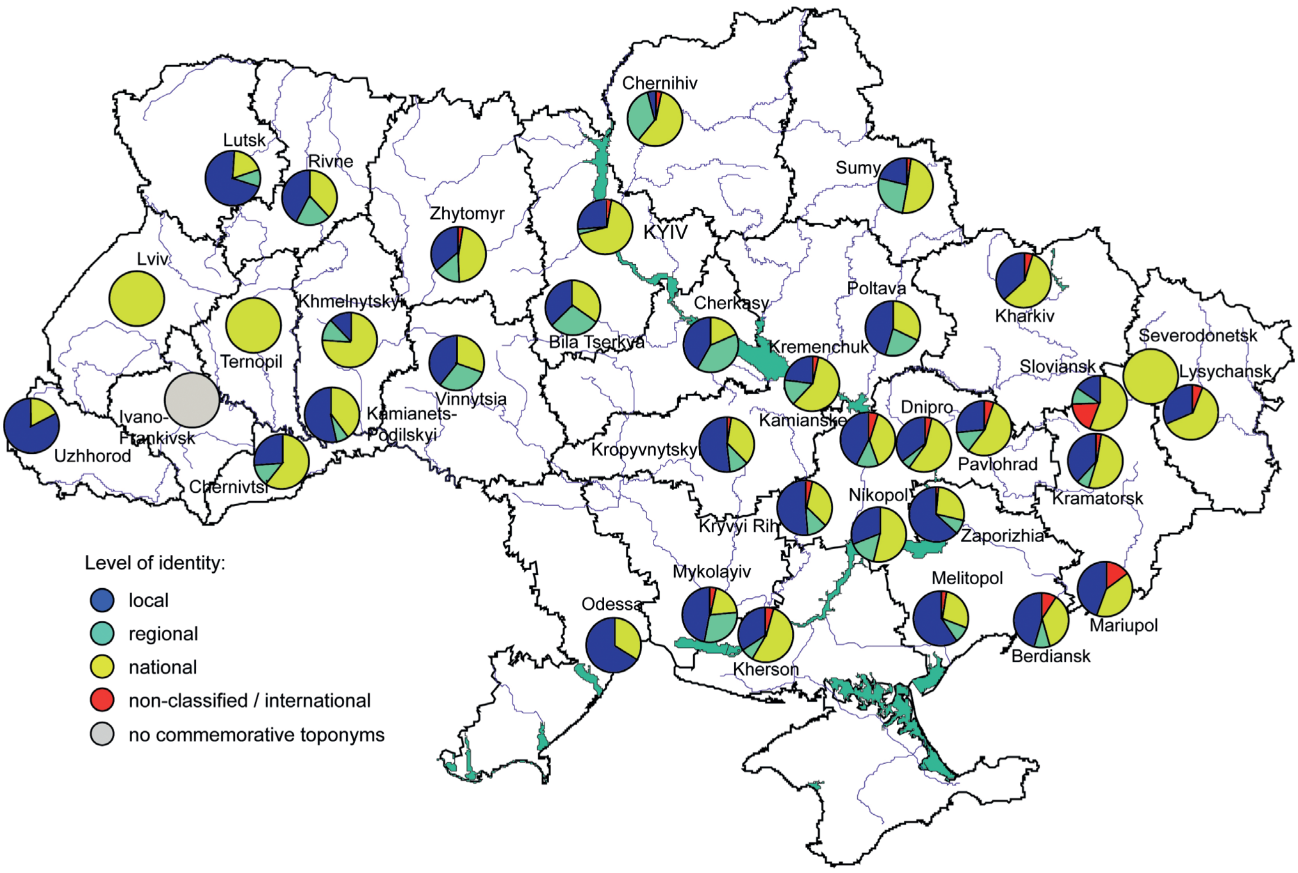

Fig. 5 Toponymy by levels of identity.

This is a testimony of certain cosmopolitanism in the selection of new names, while in the rest of Ukraine a clear priority is given to memorialization of the very national culture. Interestingly, similar characteristics (a high proportion of national and non-specific names) are also observed in Kyiv. This corresponds to the capital status of this city: here, street names should reflect the peculiarities of different regions of the country and, at the same time, should have a certain scale of cosmopolitanism.

What the spatial patterns discovered do reflect more: the real identity of population or local government policy of constructing identity? Given the coincidence with the electoral fractures and the ethnic composition of the population, it is very likely that the spatial structure of the new Ukrainian urban toponymy does corresponds to the real structure of identity, although in many cases the semantics of place names does not reflect the exact meaning of real identities. E.g., if the great importance of the Ukrainian liberation movements for the population of the western and central parts of the country is beyond doubt, then excessive accentuation of the Cossack era in Chernivtsi or the cities within the Black Sea region is clearly artificial. This means that local authorities have proposed such street names to satisfy the locals. That is, the ideology formulated in toponymy does not contradict the identity of the locals, and therefore is well-accepted.

The analysis, presented in the paper, would be incomplete and poor reasoning without giving some review of a public discussion related to the renaming policy, procedure and regional approaches. In this paper we will not pay attention to the general discussion on the feasibility/priority of the decommunization: our focus will be on the part of the discussion that is related to competition between different identities and layers of historical memory.

At the local level, the debate took place during the collection of suggestions on renaming specific streets and the subsequent public discussion of new street names proposed by the commissions. Often, several variants of the new street name were presented at the discretion of the community, and the discussion participants could vote for each of them. Basically, this was so-called calm discussion, that is, discussions and voting for a street name occurred smoothly and within a defined procedure. However, sometimes there was a hot discussion, when there were sharp controversies regarding specific names. In this respect, there were strong regional tendencies manifested in the greater public commitment to the memory of a certain historical epoch or the desire to distance from politicized names in favour of the neutral toponymy (topographical or figurative toponyms, or, where possible, restoration of the historical names).

For example, in Cherkasy (Central Ukraine), residents gave a steady advantage to commemoration of the XX century national liberation movements and the Cossack Epoch. At the same time, in Dnipro and Kamianske, the idea of changing the Soviet toponymy to the actual and patriotic one was questioned, given the high versatility of the political situation and the complexity for the elderly people to get accustomed to the change of the usual street names into radically opposite ones. The opinion was expressed that restored historical names, as well as timeless, non-political and even non-contextual toponyms, are the best choice (DneprChas, 17 June 2016; Novoe Vremia, 12 January 2018; Disk, 15 June 2015). "The residents themselves expressed a desire to give politically neutral place-names, and we listened to them. To avoid future renamings, we turned to names reflecting human emotions, nature, and professions" – emphasizes Serhiy Svitlenko, a member of the renaming working group, the dean of the historical Faculty at Oles Gonchar Dnipro National University. As a result, 130 toponyms in Dnipro have got neutral names (Dnepropetrovskaya Panorama, 29 June 2016).

It is revealing that the proposals to rename Lenin and Moscow streets in Dnipro in honour of nationalist leader Stepan Bandera were rejected: the preponderance was given to the historical name (Voskresenska) and to the commemoration of the Kyivan Rus Prince Volodymyr Monomakh, respectively. Instead of Roman Shukhevych, another nationalist leader (proposal of the renaming commission), one of the streets was named after Leonid Stromtsov, director of the Dnipro machine-building plant in Soviet times (1954–1983) (DepoDnipro, 14 August 2015). The inhabitants of some streets stood up against erasing the names of the Soviet figures and replacing them with some Cossack leaders. Residents of one of the streets created an initiative group and said they did not want to live on the street named after hetman Ivan Vyhovsky, since his period of rule belongs to the Ruin period of Ukrainian history; the initiative group insisted that the street should retake its historical name instead (Polihonna), which was finally done (Radio Svoboda, 18 August 2015). However, the Prospect Kalinina (Kalinin Avenue) was renamed in honour of Serhiy Nigoyan, the first of the protesters killed in Ukrainian capital during Euromaydan; this decision was supported by the community during a public discussion, then challenged by opponents in court, but protected due to a clear and strong position of the city authorities (tyzhden.ua, 15 December 2017).

Particularly keen discussion of a somewhat different nature unfolded in Dnipro regarding the renaming of Prospekt Karla Marksa (Karl Marx Avenue), the main street of the city. Outstanding personalities have become competitors in the process of changing the toponymy. Olexandr Pol (1832–1890), archaeologist, entrepreneur, philanthropist and active public figure, thoroughly studied local iron ore basin for 15 years. The director of the Olexander Pol Institute in Dnipro believes that "the region, which was purely agrarian, became a powerful industrial centre thanks to Pol and his followers, therefore, it is logical to commemorate him in the name of the central avenue". The choice in favour of Olexandr Pol means the emphasis on the industrial history of the city. Proponents of Dmytro Yavornytsky (1855–1940) had their arguments: an outstanding scientist dedicated more than 30 years of his life to the Dnieper region; he is considered to be the founder of archeology in Ukraine and the father of the study of the Ukrainian Cossacks; he wrote the first history book of the city, was the founder of a local university, etc. The Head of the Dmytro Yavornytsky Memorial Museum said: "It is difficult to find a significant event in the life of the then-city [Dnipro], to which Yavornytsky would not be involved." Therefore, the choice in favour of Yavornytsky is an attempt to emphasize the continuity of the city with the Cossack epoch. In fact, it was the choice of the city's brand, the personification of its essence. Finally, Yavornytsky won this toponymic battle, and toponyms associated with the Cossack epoch constitute the relative majority in Dnipro. However, another central street of the city, Prospekt Kirova (Kirov Avenue) was renamed in the honour of Olexandr Pol (RadioSvoboda, 03 August 2015).

It is revealing that both these figures are closely connected with the city of Dnipro and the surrounding region, which cannot be said about their toponymic predecessors (Kirov and Karl Marx). In many cities, decommunization committees decided to strictly follow the will of the community and not offer new names other than those proposed by the residents. This led to an emphasis on the local historical and cultural context in cities with developed local identity. For example, in Vinnytsia (where we were able to get an official comment from a representative of the decommunization committee) there were almost no streets renamed in honour of the OUN-UIA personalities, in particular Stepan Bandera. This is not due to the rejection by local community, but because their activities are not directly related to the city and the surrounding region. The only exception was Omelian Hrabets, a UIA-South partisan group commander, who acted precisely in Podolia, including the Vinnytsia region. Given the overwhelming opinion of the residents, the former Red Partisan Street in Vinnytsia was renamed in honour of Vladislav Gorodetsky, a Ukrainian architect of Polish origin born in Podolia, although representatives of local nationalist movements demanded that the street should be radically renamed in honour of the UIA heroes. The approach was similar in Kryvyi Rih, Zaporizhia, Lutsk, etc.; in this way a number of sharp conflicts regarding new street names were minimized, but urban toponymy was limited mostly to the local and regional context. Cited above Serhiy Svitlenko (from Dnipro) recognized this problem emphasizing that "Inventing new toponymy, we need to look beyond the Middle Dnieper elements and reflect also the history of Western

Ukraine; this is very important politically" (Litza, 22 July 2015).

A striking example of the conflict between the official decommunization approach and local government's policy was the already mentioned situation in Odessa, when the city council initially self-abolished from the renaming process, and then cancelled some of the renamings carried out according to the order of the Head of the Odessa Regional State Administration, and turned back communist toponymy to the city map. However, this decision of the city council was appealed by the prosecutor's office and investigated by the Security Service of Ukraine.

At the national level, the debate on decommunization, covered in media, is confined almost exclusively to the issue of commemorating the memory of the OUN-UIA, as well as other figures that contributed to the formation of the OUN's ideology. This discussion is quite important because decommunization legislation prohibits to glorify both Soviet and Nazi regimes. This means that if an organization or some of its representatives were Nazi collaborators during Second World War, the streets cannot be renamed after them. This moves the discussion from the standpoints of solely personal or collective preferences to the legal dimension.

The official position, transmitted by the Institute of National Remembrance, is that the OUN-UIA members were not collaborators. Some ONU leaders initially sought to use the military and political power of the Third Reich to gain an independent Ukrainian state (on June 30, 1941, the OUN, under the leadership of Bandera, proclaimed the Act of Restoration of Ukraine's Independence with numerous diplomatic laudations to Hitler), but for several weeks, seeing the real Nazi policy towards Ukrainians, understood the mistake. A number of OUN figures, including Stepan Bandera, even found themselves in the Nazi concentration camps soon afterwards and could not physically control either the political or the militaristic formations of the nationalists. The soldiers of UIA, created only in 1942, fought from the outset both against the Red Army and the Nazis.

However, not all citizens and experts share the official statements. The attempt to rename some streets in eastern and southern Ukraine in honour of the OUN-UIA leaders provoked public resistance. Particularly strong discussion in media was induced by the renaming of Moskovskyi Prospekt (Moscow Avenue) in Kyiv in 2016 in honour of Bandera. Online media Gordon asked for opinions of several well-known history experts regarding this fact (Gordon, 8 July 2016). The experts had quite different, individual assessments, but mostly agree the following positions:

1. This renaming in Kyiv was done in the teeth of Russia as the OUN leader now turned into the symbol of struggle against the current Russian policy. The radical renaming of streets is not a mistake but rather stupidity, as this may scare away many Ukrainians who may potentially be patriots but are not (yet) ready to accept Bandera. Such renaming will only increase the distance between the capital and the southeast regions.
2. True history of the OUN-UIA and biographies of their leaders have not yet been rigorously investigated: they are either idealized or demonized with political goals. Some of nationalists did collaborate with the Nazis, but they did want, by this medium, to create an independent Ukraine and gave their lives for this.
3. Since the purpose of decommunization is to create a new value matrix, new toponymy must unite society rather than split it. Unfortunately, Ukraine still lacks an integral historical policy. A mature democratic society should seek (and preferably find) for a consensus, especially in the choice of national heroes. In the historical pantheon of Ukraine there are many worthy people who do not split the society, e.g., dissidents, human rights activists, and writers.
4. Preferably new street names should be related to local specifics, the history of the city, and its outstanding inhabitants. E.g. we may have different attitudes to Bandera, but the important thing in this particular case is that he had nothing common with Ukrainian capital. A good example may be the street renamed after Symon Petliura, a Head of Ukrainian Peoples Republic, who lived in Kyiv and even often drove along "his" street to the railway station.
5. Since military and political regimes and dominant ideologies are constantly changing, it is desirable to retrieve old historical toponymy or give preference to neutral place-names. However, the complete refusal to honour the national heroes will interfere the building of national memory and lead to preservation of the existing fault lines in Ukrainian society. Therefore hiding head in the sand is equally harmful as extremely radical renaming policy.

Obviously, the main message is that new names should consolidate Ukrainian society, and renaming should not be used as a tool to erase one layer of historical memory for the benefit of another. In a similar manner, Yevhen Hendin, the Head of the University Theatre at the Oles Honchar Dnipro National University, also spoke about a similar renaming attempt in Dnipro: "I cannot understand the stressed bending across the knee, the underlined PR, sadistic from my point of view – please, take Bandera instead of Lenin. I consider it not a conciliatory, not a compromise, but a discreditable step. But, at the same time, I think that in principle such a street should be in the city" (DepoDnipro, 14 August 2015).

Renaming of Prospekt Vatutina (Vatutin Avenue) in the honour of Roman Shukhevych, a commander of the UIA, was the other, most recent and even more sensitive case from Kyiv. Shukhevych studied at the

Abwehr, served in the Wehrmacht units, and was one of organizers of anti-Polish and anti-Jewish ethnic cleansings (Rudling, 2016). Moreover, by a twist of fate, the only synagogue and the Jewish community within the Left Bank Kyiv now are located on Prospekt Shukhevycha. Avraham Shmulevich, rabbi and political scientist, commented this in the following way: "If the name of Shukhevych were given to some other street, this could eventually be forgotten, and the severity of the conflict would gradually disappear. But now a permanent reminder has been created. It's like hammering a nail into the body: it will always hurt" (Obozrevatel, 14 June 2017). Antifascist Human Rights Legal League and Jewish Human Rights Watch demanded to recognize the renaming as illegal and cancel it (Obozrevatel, 14 June 2017).

However, despite these isolated cases, the facts point to the absence of total "banderization" of Ukraine: the rush occurs about individual, isolated cases. In particular, according to Volodymyr Viatrovych, the head of the Ukrainian Institute of National Remembrance, only 34 streets out of over 50,000 renamed in the framework of decomunization, were named in the honour of Bandera. As Viatrovych commented, "attempts to create a cult of Bandera under the canons of the cult of Lenin are harmful both for understanding his place and role in history and for proper commemoration. But even more harmful are attempts to invent this cult artificially and to represent decommunization as "banderization".

6. Conclusions

In Ukraine, there are three main regional strategies for renaming urban toponymy in the framework of decommunization, which reflect the goals and tactics of local governments. The essence of these strategies is to update the various pages of national history and politicize/depoliticize of toponymy. If the political elites of the western and central parts of the country position Ukraine as a direct ideological successor to the Ukrainian People Republic, the oppositionally-minded political elites of the south-east made an attempt to depoliticize the local toponymy as much as possible and propagandize Ukraine as the successor of the Cossack state (in the medium-term perspective, this renaming can indeed lead to a shift in the collective memory in the south-east). In the far east (in Donbass and in Zaporizhia oblast), the new place-names maintain a close connection with the Soviet era. Spatial patterns of these strategies are well correlated with the existing electoral differentiation and the existing ethnic composition of the Ukrainian population.

Cossack epoch is considered by the Ukrainian political elites as the most compromise and intermediary stratum of national identity, which does not cause rejection in society regardless of the political preferences and ideological beliefs. The similar role belongs to the epoch of Ukrainian Peoples Republic, which is widely represented in the toponymy throughout the country, except for the extreme east. At the same time, the Ukrainian Insurgent Army and the Soviet legacy remain topics dividing the Ukrainian society, including political elites.

The street renaming process in post-revolutionary Ukraine is both national and regional-specific; however, it shares some common characteristics with other post-Socialist European countries: 1. appeal to the pre-socialist period arouse in order to reinforce the national identity; 2. the decrease of the political and military place-names in favour of connotations associated with local and national culture and heritage; 3. increasing role of local and regional toponymy.

Acknowledgements

This research was made within the framework of scientific project No 16BP050-02 "Spatial Transformation in Ukraine: Models of Urban Modernization and Planning" funded by the Ministry of Education and Science of Ukraine.

References

Alderman, D. H. (2003): Street names and the scaling of memory: the politics of commemorating Martin Luther King, Jr within the African American community. Area 35(2), 163–173, https://doi.org/10.1111/1475-4762.00250.

Azaryahu, M., Kook, R. (2002): Mapping the nation: street names and Arab Palestinian identity: three case studies. Nations and Nationalism 8(2), 195–213, https://doi.org/10.1111/1469-8219.00046.

Azaryahu, M. (2012): Renaming the past in post-Nazi Germany: Insights into the politics of street naming in Mannheim and Potsdam. Cultural Geographies 19(3), 385–400, https://doi.org/10.1177/1474474011427267.

Azaryahu, M. (2011): The critical turn and beyond: the case of commemorative street naming. An International E-Journal for Critical Geographies 10(1), 28–33.

Balode, L. (2012): Place-names in the urbanonymy of Riga: mark of identity. Studii si Cercetari de Onomastica si Lexicologie 5(1–2), 17–28.

Barrington, L. W. (2002): Examining rival theories of demographic influences on political support: The power of regional, ethnic, and linguistic divisions in Ukraine. European Journal of Political Research 41(4), 455–491, https://doi.org/10.1111/1475-6765.00019.

Barrington, L. W., Herron, E. S. (2004): One Ukraine or many? Regionalism in Ukraine and its political consequences. Nationalities Papers 32(1), 53–86, https://doi.org/10.1080/0090599042000186179.

Begonja, Z. (2006): Odonimi kao ideološke manifestacije na primjeru Zadra u XX stoljeću. Radovi Zavoda za povijesne znanosti HAZU u Zadru 48, 703–720.

Bertoša, M. (2008): Pulsko ulično nazivlje: civilizacijski slojevi, tragovi identiteta. In: Stradarij grada Pule, Krizmanić, A. (ed.). Pula, Histria Croatica CASH, pp. 5–29.
Birch, S. (2000): Interpreting the regional effect in Ukrainian politics. Europe–Asia Studies, 52 (6), 1017–1041, https://doi.org/10.1080/09668130050143815.
Bucher, S., Matlovič, R., Lukáčová, A., Harizal, B., Matlovičová, K., Kolesárová, J., Čermákov, L., Michalko, M. (2013): The perception of identity through urban toponyms in the regional cities of Slovakia. Anthropological Notebooks 19(3), 23–40.
Diesen, G. Keane, C. (2017): The two-tiered division of Ukraine: Historical narratives in nation-building and region-building. Journal of Balkan and Near Eastern Studies 19(3), 313–329, https://doi.org/10.1080/19448953.2017.1277087.
Crlenjak, B. (2005): Razvitak vukovarskih ulica. Vukovar, Gradski muzej.
Crljenko, I. (2006): Regionalni identitet u urbanom pejsažu sjevernog hrvatskog primorja: prostorna i značenjska analiza gradske toponimije. Unpublished MA Thesis. Zagreb, University of Zagreb.
Crljenko, I. (2012): The renaming of streets and squares in post-Socialist Croatian towns. Language and Society 3, 230–241.
David, J. (2011): Smrdov, Brežněves a Rychlonožkova ulice. Kapitoly z moderní české toponymie. Praha: Academia.
David, J. (2013): Street names – between ideology and cultural heritage. Acta Onomastica 54(1), 53–61.
David, J., Mácha, P. (2014): Názvy míst. Paměť, identita a kulturní dědictví. Ostrava.
How and why ordinary streets of the Dnipro got unusual names? Dnepropetrovskaya Panorama, 29 June 2016. (In Ukrainian). Access mode: http://dnpr.com.ua/content/yak-i-chomu-zvichayni-vulici-dnipra-otrimali-nezvichni-nazvi
In Dnipro finally decided on the issue of renaming the Kalinin avenue. Tyzhden.UA, 17 December 2017. (In Ukrainian). Access mode: http://tyzhden.ua/News/205987
Gill, G. (2005): Changing symbols: the renovation of Moscow place names. The Russian Review 64, 480–503, https://doi.org/10.1111/j.1467-9434.2005.00371.x.
Gnatiuk, O. (2015): Perception of local geographical specificity by the population of Podolia. Ekonomichna ta Sotsialna Geografiya 1(71), 88–99, https://doi.org/10.17721/2413-7154/2015.71.88-99.
Gnatiuk, O. (2016): National and local patriotism in spatial dimension: case of Podolia, Ukraine. Human Geography Journal 21(2), 75–79.
Gonzalez Faraco, J. C., Murphy, M. D. (1997): Street names and political regimes in an Andalusian town. Ethnology 36(2), 123–148, https://doi.org/10.2307/3774079.
Graham, B., Ashworth, G., Tunbridge, J. (2000): A geography of heritage: power, culture and economy. London, Arnold.
Guyot, S., Seethal, C. (2007): Identity of place, places of identities, change of place names in post-apartheid South Africa. The South African Geographical Journal 89(1), 55–63, https://doi.org/10.1080/03736245.2007.9713873.
Kadmon, N. (2004): Toponymy and geopolitics: the political use – and misuse – of geographical names. The Cartographic Journal 41, 85–87, https://doi.org/10.1179/000870404X12897.
Knappová, M. (2013): Nová pražská deantroponymická urbanonyma. Acta Onomastica 54, 123–134.
Komarov, R. (2008): Hero and monument to the hero as factors of identity: an example of Donetsk. East-West: Historical and culturological collection 9–10, 91–99. (In Ukrainian).
Krizmanić, A. (2008): Razvitak uličnog nazivlja: imena ulica, trgova i parkova 2008 s prijašnjim nazivima. In: Stradarij grada Pule, Krizmanić, A. (ed.). Pula, Histria Croatica CASH, 85–356.
Light, D., Nicolae, I., Suditus, B. (2002): Toponymy and the communist city: street names in Bucharest 1948–1965. GeoJournal 56(2), 135–44.
Light, D. (2004): Street names in Bucharest, 1990–1997: exploring the modern historical geographies of post-socialist change. Journal of Historical Geography 30(1), 154–172, https://doi.org/10.1016/S0305-7488(02)00102-0.
Malešić, N. Preimenovanje ulica i trgova grada Rijeke (II. dio), [Online], Available: http:www.geografija.hr/clanci/1523/preimenovanje-ulica-i-trgova-grada-rijeke-ii-dio [Accessed 5 August 2017].
Melnychuk, A., Gnatiuk, O. (2012): Types of regional identity of the Vinnytsia inhabitants. Bulletin of Drahomanov National Pedagogical University. Series 4: Geography and Modernity 15(27), 124–13 (In Ukrainian).
Melnychuk, A., Gnatiuk, O., Rastvorova, M. (2014): Use of territorial identity markers in geographical researches. Scientific Annals of "Alexandru Ioan Cuza" University of Iasi – Geography Series 60(1), 157–184.
Mikheyeva, O. (2008): Regional identity and political dimension (the case of Donetsk). East-West 9–10, 108–114. (In Ukrainian).
Moscow Avenue of Bandera. Historians commented on the scandalous renaming. Gordon, 8 July 2016. (In Russian). Access mode: http://gordonua.com/publications/moskovskiy-prospekt-imeni-bandery-istoriki-prokommentirovali-skandalnoe-pereimenovanie-139888.html
Nagorna, L. P. (2008): Regional identity: the Ukrainian context. Kyiv, Institute of political and ethnic studies. (In Ukrainian).
Odaloš, P., Majtán, M. (1996): Urbanonymá v kontexte histórie a súčasnosti. Materiály z onomastického kolokvia (Banská Bystrica 3–5 septembra 1996). Banská Bystrica – Bratislava.
Only 34 among the 50.000 toponyms were renamed after Bandera. Ukrainian Institute of National Remembrance. (In Ukrainian). Access mode: http://www.memory.gov.ua/news/z-50-tisyach-pereimenovanikh-obektiv-toponimiki-lishe-34-buli-nazvani-na-chest-banderi
Osipian, A. L., Osipian, A. L. (2012): Regional diversity and divided memories in Ukraine: contested past as electoral resource, 2004–2010. East European Politics and Societies 26(3), 616–642, https://doi.org/10.1177/0888325412447642.
Palonen, E. (2008): The city-text in post-communist Budapest: street names, memorials, and the politics of commemoration. GeoJournal 73 (3), 219–230.
People in Dnipropetrovsk do not want to live on the streets named after Bandera and Shukhevych. DepoDnipro, 14 August 2015. (In Ukrainian). Access mode: https://dnipro.depo.ua/ukr/dnipro

/dnipropetrovtsi-ne-hochut-zhiti-na-vulitsyah-banderi-i-shuhevicha-14082015093500
Rewakowicz, M. G. (2010): Geography matters: Regionalism and identities in contemporary Ukrainian prose. Canadian-American Slavic Studies 44(1–2), 82–101, https://doi.org/10.1163/221023910X512813.
Rihtman-Augustin, D. (2000): Ulice moga grada: antropologija domaćeg terena. Beograd, Biblioteka XX vek.
Robinson, G. M., Engelstoft, S., Pobric, A. (2001): Remaking Sarajevo: Bosnian Nationalism after the Dayton Accord. Political Geography 20, 957–980, https://doi.org/10.1016/S0962-6298(01)00040-3.
Rose-Redwood, R. S., Alderman, D., Azaryahu, M. (2010): Geographies of toponymic inscription: new directions in critical place-name studies. Progress in Human Geography 34(4), 453–470, https://doi.org/10.1177/0309132509351042.
Sänger, J. (2006): Heldenkult und Heimatliebe. Straßen- und Ehrennamen im offiziellen Gedächtnis der DDR. Berlin: Ch. Links Verlag.
Shmulevich, A. (2017): Shukhevych Avenue – one-hundred-percent-hit. Obozrevatel, 19 June 2017. (In Russian). Access mode: https://www.obozrevatel.com/society/78540-prospekt-shuhevicha-stoprotsentnoe-popadanie.htm
Stanić, J., Šakaja, L., Slavuj, L. (2009): Preimenovanja zagrebačkih ulica i trgova. Migracijske i etničke teme, 25, (1–2), 89–124.
Stiperski, Z., Lorber, L., Heršak, E., Ptaček, P., Górka, Z., Kołoś, A., Lončar, J., Faričić, J, Miličević, M., Vujaković, A., Hruška, A. (2015): Identity through urban nomenclature: eight Central European cities. Geografisk Tidsskrift – Danish Journal of Geography 111(2), 181–194.
Streets of Dnipro. Secrets of the names. DneprChas, 17 June 2016. (In Russian). Access mode: https://dpchas.com.ua/dnepropetrovskaya-solyanka/ulicy-dnepra-sekrety-nazvaniy
Yeoh, B. (1992): Street names in colonial Singapore. Geographical Review 82, 313–322, https://doi.org/10.2307/215354.
Yeoh, B. (1996): Street-naming and nation-building: toponymic inscriptions of nationhood in Singapore. Area 28, 298–307.

Fertility factors in Czechia according to the results of the 2011 census

Roman Kurkin[1], Branislav Šprocha[2], Luděk Šídlo[3,*], Jiřina Kocourková[3]

[1] Czech Statistical Office, Prague, Czechia
[2] Center of Social and Psychological Sciences SAS, INFOSTAT – Research Demographic Center, Slovakia
[3] Charles University, Prague, Faculty of Science, Department of Demography and Geodemography, Czechia
* Corresponding author: ludek.sidlo@natur.cuni.cz

ABSTRACT

In the last quarter century, women's reproductive behaviour underwent a marked change in Czechia. The 2011 census showed a fall in cohort fertility below two children per woman by the end of her reproductive span. Factors behind changes in fertility, particularly from a cohort perspective, have not been sufficiently analyzed. The aim of this article is to determine the main factors influencing cohort fertility in Czechia. The main objective is to test the impact of the most frequently discussed factors of fertility at the individual level. The analysis is based on anonymised individual-level data from the 2011 Czech Population and Housing Census. We used the method of causal modelling to monitor the impact of various factors on cohort fertility. It was confirmed, that the key factor behind fertility levels was the marital status (married women are more likely to become mothers than single women). Other important factors included woman's income (a higher income raises the chance of remaining childless or having only one child) and achieved educational level (the level of childlessness increases as the level of education rises). Future fertility rates in Czechia will depend mainly on the extent to which university educated women will be capable to reconcile work and family life in order to fulfil their reproductive ambitions. By analysing the differences in cohort fertility among various population groups and identifying the factors that may affect fertility levels across these subpopulations, it is possible to gain a better understanding of the mechanisms behind changes of reproductive behaviour.

KEYWORDS

cohort fertility; fertility factors; Czechia; population census; logistic regression

Received: 12 June 2017
Accepted: 28 March 2018
Published online: 22 June 2018

Kurkin, R., Šprocha, B., Šídlo, L., Kocourková, J. (2018): Fertility factors in Czechia according to the results of the 2011 census. AUC Geographica 53(2), 137–148
https://doi.org/10.14712/23361980.2018.14

1. Introduction

The marked changes in reproductive behaviour seen among women in Czechia over the last twenty-five years have been attributed to economic transformation and value changes associated with the second demographic transition (e.g. Kučera and Fialová 1996; Fialová and Kučera 1997; Rychtaříková 2000; Rabušic 2001; Sobotka, Zeman and Kantorová 2003; Sobotka et al. 2008). In the early 1990s Czechia was witness to a dramatic fall in total fertility as a result of fertility postponement (Sobotka 2003, 2004a, 2004b) and diversifying reproductive behaviour (Kantorová 2004; Sobotka 2004a). The previous reproductive model was no longer compatible with the new conditions, and it was rapidly abandoned by cohorts of young women (born towards the end of the 1960s, see Sobotka et al. 2008; Šprocha 2014). Czechia (and former GDR, Hungary and Slovenia) was one of the former Eastern Bloc countries in which the response to the changing conditions brought about by the end of communism was quite fast (Sobotka 2003, 2004a, 2011). A relatively large amount of attention has been devoted to analysing the changes in reproductive behaviour in Czechia (e.g. Rychtaříková 2000, 2010; Sobotka, Zeman and Kantorová 2003; Sobotka et al. 2008), while less attention has been paid to the factors behind fertility, probably owing to difficulties in accessing the data and more complicated methodologies used to analyse them.

Since the knowledge of conditionalities behind the reproductive intentions of women and couples plays a key role for decision-making processes (e.g. for setting family policies), the main aim of this article is to test the impact of selected main factors (marital status, income and education) influencing cohort fertility in Czechia in connection with the following three basic hypotheses:

H1) A married woman is more likely than a single woman to become a mother at least once, and she is also more likely to give birth to more than two children.

H2) As the woman's income rises so too does the chance of her remaining childless or of having only one child.

H3) As education levels increase, the childless rate rises and the chance of having more than two children decreases.

2. Theoretical framework

There are discussed many types of factors of fertility in the literature. The extent to which they exert an influence depends on the population, the social and economic situation, historical traditions and the internal structure of the country's population, for example, economic activity, the position of women on the labour market, education, and so on (for Czechia see e.g. Klasen and Launov 2006; Kostelecký and Vobecká 2009; Šídlo 2008).

The last global financial crisis (2008–), which led among other things to a deterioration in the situation for young people on the labour market (European Commission 2013), has reignited discussion on fertility factors (particularly structural ones) (e.g. Sobotka et al. 2011; Goldstein et al. 2013) and has been accompanied by studies investigating the effects of insecurity on the labour market and employment (e.g. Pailhé and Solaz 2012; Matysiak and Vignoli 2013). Important characteristics include level of active *participation on the formal labour market* and its inherent character (full-time, part-time employment, and so forth). Given differences in traditional gender conceptions, these factors are also analysed from a gender perspective. The specific character of the country also plays a role, particularly in relation to the nature of the labour market, social and family policies, dichotomous perceptions on the status of men and women in society and on the labour market, and so on. Over time these shape a country's long-term historical and political development (Lundström and Andersson 2012). Economic activity and the labour market are closely connected to *income levels*, which are another important factor in explaining the variables behind fertility. Economic theory holds that higher incomes lead to an increasing concern for the quality of children's lives as against the quantity of children (Becker 1960). Alongside gendered aspects of the labour market and family environment, income levels may lead to differences in fertility levels. Higher incomes have two different effects. They increase the demand for children and hence may positively influence fertility (the income effect). But they can also increase the cost of time spent with children and so may negatively impact on fertility (the opportunity cost of having a child; Cette, Dromel and Méda 2007; Pailhé and Solaz 2012). The opposite effect may be seen in those on low incomes – demand falls (income is spent on essential items) and reproduction is postponed (income effect). On the other hand the cost of time spent with children falls (cost effect), which may have a positive impact on fertility. Here *state family and social policies* are important. The United Kingdom is a typical example of a country with a historically large difference in fertility levels in relation to income, which is then exacerbated by family (and social) policies targeted at low-income families (Sigle-Rushton 2009). In traditional breadwinner countries, the man is usually responsible for bringing in the majority of the family's income and so his position on the labour market will have a greater impact on fertility than that of the woman. The income effect will have a greater impact than the cost effect, and male employment is therefore positively associated with fertility (Pailhé and Solaz 2012). By contrast in societies where the man is seen as the breadwinner, female employment has a demonstrably negative effect on fertility levels (for example, Italy; see Matysiak and Vignoli 2013). In these countries the absence of such policies is considered to be one of the reasons for lower fertility levels (Neels and De Wachter 2010). However, some studies of former Eastern Bloc countries (e.g. Matysiak and Vignoli 2013), which also experience problems owing to harmonisation of childcare and work, have not clearly identified *unemployment* as a negative factor in fertility. Czechia already had a high level of female employment under socialism (Kantorová 2004) in addition to intricate links between the labour market and family sphere (Ettlerová and Šťastná 2006); however, it has been confirmed that female employment has a negative impact on fertility levels and on the postponement of childbearing (Klasen and Launov 2006). Women in full-time work on open-ended contracts are less likely to have dependent children than women in part-time work (Kurtinová 2015).

In northern countries with high levels of gender equality, a woman's *income* correlates positively to fertility level, while her partner's income has less impact (Neyer 2009).The negative effect of insecurity on the labour market is associated with delayed reproduction in Sweden, as confirmed in research by Lundström and Andersson (2012).The situation in Belgium is similar, where unemployed women have a lower chance of entering into motherhood than working women. The difference is particularly evident in university educated women who tend to wait for better positions on the labour market (Neels and De Wachter 2010). According to research by Pailhé and Solaz (2012) the situation differs to some extent in France, where insecurity on the labour market does not have such a great effect on female fertility. However, there is still a tendency for women to postpone first order births because of insecure employment or short-term employment contracts. It is thought that generous family state benefits and high rates of unemployment benefit are behind the reduced impact of labour market insecurity on the timing of first order births and cohort fertility in France. Other research has also shown that the link between female unemployment and the timing and intensity of fertility is ambiguous. It has been demonstrated that in the former GDR, unemployment among women born after the fall of the Berlin Wall correlates positively to first order births, but male unemployment has no significant effect (Özcan, Mayer, Luedicke 2010). Data from the 2005 German "Gender and Generation Study" indicate that labour market insecurity among men living in the former GDR has a negative influence on childbearing intentions, but this relationship was not found among women (Berninger, Weib, Wagner 2011).In Italy couples where the woman does not work have a greater chance of having at least one child than those in which the woman is employed. Hence the traditional male breadwinner social model increases the chance of the woman becoming a mother (Santarelli 2011). A study by Kalwij (2010) attempted to obtain a more comprehensive view of the relationship between female employment and fertility in its investigation of 16 former Eastern Bloc countries. However, it did not succeed in finding any more substantial links either.

Position on the labour market and income level are closely linked to *educational attainment achieved*. In postmodern societies this is one of the main differentiating factors influencing reproductive behaviour. It links a series of factors that has direct and indirect impact on reproduction. They are length of study, differences in social and cultural capital, value orientations, and so forth. The direct impact education has on fertility and especially timing relates to *length of study*. Studying is generally seen as incompatible with establishing a family (Baizán, Aassve and Billari 2003; Blossfeld and Huinink 1991; Kravdal 1994). This has also been confirmed by some studies in Czechia (Kantorová 2004; Šťastná 2009). Extended education is seen to delay economic independence and thereby entry into adulthood (e.g. Kohler, Billari and Ortega 2002). This direct impact is then augmented by socioeconomic situation, values, preferences and post-study opportunities, which exert an influence on entry into motherhood because more highly educated women often try to gain stable employment, financial security or good housing etc. before establishing a family (Sobotka et al. 2008). Becker's (1960) classical economic theory holds that there is a link between higher education and higher income levels, and the associated higher opportunity costs of non-participation in the labour market. Hence the relationship between educational level and fertility rate is negative. It has a more significant impact in societies where the female/male roles are strongly differentiated and in areas where harmonising work and family roles is difficult (Liefbroer and Corijn 1999; Rychtaříková 2004). The economic theory of fertility posits that women with higher levels of education postpone fertility and childbearing (Brolcháin and Beaujouan 2012) because the opportunity costs are higher for women at the beginning of their career. The negative relationship between education level and fertility rate has been confirmed in Europe and Czechia by a number of studies (e.g. Kalwij 2010; Klasen and Launov 2006, Rychtaříková 2004). However, it does not hold everywhere because in Belgium, for instance, the income effect has a greater impact, and among the generations of women born after 1945 a positive relationship has been found between education level and cohort fertility (Neel a De Wachter 2010). Similarly a higher level of education is linked to a greater number of children among men (e.g. Pailhé and Solaz 2012; Lappegård and Rønsen 2013) because the lost opportunity costs are not as high for them (Bartus et al. 2013). However, this relationship cannot be considered to be universal either as a study in the Netherlands and Flanders in Belgium by Liefbroer and Corijn (1999) has shown.

In the majority of cases the effect of *marital status* on fertility rate has been interpreted as unambiguous. People living as married couples have a higher chance of giving birth to a larger number of children than those in unmarried couples (e. g. Pailhé and Solaz 2012, Neels and De Wachter 2010; Hoem, Jalovaara and Mureşan2013). The results of cohort fertility studies conducted in Czechia have also confirmed this over the long term (e.g. Chromková-Manea and Rabušic 2013).

3. Data and methodology

Anonymised individual-level data from the 2011 Population and Housing Census can be used to link information on the number of live births born to a woman with other characteristics. Causal modelling, specifically binary and multinomial logistic regression, was used to determine which factors affect cohort fertility levels.

Binary logistic regression was used for first model, and the dependent variable was a live birth – either the mother had thus far not given birth to a child or had given birth to at least one. The multinomial logistic regression used in second model made it possible to expand the dependent variable to include additional categories (0, 1, 2, 3 and 4 or more children). The reference category used in this model was two children. The data set consisted of women aged 45–49 and the independent variables were women's marital status (ref. married at least once), level of income (ref. high income) and educational attainment (ref. tertiary education).

The choice of independent variables was made to ensure the characteristics were relatively stable over time. Marital status was therefore aggregated into two categories – single, and married at least once (including married, divorced and widowed). Educational attainment can be considered a predominantly stable characteristic from age group 25–29 and up. Four basic education groups entered into the models: lower secondary or lower, secondary, upper secondary, tertiary.

The main categories of income level were aggregated into four categories: 'undetermined', low, middle and high income formed out of the ten main categories of employment on the basis of estimated median gross monthly income (see tab. 1). Category 'undetermined' consisted of economically inactive and unemployed seeking for their first job. Among economically inactive women are not working pensioners (including disabled pensioners), persons at home or other dependent persons and others with own source of livelihood. So this category is very heterogeneous and their income level as well as reproduction behaviour is hard to estimate.

The outcome of the binary logistic regression is the odds ratio (in table Exp (B)) of having at least one

Tab. 1 Main classes of occupations by category of income and median gross monthly salary in 2011.

Main classes of occupations (CZ-ISCO-08)	Category by income	Median gross monthly salary (in CZK)
Armed Forces	higher income	24,123
Managers	higher income	39,966
Professionals	higher income	28,928
Technicians and associate professionals	higher income	25,486
Clerical support workers	middle income	20,554
Service and sales workers	low income	14,401
Skilled agricultural, forestry and fishery workers	low income	16,936
Crafts and related trades workers	middle income	20,335
Workers Plant and machine operators, and assemblers	middle income	20,130
Elementary occupations	low income	13,346

Data source: Statistická ročenka České republiky 2013.

child versus having none for independent variable category in relation to the reference category, assuming that the values of the other independent variables do not change. The multinomial logistic regression shows the odds ratio of having a specific number of live births versus the reference number of children (two). The independent variables were entered into both types of logistical regression (using the enter method) and the interaction among the independent variables was not included in the model.

The data set from the population census could have been taken as the basic set, however some women from the model were excluded on the grounds that they did not answered to some of the questions from which the indicator used in the model was derived. The proportion of women excluded from the model containing all the female age categories was 11.9%. In the five-year age categories analysed the proportion was between 11.4% and 16.5%, and declined as age increased.

Because of the number of women excluded from the model, the reliability and explanatory potential of the model was tested using several methods. The chi-square statistic and its significance level (Sig. of chi-square model) can be used to determine whether the independent variables significantly contribute to the model. This occurs when the null hypothesis that the regression coefficients take the value zero is refuted. Next the model is tested to calculate the proportion of variability explained by the dependent variable using independent variables. In binary logistic regression pseudo Nagelkerke R^2 is used, which corrects the Cox and Snell pseudo R^2 to the maximum value of one (Řeháková 2000). The final method used is the classification table and it evaluates the quality of the model. It involves classifying the binary dependent variable into a four-field classification table according to the number of observed and predicted values. A high proportion of correctly classified variables on the main diagonal confirm the fit of the model.

In addition to evaluating the quality of the model, we also evaluated whether the categories of independent variables were significant for explaining the model. For this purpose the Wald test was used. If the null hypothesis, that the regression coefficient is null, is refuted (at 1% or 5% significance level), then the independent variables significantly contribute to explaining the dependent variable.

4. Changes in cohort fertility and a differentiated analysis of cohort fertility

Significant changes in fertility timing associated with postponed motherhood can be seen when looking at the female population structure in younger age according to number of children and therefore also in cohort fertility. A sharp fall is evident in the lower age category. While in 1991 a mother aged 20–24 had on average 0.71 children, ten years later this had fallen to 0.30 children per woman in this age group and in 2011 it was only 0.16 children. The differences were less marked among the older age groups because these had subsequently caught up on their postponed reproduction.

Given the very low fertility rate in the 40 and over age group, the level of cohort fertility can be considered to be almost completed. In 2001 the mean number of children per woman aged 40–44 was still more than two children (2.05), but by the last census, conducted in 2011, cohort fertility had fallen substantially below this level (to 1.87 children) (Figure 1).

Different generations of women lived out their reproduction spans in different eras, and these were

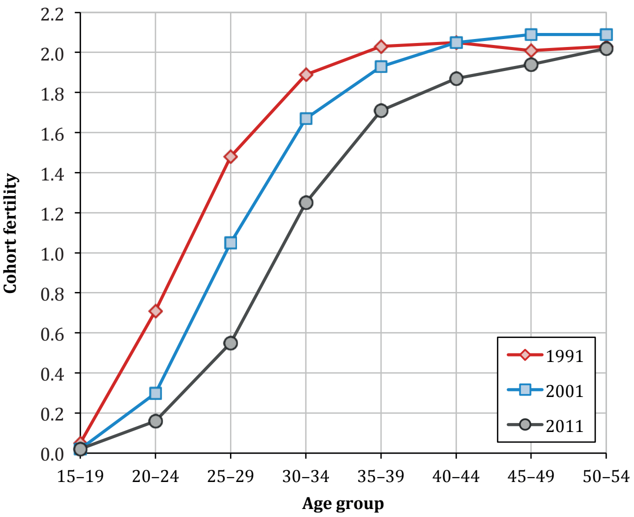

Fig. 1 Cohort fertility among women in 1991, 2001 and 2011.
Data source: Population and Housing Censuses 1991, 2001 and 2011.

subjected to a range of factors that affect completed fertility. These include economic, value, cultural and political determinants (Kurkin 2010). Completed fertility is a relatively stable indicator and so its values did not change dramatically.

Completed fertility among the generations born during the Second World War was around 2.05 children (Figure 2). The exception was the higher rate seen among the generations born in 1939 and 1940, most likely a result of the pro-population measures promised in 1963 and 1964, but that never in fact materialised.

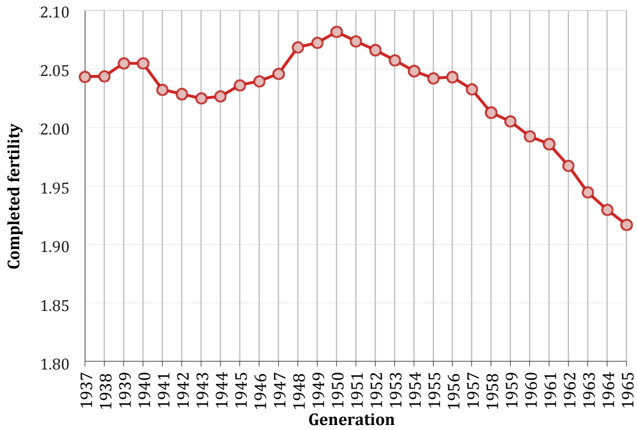

Fig. 2 Completed fertility by female generation in 2011.
Note: Mean number of children per woman who had an identified number of children for the female generations with completed reproduction in the Population and Housing Census 2011.
Data source: Population and Housing Census 2011.

Completed fertility among the generations of women born between 1945 and 1950 increased by 0.05 children per woman. This increase was a result of pro-population measures introduced in the 1970s, for example advantageous loans for newlyweds, better access to housing for families with children, higher child benefit (mainly for second and third order children), extended maternity leave and the introduction of parental benefits (Frejka 1980). The generations born from the 1950s and on experienced a decline in completed fertility as many of the pro-population measures ceased to be of value, and access to housing was still poor. The generations of women born in 1960 and later had fewer than two children on average.

The results of the 2011 Census confirm that in Czechia the two-child family model that was typical of the socialist era continues to dominate (Rychtaříková 2004). More than half of all women aged 35–39 to 70–74 had two children. Among the older age categories the two-child family model was more frequent, but because of the larger share of women with three or more children, the proportion falls to just under half – between 40 and 49%.

It is possible to calculate the parity progression ratios from the data generated from the census for women with completed reproduction. The results

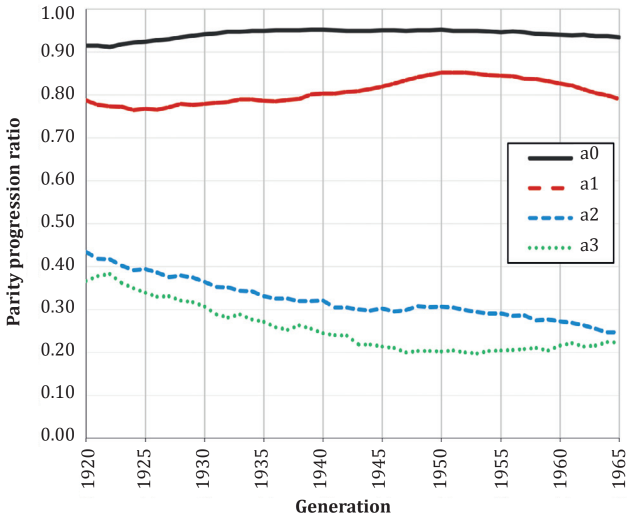

Fig. 3 Parity progression ratios by female generation in 2011.
Data source: Population and Housing Census 2011, authors' calculations.

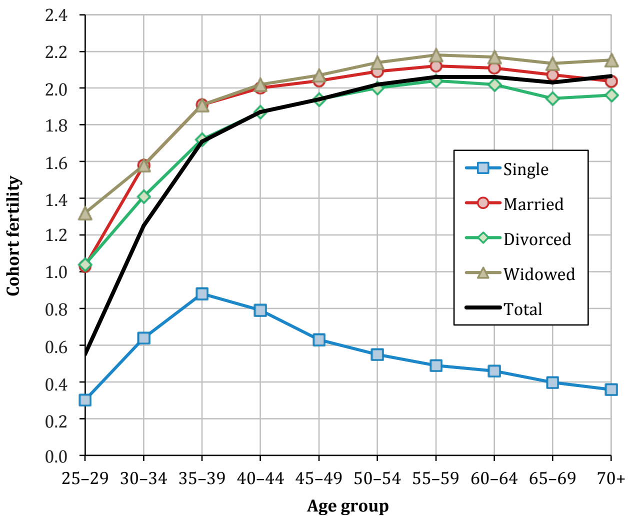

Fig. 4 Female cohort fertility by marital status and age group in 2011.
Data source: Population and Housing Census 2011.

indicate that childlessness among women born in Czechia between 1920–1985 was not common as for other countries from former West bloc (Rowland 1998). The probability of at least one child (a0) being born to women from the generations conceived between1920 and 1965 is always more than 0.90 (Figure 3), and around 0.95 among those born between 1932 and 1956.

The proportion of childless women has not risen substantially in the younger generations. The probability of having at least two children (a1) was always higher than 0.75, while the probability of having a third and fourth child (a2 and a3 respectively) fell substantially until the generation of women born in 1945, when it stabilised, and in the 1955 and subsequent generations the probability of having a third child fell, while the probability of having a fourth child grew slightly.

4.1 Cohort fertility by marital status

In Czech society entering into wedlock was traditionally linked to reproduction, and this is confirmed by the higher cohort fertility of at least one child among younger married woman (approximately under 40). It i s generally the case that married women achieved greater fertility and that this is particularly true of older widowed women. Divorce rates are highest among Czech women aged 30–39, and failed marriages may significantly affect women's reproduction plans, by shortening the timespan within which women are at greater risk of becoming pregnant, thus reducing completed fertility levels. Hence, fertility rates among divorced women are slightly lower in the post-reproductive age group (Figure 4).

Single women quite clearly have the lowest level of fertility. Interestingly, though, fertility levels among women aged 35–39 are substantially higher than among the older cohorts nearing or at the end of their reproductive lives. This is indirect evidence of the shift away from marriage and reproductions being strongly bound together, as a greater proportion of children are being born outside wedlock (almost 50% in 2015). Reproduction among single women is gradually becoming a feature of the increasingly diverse reproduction strategies and is no longer a marginal phenomenon.

4.2 Cohort fertility by educational attainment and main classes of occupations

The growing proportion of women with a upper secondary (with leaving certificate) and tertiary degree means that cohort fertility may tend to reflect the different intensity and character of reproductive behaviour of these sub-populations based on education attainment. Between 1991 and 2011, the number of women aged 25–49 with tertiary education rose from 208.1 thousand to 502.4 thousand (or from 6.0% to 13.1%).

The results of the 2011 Census show that female cohort fertility falls in almost all age groups as education levels rise (Figure 5). In the youngest group of women whose reproductive span ended between the ages of 45 to 49, cohort fertility among university educated women was 1.76 children per woman, while among women with lower secondary education or lower it was 25% higher at 2.22 children. Women with secondary without certificate were the only other group in this age category to have more than two children per woman. In the older age groups, the differences in cohort fertility between the most and least educated were more entrenched. Fertility in women aged 70 and over with lower secondary education or lower was one and a half times higher than among women with higher education. Attaining a higher education (particularly to degree level) was a privilege that distinguished groups of women through specific characteristics including lower levels of fertility.

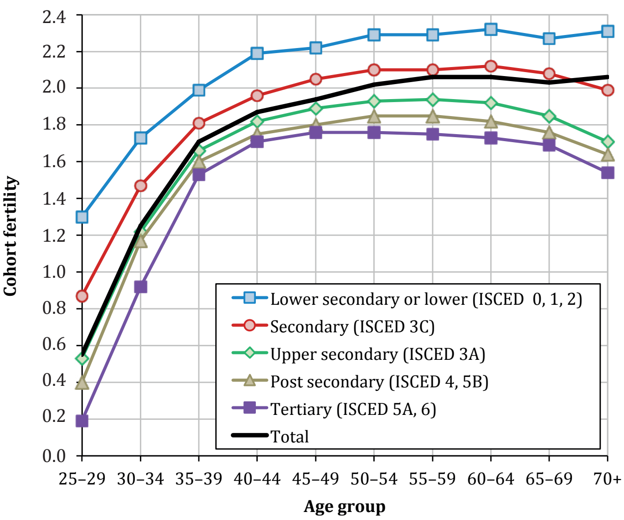

Fig. 5 Cohort fertility by educational attainment in 2011.
Note: Categories are based on ISCED-97.
Data source: Population and Housing Census 2011.

The differences in cohort fertility among women with incomplete reproduction are greater between the educated groups because they often manifest themselves in differences in the timing of childbearing. For instance in the 25–29 age group, cohort fertility among women with lower secondary education or lower was 1.30, while for the same age group of women with upper secondary level it was 0.53 and among female with tertiary education it was 0.19 children per woman.

The differences in female cohort fertility according to main classes of occupations were less marked than by levels of education. The lowest fertility levels in the youngest age group to have completed reproduction (45–49 years) were achieved by managers, professionals (e.g. lawyers, teachers, scientists, doctors, managers) and technicians and associate professionals (up to 1.85 children per woman). These are the first three main classes of occupations. By contrast elementary occupations (no. 9) and skilled agricultural workers (no. 6, Table 1) had rates high above the basic reproductive level (2.10 children per woman).

Women employed in the first three of the main classes of occupations had the lowest level of fertility even among the oldest age group, the 50+ category. By contrast those employed in categories 9 (elementary occupations) and 6 (*skilled agricultural workers*) had the highest fertility levels. The differences between the groups of women with lower and higher fertility levels increased in the older age groups. For example, in the 45–49 age group the highest level of fertility was 24% higher than the lowest, while among women aged 70 and over it was 45% higher.

Marked differences in cohort fertility by main classes of occupations were identified in the youngest age groups owing to differences in the timing of births. For instance, in the 25–29 age group cohort fertility among elementary occupations reached 0.71 children per woman, while among executives and managers it was 0.20 children.

Tab. 2 Cohort fertility by main classes of occupations in 2011.

Age group	1	2	3	4	5	6	7	8	9	88	99	Total
15–19	0.05	0.02	0.02	0.01	0.01	0.06	0.03	0.03	0.04	0.02	0.08	0.02
20–24	0.06	0.05	0.06	0.05	0.10	0.13	0.13	0.13	0.17	0.19	0.42	0.16
25–29	0.20	0.21	0.25	0.27	0.43	0.60	0.55	0.55	0.71	1.12	0.86	0.55
30–34	0.75	0.89	0.97	0.99	1.15	1.38	1.28	1.29	1.44	1.70	1.37	1.25
35–39	1.41	1.56	1.58	1.59	1.66	1.90	1.74	1.74	1.83	2.00	1.78	1.71
40–44	1.70	1.77	1.77	1.79	1.87	2.10	1.92	1.95	2.01	2.05	2.01	1.87
45–49	1.82	1.84	1.85	1.88	1.99	2.26	2.05	2.08	2.12	1.94	2.09	1.94
50–54	1.88	1.87	1.91	1.95	2.06	2.40	2.14	2.16	2.21	2.01	2.14	2.02
55–59	1.87	1.83	1.87	1.87	2.00	2.28	2.01	2.02	2.11	2.24	2.03	2.06
60–64	1.76	1.76	1.79	1.85	2.02	2.49	2.02	2.02	2.23	2.09	2.07	2.06
65–69	1.77	1.73	1.77	1.83	1.98	2.50	1.96	2.08	2.15	2.04	1.98	2.03
70+	1.85	1.61	1.71	1.84	1.92	2.34	1.99	1.89	2.08	2.07	1.93	2.06
Not identified	1.15	1.40	1.35	1.47	1.35	2.14	1.67	1.47	1.75	1.24	1.60	1.33
Total	1.41	1.34	1.40	1.28	1.47	1.93	1.64	1.66	1.90	1.69	1.56	1.58
No. of women in thousands	96	444	479	187	421	21	101	163	118	2 309	254	4 602

Note: Main classes of occupations (CZ-ISCO-08): 1 – Managers; 2 – Professionals; 3 – Technicians and associate professionals; 4 – Clerical support workers; 5 – Service and sales workers; 6 – Skilled agricultural, forestry and fishery workers; 7 – Craft and related trades workers; 8 – Plant and machine operators, and assemblers; 9 – Elementary occupations; 88 – Economically inactive and those seeking first work; 99 – Not identified. Armed forces occupations are not listed because the total number of service women aged 15 and over was around 2,000.
Data source: Population and Housing Census 2011.

4.3 Factors of cohort fertility

The significance level of the chi square in first model was zero, so the independent variables significantly contributed to explaining the dependent variable. Model 1 contributed to explaining variability at level 0.352. The proportion of dependent variable values correctly attributed in the classification table was more than 95%. According to the Wald test the categories of almost all the independent variables were significant at 1% significance level; only mothers on low and middle incomes did not contribute substantially to the model (see tab. 3).

Single women aged 45–49 had a 0.020 times lower chance of having at least one live birth than women married at least once (see tab. 3). The odds ratio of having at least one child is substantially higher for women on higher incomes than women whose main classes of income were ‘undetermined’. On the other side, according to the Wald test category mothers on low income was no significant and category middle income was at 5% significance level. In this point of view, the income effect on completed fertility of women at the end of the reproductive span seems to be relatively small.

In model 1 women with secondary education had the highest odds ratio of having a child (1.814) compared to women with tertiary education. Women with an upper secondary school had slightly less chance (1.428), followed by women with the lowest education level.

Tab. 3 Binary logistic regression, number of live births, model 1.

Independent variable	Binary dependent variable at least one child vs. none Exp (B)
Marital status	
Single	0.020**
Married at least once	1
Income level	
Undetermined	0.500**
Low income	0.964
Middle income	0.934*
High income	1
Educational attainment	
Lower secondary or lower	1.322**
Secondary	1.814**
Upper secondary	1.428**
Tertiary education	1
Constant	26.806**
Tests	
Sig. Chi square of model	0.000
R^2 (Nagelkerke)	0.352
Classification table (in %)	95.3

Note:* at 5% significance level, ** at 1% significance level.
Date source: Population and Housing Census 2011, author’s calculations using SPSS 16.0.

As in model 1, the independent variables in the multinomial logistic regression model significantly contributed to the model (significance level was zero). The proportion of explained variability was almost 0.2, which is lower than in model 1. In model 2 the proportion of correctly attributed dependent variables in the classification model was also lower (60.1%). The Wald test showed that some independent variables (upper secondary, middle income) did not contribute significantly in the model (to 1% of significance level) (see tab. 4).

The odds ratio of a single woman being childless or having one child as opposed to two children was substantially higher than among women married at least once (see tab. 4). The chance of having three rather than two children was higher among women married at least once, while the odds ratio of having four or more children versus reference two children was higher among single women.

Tab. 4 Multinomial logistic regression, number of live births, model 2.

Independent variables	Nominal dependent variable – number of children (ref. = 2 children)			
	0 vs. 2	1 vs. 2	3 vs. 2	4+ vs. 2
Marital status				
Single	137.929**	10.350**	0.822**	1.678**
Married at least once	1	1	1	1
Income level				
Undetermined	2.590**	1.461**	1.674**	3.888**
Low income	1.096**	1.042**	1.332**	1.674**
Middle income	1.074*	0.931**	1.273**	1.344**
High income	1	1	1	1
Educational attainment				
Lower secondary or lower	0.862**	0.759**	2.216**	4.488**
Secondary	0.496**	0.645**	1.353**	1.346**
Upper secondary	0.656**	0.833**	1.062**	0.990
Tertiary education	1	1	1	1
Tests				
Sig. chi square of model	0.000			
R^2 (Nagelkerke)	0.199			
Classification table (in %)	60.1			

Note: * to 5% of significance level, ** to 1% of significance level.
Data source: Population and Housing Census 2011, authors’ calculations using SPSS 16.0.

Women whose main classes of occupations were ‘undetermined’ had the highest odds ratio of having three, four or more children as opposed to two than women in the other occupations categories. This was followed by women on low and middle incomes for whom the odds ratio was slightly lower, and this tended to be lowest among women on higher incomes. The odds ratio of remaining childless as opposed to having the reference two children was higher among women whose main classes of occupations was 'undetermined'. In all cases the chance of having three or four or more children versus two fell among women with a higher level of education.

The chance of being childless or having one child as opposed to the reference two rose through the education levels starting from secondary school education to higher education. Women with lower secondary school education or lower were again the exception because they were more likely to be childless or have only one child compared to women with secondary or upper secondary education (see tab. 3).

5. Discussion

The study confirmed that marital status is a key factor in fertility in Czechia. In both models the variable was significant at the 5% significance level and women who had married at least once had a substantially greater chance of having a child than not. The chance of having a higher number of children (3 or 4+) as opposed to the reference two children was higher in case of three children and lower for four or more children. The effect of marital status of woman on fertility depends on the country context and on the prevalence of non-marital cohabitation (Kuhnt and Trappe 2013). While being married has only a modest positive effect on fertility in France (Toulemon and Testa 2005), the impact is sizeable in the Netherlands (Balbo and Mills 2011). Hypothesis H1 can therefore be confirmed.

Woman's socioeconomic situation was not found to be very important determinant of fertility at the end of the reproductive age. The chance of a woman having a child as opposed to not having a child falls as income increases, as hypothesis H2 states was not the case among the 45–49 age group. On contrary, the odds ratio of remaining childless as opposed to having the reference two children in multinomial logistic regression was higher among women whose main classes of occupations was 'undetermined' This is in line with findings of Spéder and Kapitány (2009) that the likelihood to have a child in Hungary declines sharply when women or her partner is unemployed. Findings of other studies also confirm that among childless women those with a higher income are more likely to give birth whereas the unemployed are prevented from realizing their childbearing intentions (Berrington 2004). It was shown that as the woman's income level increased she had less chance of having a greater number of children than the reference two in Czechia. Hypothesis H2 can therefore be considered to be only partially confirmed.

The effect of education depends on other variables in the model. If no income variables are included, then it can mediate "income effect" (Spéder 2010). Accordingly, mechanisms discussed on the effect of income variable work similarly for the education effect (Quesnel-Vallée and Morgan 2003). The odds ratio of a woman giving birth to a child falls as the highest level of education increases among women, beginning with those with secondary school education but no leaving certificate and rising to those with tertiary education. The group of women with the lowest level of education (lower secondary or lower) was third highest. The chance of having a greater number of children (3 or 4 and more) as opposed to two children falls in relation to increasing level of education. Similarly to the previous hypothesis, H3 can be confirmed if we overlook the exceptional case of woman with the lowest level of education that were at both extremes – they tend to be childless more frequently than other women and also have the highest number of children. Both models confirmed the statistically significant effect of the educational attainment on the number of live births. The research findings regarding education have been ambiguous and seem to point to the relevance of country context (Kuhnt and Trappe 2013). It appears that women with higher educational levels are more likely to realize their fertility intention in France (Toulemon and Testa 2005) which suggest the importance of family policies to support work-life balance. France has a long history of policies that encourage fertility, including extensive preschool daycare facilities to favour the work-life balance (Régnier-Loilier and Vignoli 2011). On contrary, in United Kingdom educational differences in cohort fertility are significant and suggest the strong educational gradient in fertility among British women (Berrington et al. 2015). Completed fertility is smaller for higher educated women because a higher proportion of women with tertiary education remain childless or have fewer children than low-educated women. As family policies in United Kingdom are concentrated on low-income families higher educated women are not supported in work-life balance by family policies like in France.

6. Conclusion

Fall in cohort fertility among the generation of women with completed reproduction spans to under two children per woman, signalling the importance of understanding fertility differences in Czechia. Analysing differences in cohort fertility values among different population groups and identifying factors that may influence fertility levels in these sub-populations enables us to better understand the mechanisms behind reproductive behaviour. These findings can then be used to modify policies aimed at reducing, where possible, any subsequent falls beyond this figure among future generations (Šprocha 2014).

Given changes in fertility levels and the analysis above, one can postulate that the future fertility rate in Czechia will strongly depend on how women with higher education, assuming they increase as a proportion of the population, are able to harmonise work and family life so they can fulfil their reproductive ambitions.

Acknowledgements

This article as written with the support of GAČR grant (no. 15-09443S) "Risks associated with the postponement of childbearing: A new role for family policies?".

References

Baizán, P., Aassve, A., Billari, F. C. (2003): Cohabitation, marriage, and first birth: The interrelationship of family formation events in Spain. European Journal of Population 19(2), 147–169, https://doi.org/10.1023/A:1023343001627.
Balbo, N., Mills, M. (2011): The influence of the family network on the realisation of fertility intentions. Vienna Yearbook of Population Research, 179–205.
Bartus, T., Murinkó, L., Szalma, I., Szél, B. (2013): The effect of education on second births in Hungary: A test of the time-squeeze, self-selection, and partner-effect hypotheses. Demographic Research 28(1), 1–32, https://doi.org/10.4054/DemRes.2013.28.1.
Becker, G. S. (1960): An economic analysis of fertility. In: National Bureau of Economic Research (ed.). Demographic and Economic Change in Developed Countries. Princeton: Princeton University Press, pp. 209–231 (NBER Conference Series vol. 11).
Berrington, A. (2004): Perpetual postponers? Women's, men's and couple's fertility intentions and subsequent fertility behaviour. Population trends, 117, 9–19.
Berrington, A., Stone, J., Beaujouan, E. (2015): Educational differences in timing and quantum of childbearing in Britain: A study of cohorts born 1940–1969. Demographic Research 33, 733–764, https://doi.org/10.4054/DemRes.2015.33.26.
Berninger, I., Weiß, B., Wagner, M. (2011): On the links between employment, partnership quality, and the desire to have a first child: The case of West Germany. Demographic Research 24(24), 579–610, https://doi.org/10.4054/DemRes.2011.24.24.
Bhrolcháin, M. N., Beaujouan, É. (2012): Fertility postponement is largely due to rising educational enrolment. Population Studies 66(3), 311–327, https://doi.org/10.1080/00324728.2012.697569.
Blossfeld, H. P., Huinink, J. (1991): Human capital investments or norms of role transition? How women's schooling and career affect the process of family formation. American Journal of Sociology 97(1), 143–168, https://doi.org/10.1086/229743.
Cette, G., Dromel, N., Méda, D. (2007): Opportunity costs of having a child, financial constraints and fertility. Applied Economics Letter 14(4), 239–244, https://doi.org/10.1080/13504850500425535.
Ettlerová, S., Šťastná, A. (2006): Harmonizace rodinných a pracovních povinností rodičů se závislými dětmi. Demografie 48(1), 12–21.
European Commision (2014): EU measures to tackle young unemployment [electronic resource]. Available at http://europa.eu/rapid/press-release_MEMO-14-466_en.htm (cited on December, 20 2014).
Fialová, L., Kučera, M. (1997): The Main Features of Population Development in the Czech Republic during the Transformation of Society. Czech Sociological Review 5(1), 93–112.
Frejka, T. (1980): Fertility trends and policies: Czechoslovakia in the 1970s. Population and Development Review 6(1), 65–93, https://doi.org/10.2307/1972658.
Goldstein, J. R., Kreyenfeld, M., Jasilioniene, A., Karaman Örsal, D. (2013): Fertility reactions to the 'Great Recession' in Europe: Recent evidence from order-specific data. Demographic Research 29(4), 85–104, https://doi.org/10.4054/DemRes.2013.29.4.
Hoem, J. M., Jalovaara, M., Mureşan, C. (2013): Recent fertility patterns of Finnish women by union status: A descriptive account. Demographic Research 28(14), 409–420, https://doi.org/10.4054/DemRes.2013.28.14.
Chromková-Manea, B., Rabušic, L. (2013): Male Fertility in the Czech Republic – First Empirical Evidence. Demografie 55(4), 275–290.
Kalwij, A. (2010): The Impact of Family Policy Expenditure on Fertility in Western Europe. Demography 47(2), 503–519, https://doi.org/10.1353/dem.0.0104.
Kantorová, V. (2004): Education and Entry into Motherhood: The Czech Republic during State Socialism and the Transition Period (1970–1997). Demographic Research (special collection) 3(10), 245–274, https://doi.org/10.4054/DemRes.2004.S3.10.
Klasen, S., Launov, A. (2006): Analysis of the determinants of fertility decline in the Czech Republic. Journal of Population Economics 19(1), 25–54, https://doi.org/10.1007/s00148-005-0030-2.
Kohler, H.-P., Billari, F. C., Ortega, J. A. (2002): The emergence of lowest-low fertility in Europe during the 1990s. Population and Development Review 28(4), 641–680, https://doi.org/10.1111/j.1728-4457.2002.00641.x.
Kostelecký, T., Vobecká, J. (2009): Housing Affordability in Czech Regions and Demographic Behaviour – Does Housing Affordability Impact Fertility? Czech Sociological Review 45(6), 1191–1214.
Kravdal, Ø. (1994): The importance of economic activity, economic potential and economic resources for the timing of first births in Norway. Population Studies 48(2), 249–267, https://doi.org/10.1080/0032472031000147786.
Kučera, M., Fialová, L. (1996): Demografické chování obyvatelstva České republiky během přeměny společnosti po roce 1989. Working Papers 96(1). Praha: Sociologický ústav AV ČR.
Kuhnt, A. K., Trappe, H. (2013): Easier said than done: childbearing intentions and thein realization in a short term perspective. MPIDR Working paper WP 2013-018.
Kurkin, R. (2010): Vývoj plodnosti ve státech a regionech Evropské unie po roce 1991. [Diploma thesis]. Praha: Univerzita Karlova, Přírodovědecká fakulta, Katedra demografie a geodemografie.
Kurtinová, O. (2015): A Brief Insight into Gender Inequalities in the Czech Labour Market. Demografie 57(1), 5–20.

Lappegård, T., Rønsen, M. (2013): Socioeconomic Differences in Multipartner Fertility Among Norwegian Men. Demography 50(3), 1135–1153, https://doi.org/10.1007/s13524-012-0165-1.

Liefbroer, A. C., Corijn, M. (1999): Who, What, Where, and When? Specifying the Impact of Educational Attainment and Labour Force Participation on Family Formation. European Journal of Population 15(1), 45–75, https://doi.org/10.1023/A:1006137104191.

Lundström, K. E., Andersson, G. (2012): Labor market status, migrant status, and first childbearing in Sweden. Demographic Research 27(25), 719–742, https://doi.org/10.4054/DemRes.2012.27.25.

Matysiak, A., Vignoli, D. (2013): Diverse Effects of Women's Employment on Fertility: Insights from Italy and Poland. European Journal of Population 29(3), 273–302, https://doi.org/10.1007/s10680-013-9287-4.

Neels, K., De Wachter, D. (2010): Postponement and recuperation of Belgian fertility: how are they related to rising female educational attainment? Vienna Yearbook of Population Research 2010, 8(special issue on "Education and Demography"), 77–106.

Neyer, G. (2009): Rodinná politika a plodnost v Evropě: Pronatalitní politika v souvislosti s politikou genderovou, politikou zaměstnanosti a opatření týkajícími se péče o dítě. Demografie 51(4), 235–251.

Özcan, B., Mayer, K. U., Luedicke, J. (2010): The impact of unemployment on the transition to parenthood. Demographic Research 23(29), 807–846, https://doi.org/10.4054/DemRes.2010.23.29.

Pailhé, A., Solaz, A. (2012): The influence of employment uncertainty on childbearing in France: A tempo or quantum effect? Demographic Research 26(1), 1–40, https://doi.org/10.4054/DemRes.2012.26.1.

Quesnel-Vallée, A., Morgan, S.P. (2003): Missing the target? Correspondence of fertility intentions and behaviour in the US. Population Research and Policy Review 22(5–6), 497–525, https://doi.org/10.1023/B:POPU.0000021074.33415.c1.

Rabušic, L. (2001): Value change and demographic behaviour in the Czech Republic. Czech Sociological Review 9(1), 99–122.

Rychtaříková, J. (2010): Pokles porodnosti-hlavní faktor demografické změny. In: Burcin, B., Fialová, L., Rychtaříková, J. et al. Demografická situace České republiky. Proměny a kontexty 1993–2008. Praha: SLON, 47–63.

Rychtaříková, J. (2001): Demographic transition or demographic shock in recent population development in the Czech Republic? Acta Universitatis Carolinae Geographica 35(1), 89–102.

Rychtaříková, J. (2004): Změny generační plodnosti v České republice se zaměřením na vzdělání žen. Demografie 46(2), 77–90.

Régnier-Loilier, A., Vignoli, D. (2011): Fertility Intentions and Obstacles to their realization in France and Italy. Population-E 66(2), 361–390, https://doi.org/10.3917/pope.1102.0361.

Řeháková, B. (2000): Nebojte se logistické regrese. Sociologický časopis 36(4), 475–492.

Santerelli, E. (2011): Economic resources and the first child in Italy: A focus on income and job stability. Demographic Research 25(9), 311–336, https://doi.org/10.4054/DemRes.2011.25.9.

Sigle-Rushton, W. (2009): Plodnost v Anglii a Walesu – nečekané souvislosti rodinné politiky? Demografie 51(4), 258–265.

Sobotka, T. (2003): Re-emerging diversity: Rapid fertility changes in Central and Eastern Europe after the collapse of the communist regimes. Population-E 58(4–5), 451–486, https://doi.org/10.3917/pope.304.0451.

Sobotka, T. (2004a): Is lowest-low fertility in Europe explained by the postponement of childbearing? Population and Development Review 30(2), 195–220, https://doi.org/10.1111/j.1728-4457.2004.010_1.x.

Sobotka, T. (2004b): Postponement of childbearing and low fertility in Europe. [PhD thesis]. Amsterdam: Dutch University Press: University of Groningen.

Sobotka, T. (2011): Fertility in Central and Eastern Europe after 1989. Collapse and gradual recovery. Historical Social Research (Special issue Fertility in the 20thCentury: trends, policies, theories, discourses) 36(2), 246–296.

Sobotka, T., Skirbekk, V., Philipov, D. (2011): Economic Recession and Fertility in the Developed World. Population and Development Review 37(2), 267–306, https://doi.org/10.1111/j.1728-4457.2011.00411.x.

Sobotka, T. et al. (2008): Czech Republic: A rapid transformation of fertility and family behaviour after the collapse of state socialism. Demographic Research 19(14), 403–454, https://doi.org/10.4054/DemRes.2008.19.14.

Sobotka, T., Zeman, K., Kantorová, V. (2003): Demographic Shifts in the Czech Republic after 1989: A Second Demographic Transition View. European Journal of Population 19(3), 249–277, https://doi.org/10.1023/A:1024913321935.

Spéder. Z., Kapitány, B. (2009): How are Time-dependent Childbearing Intentions Realized? Realization, Postponement, Abandonment, Bringing Forward. European Journal of Population 25, 503–523, https://doi.org/10.1007/s10680-009-9189-7.

Šídlo, L. (2008): Faktory ovlivňující regionální diferenciaci plodnosti v Česku na počátku 21. století. Demografie 50(3), 153–172.

Šprocha, B. (2014): Odkladanie a rekuperácie plodnosti v kohortnej perspektíve v Českej republike a na Slovensku. Demografie 56(3), 219–233.

Šťastná, A. (2009): Zakládání rodiny a narození prvního a druhého dítěte. In: Kuchařová, V., Šťastná, A. (eds.). Partnerství, rodina a mezigenerační vztahy v české společnosti. Praha: VÚPSV, 29–49.

Toulemon, L., Testa, M. R. (2005): Fertility intentions and actual fertility: a complex relationship. Population and Societies, p. 415.

Data sources

Demografická ročenka České republiky 1991–2012 [Demographic Yearbook of the Czech Republic 1991–2012]. Praha: Český statistický úřad [Czech Statistical Office]. https://czso.cz/csu/czso/demograficka-rocenka-ceske-republiky-2012-qdv084etsg

Pohyb obyvatelstva v Českých zemích 1785–2015 [Population changes in the Czech lands 1785–2013]. Praha: Český statistický úřad [Czech Statistical Office]. https://www.czso.cz/csu/czso/obyvatelstvo_hu

Population and Housing Censuses 1991, 2001 and 2011. Internal databases with census data. Praha: Český statistický úřad [Czech Statistical Office].

Statistická ročenka České republiky 2013 [Statistical Yearbook of the Czech Republic]. Praha: Český statistický úřad [Czech Statistical Office]. https://www.czso.cz/csu/czso/statisticka-rocenka-ceske-republiky-2013-pxygncc90n

Spatial prediction of soil infiltration using functional geostatistics

Diego Leonardo Cortes-D, Jesus Hernan Camacho-Tamayo*, Ramon Giraldo

Universidad Nacional de Colombia, Department of Civil Engineering and Agricultural, Colombia
* Corresponding author: jhcamachot@unal.edu.co

ABSTRACT

The infiltration of water into the soil is a necessary parameter for irrigation systems design. Characterizing its spatial behavior allows a site-specific management of water according to soil conditions and crop requirements. The aim of this study is to establish the spatial distribution of infiltration in an Andisol by means of two geostatistical approaches: on the one hand by means of functional kriging, taking as input infiltration curves (obtained after a smoothing stage), and on the other hand by using classical ordinary kriging on the parameters of the Kostiakov and Phillip models. The comparison between these methodologies is carried out taking as a criterion the sum of squared errors of a leave-one-out cross-validation analysis. The results show a high correlation between observations and predictions (R^2 values around 99%), which indicates that the use of functional geostatistics in this context could be a good alternative. Moreover, from a descriptive point of view, we can point out that the contour maps of basic infiltration (BI), cumulative infiltration (Ci), saturated hydraulic conductivity (Ks), and sorptivity (S) obtained with the observed data, as well as the predictions by functional geostatistics, show a very similar behavior, which empirically validates the use of this methodology.

KEYWORDS

functional data; functional kriging; Kostiakov; Philip; spatial variability

Received: 29 October 2016
Accepted: 19 June 2018
Published online: 27 August 2018

Cortes-D, D. L., Camacho-Tamayo, J.-H., Giraldo, R. (2018): Spatial prediction of soil infiltration using functional geostatistics. AUC Geographica 53(2), 149–155
https://doi.org/10.14712/23361980.2018.15

1. Introduction

Soil management does not usually consider the particularities of the terrain or its physical properties; thus it is important to perform an evaluation based on the concepts of sustainability (Martins et al. 2010). In efficient and sustainable agricultural production systems, it is important to understand the behavior of water within the soil in order to be able to use this resource reasonably.

Infiltration is a hydrodynamic attribute that concerns the movement of water and that is closely related to the capillarity processes and forces associated with the adhesion and cohesion of soil particles (Orjuela-Matta et al. 2010). Moreover, infiltration is a basic parameter that must be accounted for when designing and implementing irrigation systems (Chowdary et al. 2006; Machiwal et al. 2006). Other attributes associated with soil water movement include hydraulic conductivity and sorptivity (S). Hydraulic conductivity is associated with the resistance of soil pores relative to the soil itself, is used to solve drainage and soil conservation problems, and depends not only on soil properties but also on the soil water content (Jačka et al. 2016). Sorptivity explains the movement of water in the soil as an effect of its matric potential, since S is estimated as a measure of the ability of a porous medium to absorb or desorb fluid by capillarity (Moret-Fernández et al. 2017).

Understanding soil behavior and its rate of infiltration over time could help control soil erosion or other degradation processes, establish the availability of water for crops and the movement of substances in the soil, and be used to create strategies for watershed management (Orjuela-Matta et al. 2010).

In general, the study of soil infiltration begins with experimental trials in a finite group of sites within an area of interest. Next, the experimental data are fit to theoretical models, such as those of Kostiakov, Lewis, Horton, or Philip, among others, and the empirical parameters of the models are determined. This process concludes with a calculation of the univariate descriptive measures (localization and dispersion), obtaining distribution graphs (histograms and box plots), conducting multivariate analyses (correlation, classification, and principle components), and performing univariate geostatistical analyses (variogram estimation, kriging prediction, and creation of maps). The spatial behavior can be described (Camacho-Tamayo et al. 2013) and the management zones can be established for agricultural production by using these tools (Cucunubá-Melo et al. 2011).

Geostatistics (Isaaks, Srivastava 1989) is a field of statistics which is used in practice for predicting a variable for non-sampled locations of a region. Specifically, several variations of kriging and cokriging methods (Cressie 1993) are used to fulfill this task. These tools are considered when for each sampled site we have data for one or several properties (for example, hydraulic conductivity, or sorptivity). However, in many practical situations a large number of measurements are recorded continuously (depth, time, wavelength, etc.). In these cases, instead of univariate or multivariate data, we have (after a smoothing step) curves or functions (functional data). The extension of the classical univariate and multivariate geostatistical methods to the context of functional data has given rise a new field known as functional geostatistics (Giraldo et al. 2010). In particular, kriging and cokriging methods have been extended to in order to deal with functional data (Giraldo et al. 2009; Giraldo et al. 2010). These tools allow predicting curves rather than data of one or several variables. The application of functional geostatistics starts with the application of non-parametric smoothing techniques (e.g., kernel or B-spline regression) to convert the discrete data at each sampling site (i.e. infiltration) into continuous functions. After that, functional kriging (Giraldo et al. 2010) can be used for predicting curves (instead of scalar values) at sites for which information is lacking.

In this study, functional geostatistics is applied to infiltration curves in order to evaluate their predictive capacity. A cross-validation analysis is performed to estimate prediction errors and compare them with those obtained using standard univariate kriging methods that are used to analyze infiltration data (Giraldo et al. 2010). The infiltration curve of each site (obtained after smoothing the discrete data) is temporally removed and predicted by using functional kriging and classical ordinary kriging (in this case, the parameters of the Kostiakov and Philip models are predicted, and then they are used to construct the predicted curve).

This paper is organized as follows: Section 2 gives an overview of the data analyzed, the Kostiakov and Phillip models, and the theory of functional geostatistics. Then in Section 3, a comparison between functional and classical univariate kriging methods (based on infiltration data) is carried out.

2. Material and methods

2.1 Description of study site

This study was performed at the Marengo Agricultural Center (Centro Agropecuiario Marengo, CAM by its initials in Spanish) located in the municipality of Mosquera (Cundinamarca, Colombia) at 4°42′ N and 74°12′ W and at an altitude of 2,543 m. The mean annual temperature in this zone is 13.1 °C, and the mean annual rainfall is 665 mm (weather records of the National University of Colombia). The area described, according to temperature and precipitation characteristics, is classified according to Holdrige as a life zone of low montane dry forest (bs-MB) and the climate as cold dry (FS).

The soils of the CAM have been formed from three types of parent materials, namely lacustrine clays, volcanic ash, and alluvial sediments. The type of relief corresponds to terraces and floodplains, specifically terrace plains and overflow plains (Bolívar, Ordóñez 2014).

2.2 In-situ infiltration measurement

An evenly spaced, rigid grid with 75 recording sites was created over four plots with mainly kikuyu grass (*Pennisetum clandestinum*). These plots covered an area of 16 ha and were located using a GPS Garmin Etrex 20 receiver in real time via satellite. At each site, a double-ring infiltrometer was used to measure water infiltration over a period of 150 minutes (at 1, 2, 3, 4, 5, 10, 15, 30, 45, 60, 90, 120, and 150 minutes) using variable loads and while ensuring that the difference between the maximum and minimum reading was never greater than 100 mm.

2.3 Determination of infiltration characteristics

The results obtained from the field were fit to two theoretical infiltration models, the Kostiakov (Equation 1) and the Philip (Equation 4) models. The parameters for these models were estimated by using the R software (R Development Core Team 2011) and then were used to calculate S, K_s, B_I, and C_i and to carry out conventional statistical and geostatistical analyses.

Kostiakov model

$$I(t) = a \cdot t^b \tag{1}$$

where
$I(t)$ is the cumulative infiltration content (cm) at time t and a and b are empirically-derived fit coefficients.

$$-600(b-1) = t_{base} \tag{2}$$

$$B_I = 60 \cdot a \cdot b \cdot t_{base}^{\,b-1} \tag{3}$$

with B_I as the basic infiltration (cm h^{-1}).

Philip model

$$I(t) = S \cdot t^{1/2} + K_s \cdot t \tag{4}$$

where S depends on the initial soil moisture conditions (cm h$^{-1/2}$) and K_s is the saturated hydraulic conductivity of the soil (cm h^{-1}).

2.4 Functional geostatistics

Functional geostatistics (Giraldo et al. 2010) offers methods for spatial prediction of curves. Usually, the discrete data (for example infiltration records) at each sampling site are previously recorded in continuous curves after smoothing them by using basis functions (Ramsay, Silverman 2005). Afterward, they are used for predicting a whole curve at unsampled sites. A cross-validation analysis was carried out. Each of the 75 curves was temporarily removed and predicted based on the remaining 74 by using functional kriging (Giraldo 2009). Thus observed and predicted curves (by using functional kriging) were obtained for each of the 75 sites. The comparison between them (cross-validation) was used for evaluating the quality of this approach and also to compare with the results obtained by means of classical geostatistics.

The functional kriging predictor is defined as follows (Giraldo 2009):

$$\hat{\chi}(s_0) = \sum_{i=1}^{n} \lambda_i \chi(s_i);\ \lambda_1, \ldots, \lambda_n \in \mathbb{R} \tag{5}$$

where $\hat{\chi}(s_0)$ is the predicted curve at site s_0, $\chi(s_i)$ corresponds to the curve observed at site s_i, and i = 1.2, … , n and λ_i = 1, … , n, are the weighting parameters that indicate the contributions of each observed curve to the predicted curve.

The λ_i parameters are estimated by solving the following system of equations (Giraldo et al. 2010):

$$\begin{pmatrix} \int_T \gamma_t(\|s_1 - s_1\|)dt & \cdots & \int_T \gamma_t(\|s_1 - s_n\|)dt & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \int_T \gamma_t(\|s_n - s_1\|)dt & \cdots & \int_T \gamma_t(\|s_n - s_n\|)dt & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} \begin{pmatrix} \int_T \gamma_t(\|s_0 - s_1\|)dt \\ \vdots \\ \int_T \gamma_t(\|s_0 - s_n\|)dt \\ 1 \end{pmatrix} \tag{6}$$

where the integrals correspond to the trace-variogram function (Giraldo et al. 2010) evaluated for both the distances between observation sites (left matrix) and the distances between the observation site and the prediction site (right vector). The trace-variogram function is estimated by the method of moments (Giraldo et al. 2010):

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} \int_T \big((\chi_t(s_i) - \chi_t(s_j)\big)^2 dt \tag{7}$$

where $\chi_t(s_i)$ is the curve at site i and $N(h) = \{(s_i,s_j): \|s_i - s_j\| = h\}$ is the number of pairs of sites separated by a distance h. Once the trace-variogram function for a sequence of K values is estimated, a parametric semivariance model (spherical, Gaussian, exponential, etc.) is fit to the scatterplot in order to make an estimation at any possible distance between sites.

The results achieved with functional geostatistics were compared with those obtained by using conventional geostatistical methods (estimating the semivariogram and using classical ordinary kriging). For this process, the variables B_I, C_I, S, and K_s from the theoretical infiltration models were considered. For each case (each variable), the semivariance function was estimated (Cressie 1993):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{n(h)} [z(x_i) - z(x_i+h)]^2 \quad (8)$$

where $z(x_i)$ is the value of each variable at a site i, $z(x_i + h)$ is the value of the same variable at a point with a distance h away from the previous point, and $N(h)$ is the number of pairs of data separated by the distance h. This empirical semivariogram is fit using one of the above-mentioned theoretical semivariance models. Various criteria are considered for selecting the best model, including the coefficient of determination (R^2), the least sum of squared errors (LSSE), and the cross-validation correlation (CVC). The shared parameters among the theoretical semivariance models include the nugget (C_0), which is a discontinuity in the semivariogram at the origin, the variance of the process (C), and the reach (r), which is the distance until there is a spatial correlation. The nugget-variance ratio, $C/(C_0 + C)$, is also often used as a criterion for model selection. This parameter establishes the degree of spatial dependence (DSD) expressed by the studied attribute. Cambardella (1994) states that if the DSD is greater than 75% the dependence is strong, if the DSD is between 25% and 75% the dependence is moderate, and if the DSD is less than 25% the dependence is weak.

3. Results and discussion

In Figure 2, we can see the fit of the data to a decreasing potential curve, which describes the infiltration rate. The strong change in the first moments of the curve confirms that the soil was in a state of water deficit at the time of the test. Given the trend of the curves, the results suggest that at some points the soil did not reach a condition close to saturation, after being subjected to a constant application of water, for a period of 150 minutes, especially for those sites that they registered a high infiltration rate. This behavior was observed for all the infiltration tests performed and reported by different authors (Machiwal et al. 2006; Orjuela-Matta et al. 2010; Latorre et al. 2015).

To provide an example of a functional geostatistical analysis of collected data, a curve was predicted for a site located at coordinates 984475 (W) and 1009685 (N) (Figure 1), as shown in Figure 2. To perform this prediction, the trace-semivariogram function was estimated using Equation 7. By using this function and Equation 6, the λ_i parameters for the functional kriging predictor, as defined in Equation 5, were estimated. Additionally, based on the predicted curve, the parameters for the theoretical Kostiakov and Philip models were estimated (Table 1).

The predicted curve is consistent with the behavior patterns of the observations. After 60 minutes, the water entry rate into the soil, or infiltration velocity, became constant, due to the increased matric potential of the soil. According to the observations

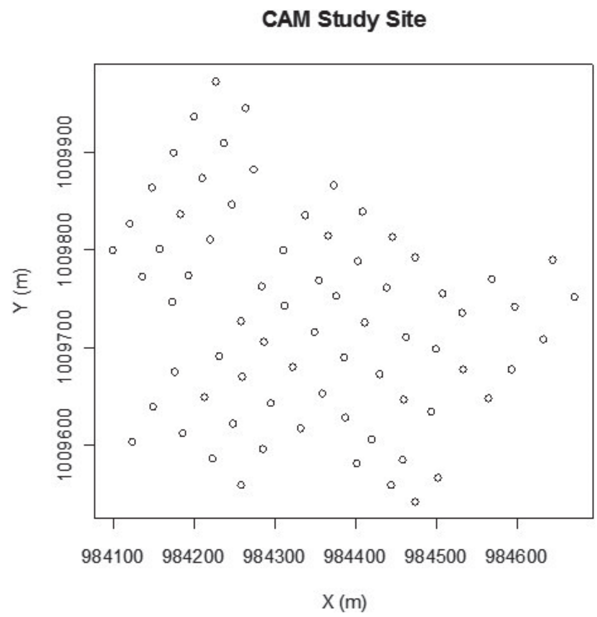

Fig. 1 Distribution of recording sites and identification of random sites where the cumulative infiltration (C_i) curve was predicted.

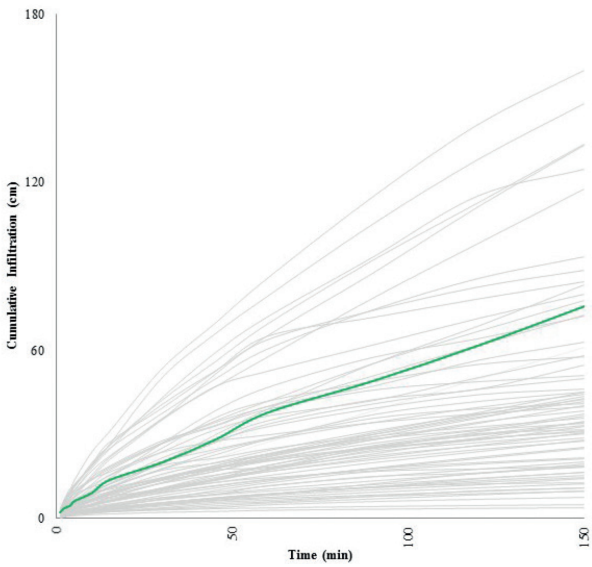

Fig. 2 Cumulative infiltration curve for the site at coordinates of 984475 (W) and 1009685 (N).

Tab. 1 Parameters estimated for Kostiakov and Philip models for the location with coordinates 984475 (W) and 1009685 (N).

Parameter	Value
a	1.53
b	0.78
B_i (cm h^{-1})	23.66
C_i (cm)	75.64
S (cm h$^{-1/2}$)	15.25
K_s (cm h^{-1})	20.48

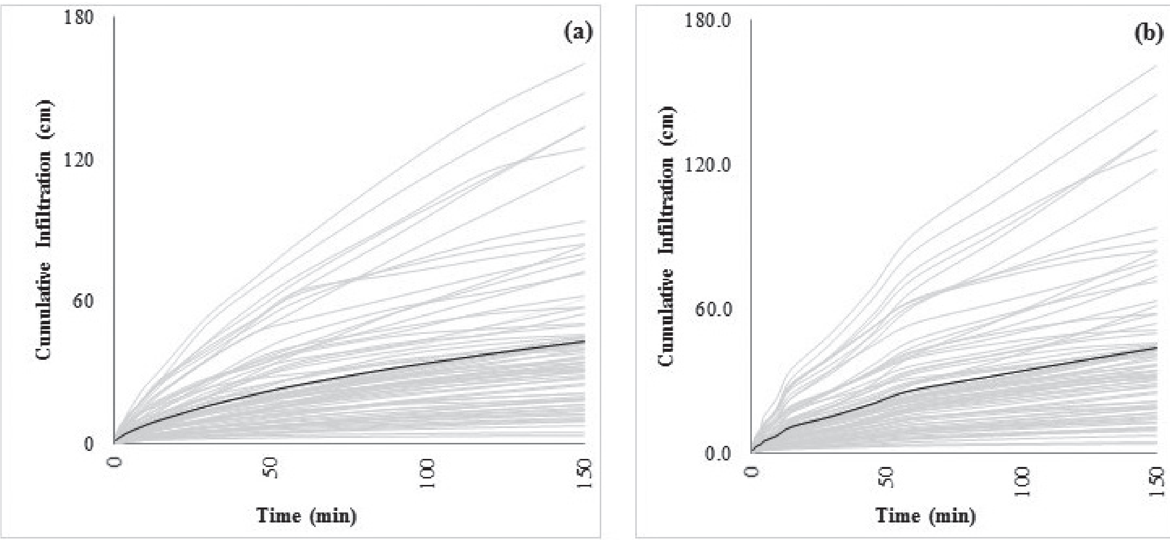

Fig. 3 (a) Cumulative infiltration curves recorded at the 75 sites. (b) Predicted curves at each site using functional kriging.

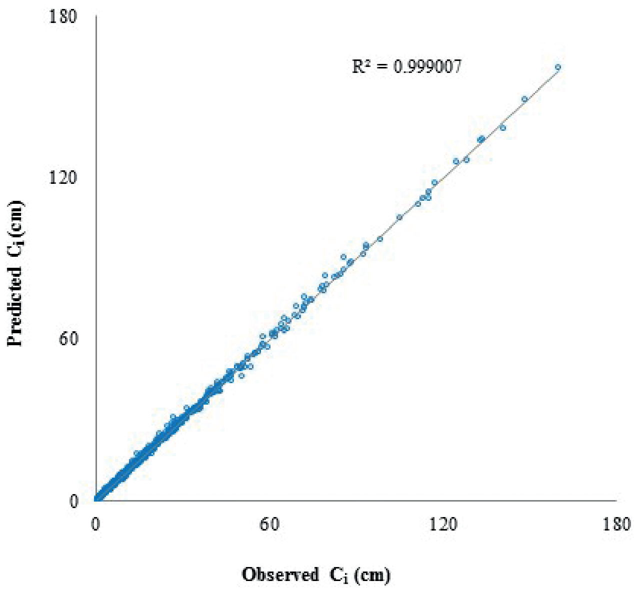

Fig. 4 Functional cross-validation of the amount of infiltration (Ci).

of Montenegro and Malagón (1990), the infiltration velocity at the unsampled site was high (12.7–25.4 cm h^{-1}), potentially due to a greater quantity of micropores (Richmon, Rillo 2006) relative to the sites with lower infiltration, which resulted in a high K_s value.

The recorded cumulative infiltration curves and the predictions at each site are shown in Figure 3. As an indicator of goodness-of-fit, a simple linear regression (Figure 4) was estimated to compare observed and predicted values. This plot shows a high correlation between them. The R^2 (around 99%) confirms that the method used allows obtaining good predictions.

We compared observed and predicted data of C_i, B_I, K_s, and S, respectively, by means of descriptive measures (Table 2), estimated parameters of the semivariogram models (Table 3), and maps of spatial distribution (Figure 5).

Tab. 2 Descriptive measures for the Kostiakov and Philip parameters estimated from the observed (Obs) and predicted (Pre) data.

Model	Parameter	Mean	Median	CV (%)	Min	Max
Kostiakov	a_{Obs}	2.09	1.59	70.87	0.34	6.44
	a_{Pre}	2.09	1.60	70.82	0.34	6.43
	b_{Obs}	0.61	0.60	16.08	0.36	0.86
	b_{Pre}	0.61	0.60	16.08	0.36	0.86
	$B_{I\,Obs}$	10.65	7.51	97.78	0.33	45.16
	$B_{I\,Pre}$	10.62	7.49	97.78	0.33	45.09
	C_{iObs}	45.08	35.35	77.82	3.90	160.00
	C_{iPre}	45.19	35.35	77.96	3.91	160.97
Philip	S_{Obs}	17.50	14.01	67.29	3.37	53.46
	S_{Pre}	18.03	14.25	67.12	3.40	53.44
	K_{sObs}	5.14	3.59	138.19	−7.65	36.16
	K_{sPre}	5.12	3.58	138.40	−7.68	36.02

In both cases (Kostiakov and Phillip models), the descriptive values in Table 2 are very similar, indicating a good performance of the functional kriging results.

Tab. 3 Parameters of the theoretical semivariogram model fit by observed (Obs) and predicted (Pre) basic infiltration values (BI), cumulative infiltration (Ci), sorptivity (S), and saturated hydraulic conductivity (Ks).

Parameter	Model	C_o	$C_o + C$	Range (m)	CVC	DSD
$B_{I\,Obs}$	Spherical	35.5	103.8	247.9	0.90	0.66
$B_{I\,Pre}$	Spherical	35.3	103.8	248.0	0.90	0.66
C_{iObs}	Exponential	393	1122	290.7	0.91	0.65
C_{iPre}	Exponential	404	1133	294.3	0.91	0.64
S_{Obs}	Exponential	0.10	124.9	94.20	0.79	1.00
S_{Pre}	Exponential	0.10	131.2	100.5	0.78	1.00
K_{sObs}	Exponential	0.10	53.20	268.5	0.82	0.99
K_{sPre}	Exponential	0.10	52.90	267.6	0.82	0.99

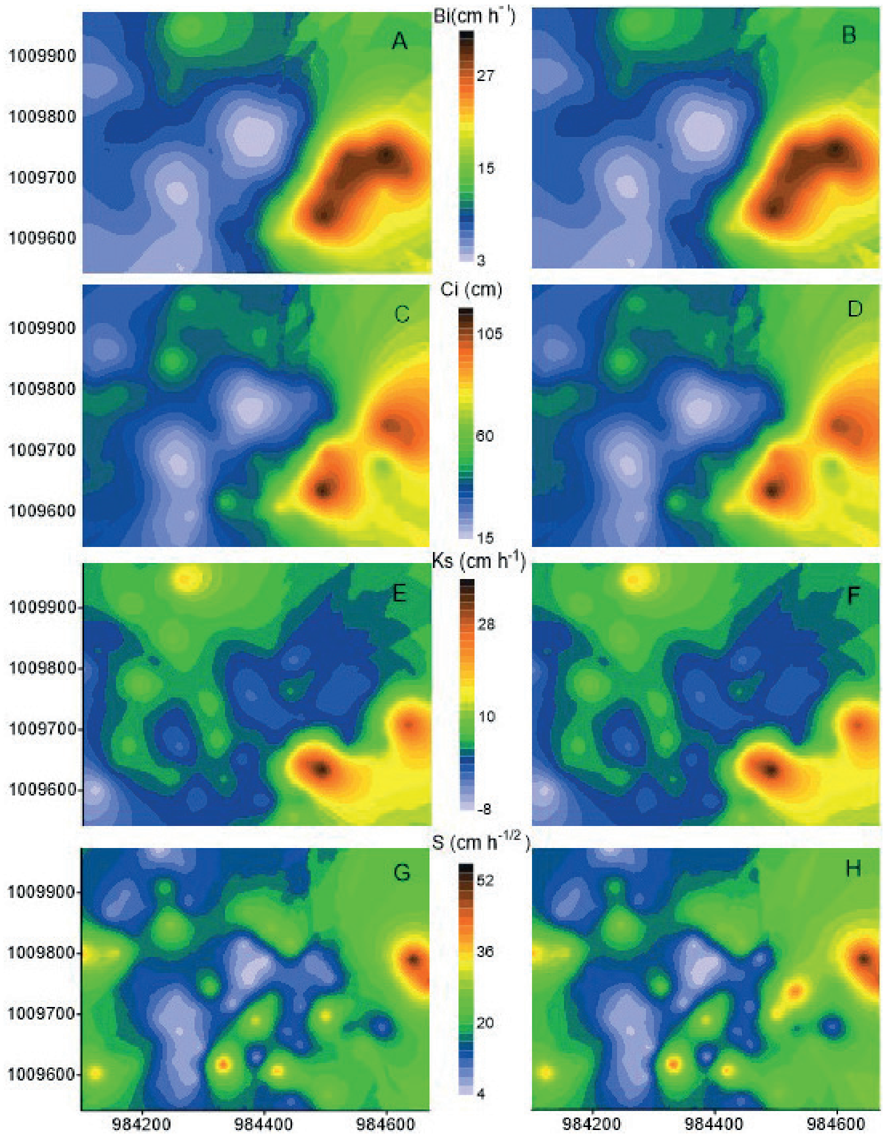

Fig. 5 Contour maps for observed (left) and predicted (right) data for (A–B) basic infiltration (B_i); (C–D) cumulative infiltration (C_i); (E–F) saturated hydraulic conductivity (K_s); and (G–H) Sorptivity (S).

The mean of parameter B_I (Table 2) indicates a moderately rapid infiltration velocity (Montenegro, Malagón 1990). However, at the location of the study, B_I varies from moderately slow to very fast. Some high infiltration velocities represent relatively dry soils due to the hydrological stress that they are subjected to. Additionally, high CV (greater than 60%) values indicate general areas or sites where infiltration behavior is very different from the other sites (Rodríguez-Vásquez et al. 2008; Martins et al. 2010).

The estimations of the semivariogram parameters (C_o, $C_o + C$ and Range, respectively) were similar (Table 3), which again indicates a good performance of the functional predictor used in this study. In both cases, spherical and exponential models were fit. In this sense, it is important to mention that some authors have reported problems for fitting semivariogram models to these properties (Rodríguez-Vásquez et al. 2008).

According to the results shown in Table 3, we can conclude that there is a moderate spatial dependence for C_i and B_I, because the values of DSD are greater than 25% and lower than 75% (Cambradella et al. 1974). This behavior is similar to that reported by Martins et al. (2010). In addition, the K_s and S show C_o values close to zero, which indicates a strong spatial dependence. A similar result was reported by Rodríguez-Vásquez et al. (2008) for the same type of soil.

The contour maps reveal high spatial variability in terms of the soil properties studied (Figure 5). The maps generated through the estimations obtained by means of functional kriging have a pattern very similar to those obtained from the recorded data (only minimal differences can be identified in each case). We can see in these maps that there is a direct relationship between the parameters (zones with high values of B_I have also high values of C_i). This is also evidenced with the parameters of the Phillip model.

We also note in these maps that there is direct relationship between K_s and B_I and C_i (zones with high values of K_s also have high B_I and C_i values). High values of K_s result from the passage of time and the saturation of soil pores with water, which cause the rate of infiltration to reach a constant value (B_I) that is similar to K_s (Gil, 2002), and consequently the soil can drain greater amounts of water and achieve greater C_i values.

4. Conclusions

The cross-validation analysis showed a high correlation between the results obtained with observed and predicted data (R^2 around 99%), which supported the use of functional geostatistics. Traditional statistical analyses demonstrated the reliability of applying FG to predict infiltration curves, given that the descriptive measures of the parameters for the theoretical infiltration model were similar for the observed data and predictions, even when the parameters displayed high spatial variability.

The estimations of the semivariogram parameters and the contours maps were also similar. The contour maps for B_I, C_i, K_s, and S for the recorded and predicted data demonstrate the wide range of these parameters. However, the maps also showed the same behavioral patterns, which confirmed the congruence between the studied models and the efficacy of predicting these parameters via FG. This was established even when the parameters displayed high spatial variability. All the analyses indicate that functional geostatistics can be a useful tool in the study of the spatial variability of soil properties.

References

Bolívar, A., Ordóñez, N. (2014). Levantamiento Agrológico del Centro Agropecuario Marengo. Instituto Geográfico Agustín Codazzi. Bogotá, Colombia.

Camacho-Tamayo, J. H., Rubiano Sanabria, Y., Santana, L. M. (2013): Management units based on the physical properties of an Oxisol. Journal of Soil Science and Plant Nutrition 13, 767–785. https://doi.org/10.4067/S0718-95162013005000061

Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., Konopka, A. E. (1994): Field-scale variability of soil properties in Central Iowa Soils. Soil Science Society of America Journal 58, 1501–1511. https://doi.org/10.2136/sssaj1994.03615995005800050033x

Chowdary, V. M., Damodhara Rao, M., Jaiswal, C. S. (2006): Study of infiltration process under different experimental conditions. Agricultural Water Management 83, 69–78. https://doi.org/10.1016/j.agwat.2005.09.001

Cressie, N. (1993): Statistics for Spatial Data. John Wiley and Sons.

Cucunubá-Melo, J. R., Alvarez Herrera, J. G., Camacho-Tamayo, J. H. (2011): Identification of agronomic management units based on physical attributes of soil. Journal of Soil Science 11, 87–99. https://doi.org/10.4067/S0718-95162011000100008

Gil, R. C. (2002). El comportamiento físico-funcional de los suelos. Buenos Aires: Instituto de Suelos. Inta Castelar, 19.

Giraldo, R. (2009). Geostatistical Analysis of Functional Data. Barcelona. PhD Thesis submitted to Universitat Politècnica de Catalunya

Giraldo, R., Delicado, P., Mateu, J. (2010): Continuous time-varying kriging for spatial prediction of functional data: an environmental application. Journal of Agricultural, Biological, and Environmental Statistics 15, 66–82. https://doi.org/10.1007/s13253-009-0012-z

Jačka, L., Pavlásek, J., Pech, P., Kuráž, V. (2016): Assessment of evaluation methods using infiltration data measured in heterogeneous mountain soils. Geoderma 276, 74–83. https://doi.org/10.1016/j.geoderma.2016.04.023

Isaaks, E., Srivastava, M. (1989): Applied Geostatistics. Oxford University Press.

Latorre, B., Peña, C., Lassabatere, L., Angulo-Jaramillo, R., Moret-Fernández, D. (2015). Estimate of soil hydraulic properties from disc infiltrometer three-dimensional infiltration curve. Numerical analysis and field application. Journal of Hydrology 527, 1–12. https://doi.org/10.1016/j.jhydrol.2015.04.015

Machiwal, D., Jha, M. K., Mal, B. C. (2006): Modelling infiltration and quantifying spatial soil variability in a wasteland of Kharagpur, India. Biosystems Engineering 95, 569–582. https://doi.org/10.1016/j.biosystemseng.2006.08.007

Martins, A. L., Moura, E., Camacho-Tamayo, K. H. (2010): Variabilidad espacial de la infiltración y su relación con algunas propiedades físicas. Ingeniería e Investigación 30, 116–123.

Montenegro, H., Malagón, D. (1990): Propiedades Físicas de los Suelos. Colombia: Instituto Geográfico Agustín Codazzi.

Moret-Fernández, D., Latorre, B., Angulo-Martínez. M. (2017). Comparison of different methods to estimate the soil sorptivity from an upward infiltration curve. CATENA 155, 86–92. https://doi.org/10.1016/j.catena.2017.03.009

Orjuela-Matta, H. M., Rubiano-Sanabria, Y., Camacho-Tamayo, J. H. (2010): Comportamiento de la infiltración en un Oxisol. Revista U.D.C.A Actualidad and Divulgación Científica 13, 31–39.

Ramsay, J., Silverman, B. (2005): Functional Data Analysis, second edition. New York: Springer. https://doi.org/10.1002/0470013192.bsa239

Richmond, P., Rillo, S. (2006): Evaluación del efecto de la compactación por el rodado de máquinas sobre algunas propiedades físicas del suelo y el cultivo de trigo en siembra directa. Argentina: INPOFOS Informaciones agronómicas, 32.

Rodríguez-Vásquez, A. F., Aristizábal-Castillo, A. M., Camacho-Tamayo, J. H. (2008): Variabilidad espacial de los modelos de infiltración de Philip y Kostiakov en un suelo ándico. Engenharia. Agrícola 28, 64–75. https://doi.org/10.1590/S0100-69162008000100007

Risk assessment of desertification using GIS in upper and lower reaches of Mond basin, Iran

Masoud Masoudi*, Parviz Jokar

Shiraz University, Department of Natural Resources and Environmental Engineering, Iran
* Corresponding author: masoudi@shirazu.ac.ir

ABSTRACT

The present paper attempts to develop a new model by considering various indicators of different types of land degradation or desertification. These types include water erosion, soil salinity, vegetation degradation, and lowering of ground water table. The indicators can be used to find areas with higher rates of degradation which are called Potential Risk Areas (risky zones) in this paper, and can also be used to estimate the probability that degradation will increase in these areas. The Mond river basin, located in the southern part of Iran, has been selected as a test area to assess the risk and kind of desertification. For this purpose two sub basins of the Khormuj and Khane-Zenian & Siakh-Darengun have been chosen for detailed study as these two provide enough variation in climatic conditions like rainfall and topography. The different kinds of data gathered from records and published reports of the different governmental offices of Iran have been used for this purpose. The thresholds for the severity classes of indicators have been established and then the hazard map for each indicator of types of desertification has been prepared in a GIS. The risk maps of water erosion, soil salinization, lowering of water table, and vegetation degradation have been produced for both sub basins. Areas on the maps are assigned to risk classes on the basis of risk scores derived by considering the cumulative effects of all indicators overlying the area in the GIS. It was possible to distinguish the areas under 'actual risk' from areas under 'potential risk' of desertification types. Also areas under potential risk are classified to subclasses with different probability level to show a statistical picture of risk in future. The final map of risk of desertification is produced by overlaying all four maps of degradation types. Between the two basins the overall environmental condition in the Khormuj sub basin is worse. Results show that potential risk areas are much widespread than areas under actual risk in the upper reaches (of both sub basins) of Mond basin, indicating further threat of land degradation or desertification in the future. The percent of areas under actual risk are much more extensive in the lower reaches (Khormuj sub basin), indicating the higher degradation at present. It is hoped that this attempt using GIS will be found applicable for other regions of the world.

KEYWORDS

desertification; GIS; indicator; actual risk; potential risk

Received: 1 January 2017
Accepted: 19 June 2018
Published online: 27 August 2018

Masoudi, M., Jokar, P. (2018): Risk assessment of desertification using GIS in upper and lower reaches of Mond basin, Iran.
AUC Geographica 53(2), 156–166
https://doi.org/10.14712/23361980.2018.16

1. Introduction

Drylands (arid, semi-arid and dry subhumid areas) are one example of a hotspot where land and populations are particularly vulnerable, both at present and into the future (Middleton et al. 2011). Drylands cover 41% of the planet's land area and are inhabited by more than 2 billion people (Middleton et al. 2011). Desertification (land degradation in drylands) has been ranked amongst the most urgent global environmental challenges (MA 2005). Indeed, land degradation will remain an important global issue for 21st century because of its impact on environment, and it effect on food security and the quality of life. Land degradation can be considered in terms of the loss of actual or potential productivity or utility as a result of natural or anthropogenic factors, it is the decline in land quality or reduction in its productivity (Zhang et al. 2014).

The United Nations Environmental Programme (UNEP) estimates 69% of the world's arid lands, excluding the very arid deserts, are under moderate to severe hazard of land degradation (Dregne 1991). The awareness about land degradation and desertification is not new. There are mentions of soil erosion dating back to classical times, and the subject was first brought to wide attention during the United States 'dust bowl' of the 1930s. In modern times, the human suffering caused by the drought in the West African Sahel zone during the 1970s led to the first UN Conference on Desertification in 1977. Since then, the problem has remained at the forefront of international discussions. In the early 1990s, desertification was defined as 'land degradation in arid, semi-arid and dry sub-humid areas resulting from various factors, including climatic variations and human activities' (UNEP 1992).

The alarming losses in economic revenues and agroecosystem services have revealed an acute need for monitoring of land degradation and analyses of its causes in order to advise decision makers on spatial targeting of land rehabilitation Measures (Dubovyk et al. 2013). Tracking trends in desertification, as well as land and soil degradation, and their links to different human and biophysical drivers is especially difficult (Eswaran et al. 2001), particularly if the information needs of stakeholders as diverse as policy makers, scientists, land managers and society at large are to be met (Vogt et al. 2011). The kind of authoritative and consensual assessments that are needed do not yet exist (UNCCD 2011). Within the coupled human-ecological system, it is necessary to identify critical variables that target both human and ecological system components. Monitoring needs to draw on indicators that measure both 'slow' and 'fast' human and biophysical variables (Reynolds et al. 2011).

Desertification involves a complex set of factors, interacting in space and time leading to a decrease in land productivity. It is closely related to many environmental factors such as climate, soil, vegetation cover, and morphology the character and intensity of which contribute to the evolution and characterization of different degradation levels (Masoudi and Amiri 2013; Barzani and Khairulmaini 2013; Masoudi 2014; Masoudi and Jokar 2015). Desertification is also strongly linked to socio-economic factors, since man's behavior and his social and economic actions can greatly influence the evolution of numerous environmental characteristics (Jafari and Bakhshandehmehr 2013).

Iran lies within the arid and semi-arid climatic belt, and in such climatic conditions the desertification processes are known to progress more speedily and pervasively. Arid and semi-arid regions cover more than 85% Iran's land, of which desert areas account for 34 million ha (FRW (Forest, Range, and Watershed Management Organization) 2004). Compared to other countries in the Middle East, the present status of desertification in Iran is alarming as about 94% of arable lands and permanent pastures are estimated to be in the process of degradation (FAO 1994).

Different models for assessing desertification such as mathematical methods, remote sensing, parametric equations, direct observation and measurement have been developed. Recently, several methods of desertification and land degradation have been used. The FAO/UNEP (1984) introduces the "Provisional Methodology for Assessment and Mapping of Desertification Hazard" which evaluates the main parameters affecting desertification processes. This method was the first major exercise that was developed to assess land degradation in arid and semi-arid regions. In this method, the hazard of desertification is assessed on the basis of five maps using the sum of numerical values of indices for desertification status map (DS), desertification rate (DR), inherent risk (IR), domestic animal pressure (AP), population pressure (PP). Hazard desertification (HD) map was calculated as:

$$DH = DS + DR + IR + AP + PP$$

It determines its severity classes like slight, moderate, severe and very severe. To produce status, rate and risk of land degradation, main types of land degradation in arid and semi-arid climate are selected as: water and wind erosion, soil salinization and vegetation degradation (FAO/UNEP 1984).

The MEDALUS model (Kosmas et al. 1999) identifies regions that are environmentally sensitive areas (ESAs). It brings the results of the physical and socio-economic aspects of desertification to bear on the identification and use of desertification indicators at various geographical scales from the local to the European. In this method, different classes of ESAs to desertification can be evaluated using various data such as landforms, soil, geology, vegetation, climate, and human actions. Each of these data is classified into various classes and a weighting factor is considered to each class. Then four main quality layers including soil, climate, vegetation, and management are evaluated. After assessing indices for each quality layer, the Environmental sensitivity index (ESI) is defined by combining the four quality layers. ESI is a composite indicator that can be used to gain an understanding of factors causing desertification risk at a point. All the data considering the four main layers are prepared in a geographical information system (GIS), and were overlain in accordance with the developed algorithm which takes the geometric mean to compile maps of ESAs to desertification.

Some other important models are GLASSOD[1] (Oldeman et al. 1991), ASSOD[2] (Van Lynden and Oldeman 1997) and LADA (Ponce Hernandez and Koohafkan 2004). In both ASSOD and GLASSOD projects, local experts assessed the relative impact of a given amount of a certain type of degradation on the productivity of the soil. This kind of assessment seems to be more realistic in finding the degree of degradation because it is more related to its impact on soil productivity.

Project of LADA has been set up by FAO, UNEP-GEF and various other partners to assess Land Degradation in Dryland Areas (LADA), in which GLASSOD is one of the methodologies to be reviewed for its potential benefits to this project. This method is based on a sequence or framework of rule of different indicator to degradation referred as DPSIR (Driving forces, Pressures, State, Impacts and Responses). The DPSIR framework is an approach to environmental hazard, developed by the European Environmental Agency, for describing, monitoring and controlling of environmental problems (EEA 1999). The approach is based on the use of DPSIR indicators, which may be direct or indirect, ecological, technical, socioeconomic or cultural causes of environmental hazard. The hazard maps of DPSIR indicators processed in the hazard assessment models give a far better opportunity to distinguish the severity classes of environmental hazard (FAO 2002; Ponce Hernandez and Koohafkan 2004). The existing **Driving forces** (D) in nature and society produce **Pressures** (P) on the natural resource that result in the current **State** (S) of land resources, with a negative **Impact** (I) on society and the environment. This, in turn, may stimulate a **Response** (R) (EEA 1999). Driving forces include those activities that may (in) directly cause the problem. Pressure indicators include those activities that may (in) directly result in an increased pressure on the natural resource. State indicators reflect the conditions of the land as well as its resilience to resist changes. Impact indicators describe the effect and impacts of the increased or reduced pressure on the natural resource. Impact indicators or change indicators measure change in either positive or negative direction (degradation or improvement). Response indicators include those activities by the land users themselves to release the pressure from the land. In some instances environmental regulations may be necessary to effect proper control of land degradation (EEA 1999; Masoudi and Amiri 2015).

Qi et al. (2013) used two landscape evaluation approaches, an integrated model and an ecological analysis method, based on landscape elements and environmental quality, respectively, to describe desertification in the Heihe River Basin of northwestern China, by evaluating the current state of the local ecosystems and environment. In total, 32 typical environment factors were selected, classifying desertification in the region into four zones.

Sepehr and Zucca (2012) introduced technique for order preference by similarity to ideal solution (TOPSIS) method as a decision-making method for the selection and integration of desertification indicators. It is a multiple criteria method to identify solutions from a finite set. TOPSIS is an algorithm for determining the most preferable choices among the possible indicators that can be developed. The simulation case study presented is related to the selection of the best set of indicators to monitor land degradation by remote sensing in three different countries (Brazil, Mozambique and Portugal), within the framework defined by the DesertWatch Extension project.

In Iran, Ekhtesasi and Mohajeri (1995) introduced the ICD[3] model for the classification of desertification in Iran. One of the advantages of this method is its capability to identify the type of desert like natural and anthropogenic deserts. ICD was developed in four steps: separation of deserts types using land use and plant types, distinguishing of desertification causes including the major and minor causes, classification of desertification and preparation of desertification maps. The model classifies the severity of desertification to five classes: slight, low, moderate, severe and very severe.

A new model for assessing hazard of desertification, namely, Iranian Model of Desertification Potential Assessment (IMDPA) is developed by Forests and Rangelands Organization of Iran**.** This model considers nine criteria or aspects of desertification, namely, climate, geology-geomorphology, soil, vegetation cover, agriculture, water, erosion (including wind and water erosion), social-economics, technology of urban development for finding areas with higher hazard of degradation. Each criterion is evaluated by three or four indicators. Total numbers of indicators introduced for the model are thirty five (Masoudi and Zakeri Nejad 2011).

This paper attempts at evolving a model for assessing risk of land degradation in southern part of Iran. For this purpose the Mond river basin for which enough data were available for variability in climate and land degradation types has been chosen. It is

1 Global Assessment of Soil Degradation

2 Assessment of Soil Degradation

3 Iranian Classification Deserts

hoped that this attempt using GIS, which is the first attempt of its kind done on Southern Iran for land degradation, will be found applicable for other regions of the world. The present work has given the opportunity to compare the intensity of different types of land degradation related to the two sub basins of Khane-Zenian & Siakh-Darengun (upper reaches of Mond River) and Khormuj sub basin (lower reaches of Mond River) which differ in elevation, climate and status of degradation. The total area covered in the GIS analysis is 1,787,000 ha.

2. Study area

The Mond basin is bounded between Lat. 27°20′ and 29°55′ N and Long. 51°09′ and 54°45′ E, it lies in the southern part of Iran (Fig. 1), covering an area of nearly 47,835 km^2. The basin is divided to five main sub basins on the basis of hydrology and topography. These basins are: Qareh Aghaj, Mond Miyani, Firuzabad, Payab and Harm-Kariyan. This Mond basin was selected as study area for this research because of availability of more data for the two sub basins, namely, the Khane-Zenian & Siakh-Darengun sub basin (upper reaches of Mond River in Qareh Aghaj basin) and the Khormuj sub basin (lower reaches of Mond River in Payab basin). This study provided a good opportunity to compare desertification in these two sub basins that differ in their physiographic and climatic status. The Khormuj sub basin covers an area of nearly 264,803 ha and the Khane-Zenian & Siakh-Darengun sub basin covers an area of nearly 160,242 ha.

The population of the entire basin is estimated to be about 2 million of which nearly half is urban. The rural population is engaged mostly in agricultural activity and cattle raising. The main river is the Mond River that flows down the Southern Zagros to the Persian Gulf. The part of the Mond River in the upper reaches is called as the Qareh Aghaj River. The elevation varies between the see level in the Payab basin to 3185 m in Kharman Kuh in the Qareh Aghaj basin. The landscape units are mountains, hills, piedmont

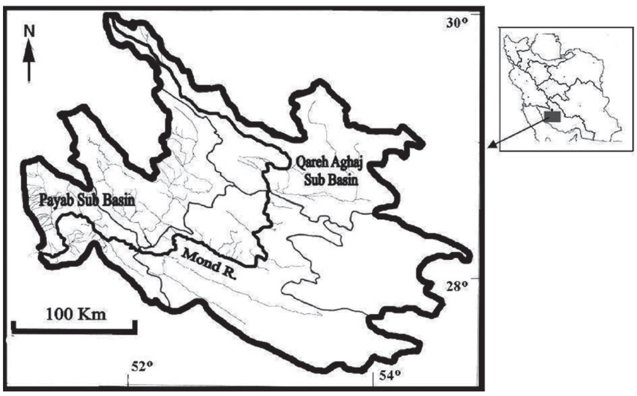

Fig. 1 Location map of Mond Basin in Iran.

and plains. The climate is arid and semi-arid in most part of the basin with a mean annual rainfall range of 150–700 mm. The main period of precipitation is during winter (60% of total rainfall). The mean annual temperature measured at the Qantareh station in Payab is 26.1 °C and for the Band Bahman in Qareh Aghaj is 14.4 °C.

3. Methods

1) Data gathering and processing for types of land degradation: The main types of land degradation in the two sub basins studied are: water erosion, wind erosion, soil salinization, lowering of ground water table and vegetation degradation. As the first step, the causes for each type of degradation have been identified. The data for this study have been gathered from the records and reports published by the different departments of the Ministries of Agriculture and Energy and the Meteorological Organization of Iran. The data obtained were of two typMasoudies: 1) numerical data and 2) thematic maps, but mainly in the map format (vector) with mostly semi-detailed scale (1 : 50,000 scale), useful for the GIS analysis. In all 18 maps have been digitized, 80 maps have been processed in the GIS. For example from a political boundary map of region after its digitizing in GIS, different socioeconomic maps were derived like density of population, percent of growth population, number of livestock and etc. The main types of data are on physiography (slope and land type), geology (rock formation), soil (depth, permeability, and erodibility factor), hydrology (water discharge), vegetation (density of vegetation cover and percentage of bare land) and climate (rainfall, evapotranspiration and temperature) and on some causes related to human activity such as over grazing, over pumping and density of population. These data especially thematic maps were verified to check quality of data with different way of verification in GIS like field observation and using topography maps. The data should not be old, because some data like vegetation cover or climate data change during time. Therefore in gathering of data this point took into consideration except those data that doesn't change during a short time like geology map or its information.

2) Preparation of indicator hazard maps: In this research nine different indicators have been selected to achieve the best model for assessing the risk of water erosion, soil salinization and vegetation degradation in both the sub basins of the Mond Basin. Also eleven indicators for lowering of water table have been selected for assessing its risk. The status map of wind erosion has also been prepared. The recommendations like by the FAO and other scientists and also the statistical parameters of the present data for local conditions have been considered for producing

Tab. 1 The indicators used in the model of risk assessment for Desertification.

Type of Desertification	Indicators	Class limits and their ratings score				
		None (1)	Slight (2)	Moderate (3)	Severe (4)	Very severe (5)
a) soil salinization	1) Depth of water table (m)	>5	3–5	1–3	0.5–1	<0.5
	2) Soil texture	Coarse soils of mountains and hills	Coarse to medium and medium	Moderately fine	Fine	Very fine (clay texture)
	3) Slope (%)	30+	15–29	5–14	1–4	<1
	4) Quality of irrigation water Ec (µmhos/cm)	<250	250–749	750–2249	2250–4999	5000+
	5) Ground water quality Ec (µmhos/cm) SAR	<250 <10	250–749 10–17	750–2249 18–25	2250–4999 26–29	5000+ 30+
	6)Efficacy of surface geology (ESG)*	<0.1	0.1–0.39	0.40–0.65	0.65–1	1+
	7) Climate	Sub humid and humid	Slightly semi-arid	Semi-arid	Arid	Very arid
	8) Dry index (P/ETP)	0.60+	0.40–0.59	0.20–0.39	0.05–0.19	<0.05
	9) Status of soil salinity EC (mmhos/cm) SAR	<4 <8	4–8 8–13	8–16 13–30	16–32 30–70	>32 >70
b) water erosion	1)Soil erodibility	<0.1	0.1–0.19	0.2–0.34	0.35–0.49	≥0.5
	2) Soil depth (cm)	Very deep or ≥ 150 cm	Deep or 90–149 cm	Semi deep to deep or 50–89 cm	Shallow to semi deep or 10–49 cm	Very shallow to shallow or <10 cm
	3) Per cent slope	<2	2–4	5–14	15–29	≥30
	4) Intensity of rainfall**	<10	10–19	20–29	30–39	≥40
	5) Total rainfall (mm)	<50	50–199	200–399	400–599	600–1000
	6) Per cent bare ground	<20	20–39	40–59	60–79	≥80
	7) Per cent vegetation cover	≥70	50–69	25–49	10–24	<10
	8) Status of water erosion	Features of erosion insignificant	Sheet and rill erosion and occasional gully erosion visible	Sheet and rill erosion moderate and occasional gully erosion visible	Fairly high abundance of features of sheet, rill and gully erosion	Highly abundant sheet, rill and gully erosion (badlands)
	9) Erodibility of surface geology	Formations resistant against water erosion and thick alluvial deposits of plains.	Formations fairly resistant against water erosion	Formation with moderate resistance against water erosion	Formations with low resistance against water erosion	Formations susceptible to water erosion like salt domes associated with layers of marl, shale.

Type of Desertification	Indicators	Class limits and their ratings score				
		None (1)	Slight (2)	Moderate (3)	Severe (4)	Very severe (5)
C) vegetation degradation	1) Potential of biomass production (kg/ha)	≥1000	650–999	350–649	100–349	<100
	2) Pressure of livestock***	≥5	1.5–5	1.0–1.5	0.5–1.0	<0.5
	3) vegetation cover (%)	≥70	50–69	25–49	10–24	<10
	4) Expansion of agricultural activity over lands suitable for natural resources	Natural resource lands (without any changes)	Irrigated lands with low limitations	Irrigated lands with limitations, or dry cultivation with low limitations	Dry cultivation with high limitations	Dry cultivation with very high limitations
	5) Rural Population Density (per sq. km)	<1	1–4	5–19	20–34	≥35
	6) Villages density (per sq. km)	0	0–0.02	0.02–0.06	0.06–0.09	≥0.09
	7) Climate	Sub humid and humid	Slightly semi-arid	Semi-arid	Arid	Very arid
	8) Coefficient variation (CV) of annual rainfall	<20	20–29	30–39	40–49	≥50
	9) Land suitability for vegetation cover	Very good	Good (suitable)	Medium	Low	Poor or very poor (Unsuitable soils)
d) lowering of ground water table	1) Over exploitation#	≥1.1	1 ≤ 1.1	0.9 ≤ 1	0.8 ≤ 0.9	<0.8
	2) Increased Consumption of ground water in the 10 years	< 1.10	1.1–1.32	1.33–1.65	1.66–1.99	≥2
	3) % Surface water consumption of total water consumption	≥75	50–74	25–49	10–24	<10
	4) Ratio of non-irrigated areas to irrigated areas	≥3	1.50–2.99	0.75–1.49	0.25–0.74	<0.25
	5) Average water consumption in irrigated areas (M^3/ha)	–	Other parts of plain	<10500	10500–16500	>16500
	6) Ratio of water exploitation from qanats## to that from wells	> 1	0.34–1	0.18–0.33	0.06–0.17	≤0.05
	7) Climate	Sub humid and humid	Slightly semi-arid	Semi-arid	Arid	Very arid
	8) Coefficient variation (CV) of annual rainfall	<20	20–29	30–39	40–49	≥50
	9) Annual rainfall, mm	≥1000	500–999	250–499	100–249	<100
	10) Influence of carbonate formations###	≥3	1.00–2.99	0.50–0.99	0.25–0.49	<0.25

Type of Desertification	Indicators	Class limits and their ratings score				
		None (1)	Slight (2)	Moderate (3)	Severe (4)	Very severe (5)
d) lowering of ground water table	11) Hydrogeology of plains	Coarse-grained texture, very thick alluvium, deep water table, excellent discharge	Medium to coarse-grained texture, thick alluvium, deep water table, good discharge	Relatively fine-grained texture, moderately thick alluvium, shallow water table, medium discharge	Fine-grained texture, thin alluvium, shallow water table, poor discharge	Fine to very fine-grained texture, very thin alluvium, shallow water table or no aquifer, very poor discharge

* ESG = ((2 × salt dome area) + (evaporates area) + (0.5 × area of formations with less evaporate material) / area of other formations

** Average of maximum for amount of rainfall in mm during 6 hours for period of 2 years

*** Pressure of livestock = potential of carrying capacity / actual density of livestock (Masoudi and Vahedi, 2014)

Over exploitation = Safe exploitation / actual extraction exploitation in MM^3 (million cube meter)

Qanat is a traditional form of freshwater extraction mostly in desert and arid areas to reflect and process water from upland area

R = extent of carbonate formations / extent of non-carbonate formations

hazardous thresholds of the indicators, revealing 'none' to 'very severe' hazardous conditions (ratings scores between 1 and 5) to assess the risk of these types of degradation (Table 1). The hazard maps have been prepared in the GIS for each indicator.

3) Producing of risk maps for each type of land degradation: To project the effect of all the indicators the hazard maps for each type of land degradation were overlaid in the GIS using the following equations, giving proper weighting for each indicator. The attributes × 2 indicate their relative importance in assessing the severity of risk. On the other hand, the indicators that have less impact were given weighting (1):

Risk score for water erosion = ((Soil depth + Slope + Status of water erosion) × 2) + Erodibility of surface geology + Intensity of rainfall + Annual rainfall + Soil erodibility + Vegetation cover + Bare ground

Risk score for soil salinization = (Status of soil salinity × 2) + Efficacy of surface geology + Quality of irrigation water + Depth of water table + Ground water quality + Soil texture + Climate + Dry index + Slope

Risk score for vegetation degradation = ((Potential of biomass production + Vegetation cover + Rural population density + Pressure of livestock) × 2) + Expansion of agricultural activity over lands suitable for natural resources + Villages density + Climate + Coefficient variation (CV) of annual rainfall + Land suitability for vegetation cover

Risk score for lowering of ground water table = ((Annual rainfall + Hydrogeology of plains + Over evacuation + Increased consumption of ground water in the 10 years + Surface water consumption + Average water consumption in irrigated areas) × 2) + Ratio of non-irrigated areas to irrigated areas + Ratio of water evacuation from qanats to that from wells + Climate + Coefficient variation (CV) of annual rainfall + Influence of carbonate formations.

The risk score arrived at enabled to subdivide the severity classes of degradation types. Five such severity classes ranging from 'none' to 'very severe' were recognized (Table 2). The GIS analysis enabled the distinction of areas under 'actual risk' from areas under 'potential risk' of land degradation types. The actual risk areas show at present a state of degradation equal to or worse than the classes assigned for the risk. The present status of hazard is determined by considering the attributes of indicators 9, 8, 3 and 10 for soil salinization, water erosion, vegetation degradation and lowering of ground water table, respectively (Table 1).

Tab. 2 Severity classes in the Risk Maps and GIS models regarding the scores of polygons.

Risk type	Class limits and their score in the GIS				
	None	Slight	Moderate	Severe	Very severe
1) Water Erosion	12–18	18.1–30	30.1–42	42.1–54	54.1–60
2) Soil Salinization	10–15	15.1–25	25.1–35	35.1–45	45.1–50
3) Vegetation Degradation	13–19.5	19.6–32.5	32.6–45.5	45.6–58.5	58.6–65
4) Lowering of Ground Water Table	17–25.5	25.6–42.5	42.6–59.5	59.6–76.5	76.6–85

Areas under potential risk have been recognized using the following criteria:

(A) Potential risk area = areas where the risk class determined > present status of hazard.

The Potential risk areas include areas that, at present, show a state of degradation lower than the classes that are predicted by the risk analysis. For example, areas under 'moderate' potential risk have at present slight or no degradation but have moderate vulnerability towards worse conditions. For calculating the probability for potential risk, the risk scores have been converted to percentage. The potential risk classes were further divided into sub classes of severity in the risk maps, based on per cent probability of potential risk, thus giving a statistical picture of the risk. The following equation was used for this purpose:

(B) % Probability of Risk in Potential Risk Areas = $[(X - a)/b] \times 100$, where a is the least score (0% probability) for each type of land degradation in Table 2, b is the numeric difference between the highest and the least scores for each type of land degradation in Table 2, and X is the risk score in each polygon. Therefore, this equation tries to stretch the risk scores between 0 and 100. For example the equation for predicting of % risk of water erosion in each polygon is: $[(X - 12)/48] \times 100$.

4) Producing of risk maps of land degradation or desertification: The final map of risk of land degradation is produced by overlaying all four maps of degradation types. To qualify the severity classes of desertification map, the maximum degree of risk among the four types of land degradation shown in each polygon was selected.

4. Results and discussion

The estimates done on the basis of observations on the current status of land degradation reflect only what has happened till date. Risk assessment, on the other hand, is based on modeling, calculations, predicting the potentially adverse situation that may arise in 10 or 50 years from now (Bridge et al. 2001). Most studies so far done in Iran like Feiznia et al. (2001) and in the world like method of USLE for water erosion or Metternicht and Zinck (1997) for soil salinity have based their estimation on the 'present status' of degradation. It should be also said that selection of such parameters which calculate the desertification hazard is more comprehensive than previous studies. Because nearly all effective factors have just been calculated (Masoudi and Amiri 2013; Barzani and Khairulmaini 2013; Masoudi 2014). There exists also confusion in the use of the term 'risk assessment' among many scientists (e.g. Norton et al. 2001; Van Der Knijff et al. 2000 and Filho et al. 2001) who actually estimated only the soil loss, using some of the models like USLE and not the risk. The different type degradation maps alone based on the present status of degradation are inadequate to predict areas under risk. It requires a comparison between the present status and data showing state of degradation in the past to find the rate of degradation. This is almost difficult in most of the cases because of unavailability of such data of the previous decades. The present model using different indicators of land degradation types has an edge to solve this problem since it finds the severity of degradation using cumulative effect of all indicators and then compares it with the present status of degradation.

In the present work, the risk assessment of desertification attempts to demarcate areas with greater probability of reaching the worst step of degradation like a change from moderate to severe state of soil

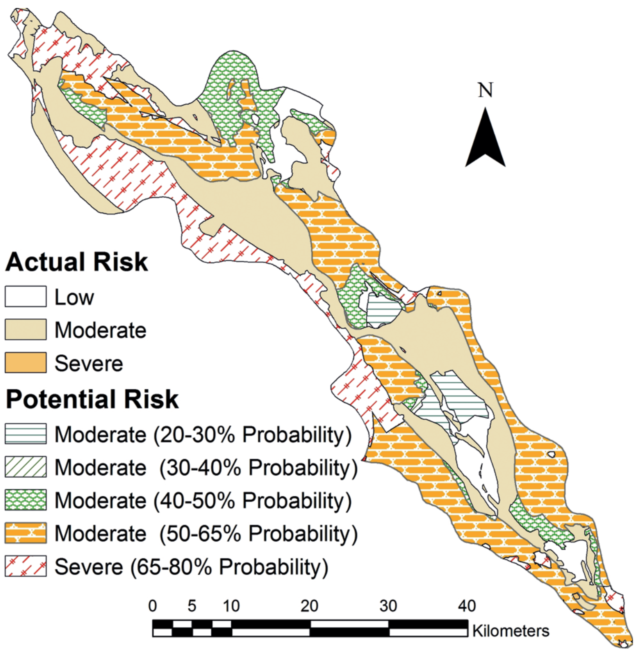

Fig. 2 Risk of water erosion in the Khane-Zenian & Siakh-Darengun sub basin.

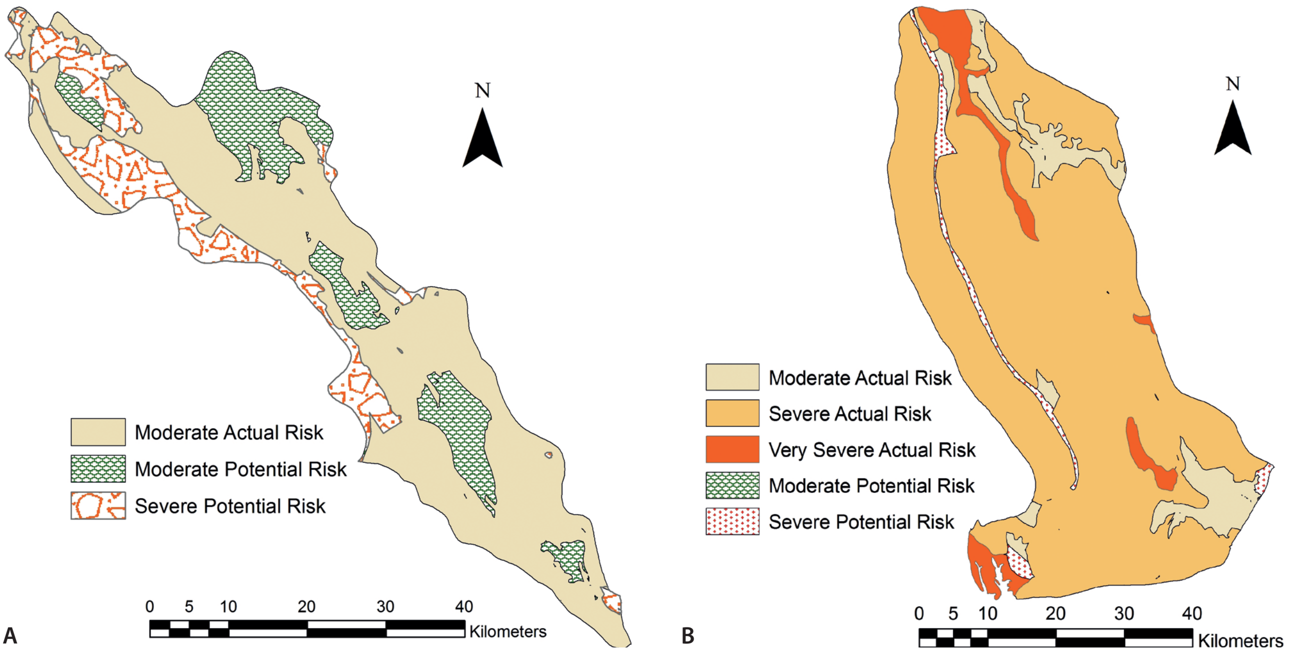

Fig. 3 Risk of desertification in the study area: (a) Khane-Zenian & Siakh-Darengun sub basin; (b) Khormuj sub basin.

salinization or other types of degradation and also measure the probability (risk) of this adverse change (Masoudi et al. 2005; Masoudi et al. 2006; Masoudi et al. 2007; Masoudi 2014). This kind of classification using two categories of 'actual risk' and 'potential risk' and its subclasses based on per cent probability in the risk maps is the first attempt of its kind for defining areas with higher risk of degradation. Preparing such risk maps may prove to be useful for regional planners, and policy makers for agricultural and environmental strategies, not only in the semi-arid and arid conditions of Southern Iran but also in other countries facing similar problem. The model can be made applicable for other countries only after little modification of some of the indicators, based on the local conditions. The risk map of water erosion is one example of this kind of methodology for assessing risk of land degradation types (Fig. 2).

To prepare the severity classes of desertification maps, the four types of desertification were overlaid in GIS. Once again from the desertification risk maps (Fig. 3) the areas under actual risk and areas under potential risk were identified. Those under actual risk have been further divided into the moderate and the severe risk classes but both define those areas in which the present state of degradation is moderate or severe for any one type of degradation. From the Fig. 3, it is concluded that in both sub basins areas under actual risk are more widespread compared to areas under potential risk. Among severity classes a greater proportion (63%) of land is under 'moderate risk' in the Khane-Zenian & Siakh-Darengun sub basin while the in the Khormuj it is 88% under 'severe and very severe risk'. This implies the obvious that the conditions in the Khormuj sub basin with arid climate are worse compared to the Khane-Zenian & Siakh-Darengun sub basin, with semi-arid climate. On the other hand, the vulnerable potential risk areas under the threat are more extensive in the Khane-Zenian & Siakh-Darengun (37%) compared to the Khormuj (3%). These results indicate that the already degraded lands with worse condition are lesser in the Khane-Zenian & Siakh-Darengun sub basin and therefore they need more attention for protection against future degradation. Also GIS analysis shows the main type of desertification in the plains and high lands of both sub basins is the vegetation degradation. This reflects the overall impacts of climatic and anthropogenic causes and soil degradation on the vegetation cover.

5. Conclusions

The Mond Basin model is the first attempt of its kind for defining the risk of desertification and can be made applicable for other areas in Iran and elsewhere. The main results of the present paper are:

1. The hazard maps of different indicators processed in the risk assessment model give a far better opportunity to distinguish the severity classes of risk of desertification. Creating some new indicators for the first time for this assessment was another achievement of the present work. Some of the new indicators are: ratio of non-irrigated areas to irrigated areas, 'Efficacy of surface geology' (ESG) and 'influence of carbonate formations'. To identify the severity classes of such indicators, the local statistical conditions of data have also been considered.

2. The model based on the statistical parameters helps to identify the areas under actual and potential risk and their sub classes based on per cent probability.

3. Areas under actual risk are widespread in both the sub basins especially in the Khormuj sub basin, indicating the more severe land degradation at present. The potential risk areas in the Khane-Zenian & Siakh-Darengun sub basin are more widespread compared to the Khormuj, indicating further threat of land degradation in future. Areas under potential risk will be the areas needing immediate attention for remedial measures for reclamation and conservation for each type of degradation.

4. Considering both actual and potential risk areas it is concluded that the areas under severe risk are dominant in the Khormuj while those under moderate risk have a greater spread in the Khane-Zenian & Siakh-Darengun. This shows between the two basins the overall environmental condition in the Khormuj sub basin is worse.

Acknowledgements

The authors are thankful to the Government Offices of Iran, for providing the data, maps and reports for this risk assessment work.

References

Barzani, M. M., Khairulmaini, O. S. (2013): Desertification risk mapping of the Zayandeh Rood Basin in Iran. Journal of Earth System Science 122(5), 1269–1282. https://doi.org/10.1007/s12040-013-0348-1

Bridges, E. M., Hannam, I. D., Oldeman, L. R., Penning de Vries, F. W. T., Scherr, S. J., Sombatpanit, S. (2001): Response to land degradation. Science publishers. INC.

Dregne, H. E. (1991): Human activities and soil degradation. New York: Marcel Decker.

Dubovyk, O., Menz, G., Conrad, C., Kan, E., Machwitz, M., Khamzina, A. (2013): Spatio-temporal analyses of cropland degradation in the irrigated lowlands of Uzbekistan using remote-sensing and logistic regression modeling. Environmental Monitoring and Assessment 185(6), 4775–4790. https://doi.org/10.1007/s10661-012-2904-6

EEA (1999): Environment in the European Union at the turn of the century, environmental assessment report No. 2. Copenhagen: EEA

Ekhtesasi, M. R., Mohajeri, S. (1995): Iranian classification of desertification method. In 2nd National Conference of Desertification and Combating Desertification Methods, Kerman, Iran.

Eswaran, H., Lal, R., Reich, P. F. (2001): Land degradation: an overview. In: Bridges, E. M., Hannam, I. D., Oldeman, L. R., Pening de Vries, F. W. T., Scherr, S. J., Sompatpanit, S. (eds.). Responses to Land Degradation. Proc. 2nd. International Conference on Land Degradation and Desertification, Khon Kaen, Thailand. Oxford Press, New Delhi, India.

FAO (1994): Land degradation in South Asia: its severity causes and effects upon the people. FAO, UNDP and UNEP report, Rome.

FAO (2002): Land Degradation Assessment in Drylands: LADA. FAO, Rome.

FAO/UNEP (1984): Provisional methodology for assessment and mapping of desertification. Food and Agriculture Organization of the United Nations, Rome.

Feiznia, S., Gooya, A.N., Ahmadi, H., Azarnivand, H. (2001): Investigation on desertification factors in Hossein-Abad Mish Mast plain and a proposal for a regional model. J of Biaban 6(2), 1–14.

Filho, C., Cochrane, T. A., Norton, L. D., Caviglione, J. H., Johansson, L. P. (2001): Land degradation assessment: tools and techniques for measuring sediment load. 3rd International conference on land degradation and meeting of the IUSS submission C- soil and water conservation. September 17–21, 2001, Rio de Janeiro, Brazil.

FRW (Forest, Range, and Watershed Management Organization) (2004): National action program for combating desertification and mitigation of drought impacts. Iran, Tehran.

Jafari, R. Bakhshandehmehr, L. (2013): Quantitative mapping and assessment of environmentally sensitive areas to desertification in central Iran. Land Degrad. Dev. 2013.

Kosmas, C., Poesen, J., Briassouli, H. (1999): Key indicators of desertification at the Environmentally Sensitive Areas (ESA) scale, The Medalus Project: Mediterranean Desertification and Land Use. Manual on Key Indicators of Desertification and Mapping Environmentally Sensitive Areas to Desertification. Project report. European Commission.

MA (2005): Millennium Ecosystem Assessment. Island Press, Washington, DC, USA.

Masoudi, M., Patwardhan, A. M., Gore, S. D. (2005): Some remedial measures for salt affected areas in parts of Mond Basin, Southern Iran. Nature Environment and Pollution Technology 4(2): 157–162.

Masoudi, M., Patwardhan, A. M., Gore, S. D. (2006): Risk assessment of water erosion for the Qareh Aghaj Sub Basin, Southern Iran. Stochastic Environmental Rese and Risk Assess 21, 15–24

Masoudi, M., Patwardhan, A. M., Gore, S. D. (2007): Risk assessment of lowering of ground water table using GIS for the Qareh Aghaj Sub Basin, Southern Iran. Journal of the Geological Society of India 70, 861–872.

Masoudi, M., Zakeri Nejad, R. (2011): A new model for assessment of erosion using desertification model of IMDPA in Mazayjan plain, Fars province. Iran. Ecol. Environ. and Conserv. 17(3), 1–6.

Masoudi, M., Amiri, E. (2013): Hazard Assessment of Current State of Vegetation Degradation Using GIS, A Case Study: Sadra Region, Iran. Journal of Ecology and Environment 36(1), 49–56. https://doi.org/10.5141/ecoenv.2013.006

Masoudi, M. (2014): Risk Assessment of Vegetation Degradation Using GIS. Journal of Agricultural Science and Technology 16, 1711–1722.

Masoudi, M., Vahedi, M. (2014): Hazard assessment of livestock pressure in Fars province, Iran. Ecology Environment and Conservation 20, 899–906.
Masoudi, M., Amiri, E. (2015): A new model for hazard evaluation of vegetation degradation using DPSIR framework, a case study: Sadra Region, Iran. Polish Journal of Ecology 63, 1–9. https://doi.org/10.3161/15052249PJE2015.63.1.001
Masoudi, M., Jokar, P. (2015): Land-use Planning using a Quantitative Model and Geographic Information System (GIS) in Shiraz Township, Iran. Ecopersia 3, 959–974.
Metternicht, G., Zinck, J. A. (1997): Spatial discrimination of salt and sodium-affected soil surfaces. International Journal of Remote Sensing 18(12), 2571–2586. https://doi.org/10.1080/014311697217486
Middleton, N., Stringer, L. C., Goudie, A. S., Thomas, D. S. G. (2011): The Forgotten Billion: MDG Achievement in the Drylands. UNDP, New York.
Norton, L. D., Filho, C. C., Cocherane, T. A., Caviglion, J. H., Fontes, H. M., Johnsson, L. P., Marenda, L. D. (2001): Monitoring the sediment loading of Itaipo lake and modeling of sheet and rill erosion hazards in watershed of Parana River: an outline of project. 10th international ISCO conference, May 1999.
Oldeman, L. R., Hakkeling, R. T. A., Sombroek, W. G. (1991): World map of the status of human-induced soil degradation: an explanatory note. Wageningen, International Soil Reference and Information Centre, Nairobi, UNEP, 27 p + 3 maps. Revised edition.
Ponce-Hernandez, R., Koohafkan, P. (2004): Methodological framework for Land Degradation Assessment in Drylands (LADA). FAO report: Rome.
Qi, F., Wei, L., Haiyang, X. (2013): Comprehensive evaluation and indicator system of land desertification in the Heihe River Basin. Nat Hazards 65(3), 1573–1588. https://doi.org/10.1007/s11069-012-0429-5
Reynolds, J. F., Grainger, A., Stafford Smith, D. M., Bastin, G., Garcia-Barrios, L., Fernandez, R. J., Janssen, M. A., Jurgens, N., Scholes, R. J., Veldkamp, A., Verstraete, M. M., von Maltitz, G., Zdruli, P. (2011): Scientific concepts for an integrated analysis of desertification. Land Degradation and Development 22, 166–183. https://doi.org/10.1002/ldr.1104
Sepehr, A., Zucca, C. (2012): Ranking desertification indicators using TOPSIS Algorithm. Nat Hazards 62, 1137–1153. https://doi.org/10.1007/s11069-012-0139-z
UNCCD (2011): Land and soil in the context of a green economy for sustainable development, food security and poverty eradication. The Submission of the UNCCD Secretariat to the Preparatory Process for the Rio+ 20 Conference, Revised Version 18 November 2011.
UNEP (United Nations Environmental Program) (1992): World Atlas of Desertification, editorial commentary by N Middleton and DSG Thomas. Arnold, London.
Van Der Knijff, J. M., Jones, R. J. A., Montanarella, L. (2000): Soil Erosion Risk Assessment in Europe. EUR 19044 EN.
Van Lynden, G. W. J., Oldeman, L. R. (1997): Assessment of the status of human-induced soil degradation in south and southeast Asia (ASSOD). International Soil Reference and Information Centre.
Vogt, J., Safriel, U., von Maltitz, G., Sokona, Y., Zougmore, R., Bastin, G., Hill, J. (2011): Monitoring and assessment of land degradation and desertification: towards new conceptual and integrated approaches. Land Degradation and Development 22, 150–165. https://doi.org/10.1002/ldr.1075
Zhang, L., Li, X. S., Yuan, Q. Z., Liu, Y. (2014): Object-based approach to national land cover mapping using HJ satellite imagery. Journal of Applied Remote Sensing 8(1), 083686. https://doi.org/10.1117/1.JRS.8.083686

The issues in methodology and data interpretation in studies of tourist attractions' attendance: annual passes

Lukáš Nekolný*, Dana Fialová

Charles University in Prague, Faculty of Science, Department of Social Geography and Regional Development, Czechia
* Corresponding author: lukas.nekolny@natur.cuni.cz

ABSTRACT

The attendance belongs to the basic quantitative indicators, which are used for assessment of the offer in the tourism. It should be used for monitoring of development, geographical structure, seasonality and from interpretational view also for monitoring of achievement or failure. This paper discusses, emphasizes and evaluates the very current topic of methodological differences in the data collection of tourist attractions attendance. The work focuses especially on the subtopic of annual passes on the example of zoos, which belong to the most visited tourist attractions worldwide and where the sale of annual cards is a typical service. Nevertheless, this article proves on the example of the zoos in German-speaking countries that the rise of attendance does not only have to show the reality. The total number of visits and the positive attendance development is in some zoos connected to the methodological approach.

KEYWORDS

methodology; tourism; attendance; annual passes; zoo

Received: 6 July 2017
Accepted: 19 June 2018
Published online: 3 September 2018

Nekolný, L., Fialová, D. (2018): The issues in methodology and data interpretation in studies of tourist attractions' attendance: annual passes.
AUC Geographica 53(2), 167–179
https://doi.org/10.14712/23361980.2018.17

1. Introduction

Tourism and recreation belong to the activities which determine the surrounding geographical area. This determination is mainly long-term, but might be short-term as well. It is connected to the maintenance and establishment of tourist attractions and infrastructure and also with a permanent occurrence of tourist players. For evaluation of tourism as an economic sphere and also for evaluation of its impacts we need to have quality data sources. Assessment, ratings and statistics are part of all human activities, including tourism and recreation. But the other branches (e.g. health care) have more quality data compared to tourism and culture (Borská 2013).

The attendance belongs to the basic quantitative indicators, which are used for assessment of the offer in this sector (Vystoupil 2007). It can be and it even should be used for monitoring of development, geographical structure, seasonality and from interpretational and pragmatic view also for monitoring of achievement or failure, which depends on growth respectively downward trend (Lew et al. 2004). High and increasing numbers of attendance can have positive effects but after reaching tolerable limits also the negative ones (e.g. situation in Barcelona) on media image and branding (Kašková, Chromý 2015; Matlovičková 2015) which affects the economic situation. Moreover, we can recognise in which part of the life cycle e.g. a tourist attraction is. It is important for its sustainability, which is a current parameter from a long-term view. After all, sustainable development of tourism is the topic of many articles and books, which were published in last decades (e.g. Swarbrooke 1998; Pásková 2014) and, furthermore, it is the basic topic of Journal of Sustainable Tourism.

But the data situation is not simple because the data base is relatively limited. In tourism, we have some information about collective accommodation establishments, nevertheless, this information is becoming insufficient because of a steadily increasing impact of shorter trips and journeys without staying overnight (Franke 2012). For this reason, it is also appropriate to check the attendance of tourist attractions. We need these data for many reasons thoroughly analysed in the next section of this article. In the statistics of this topic, there can be a lot of methodological limits, differences and distortions, which could create meaningless numbers without almost any importance. Published data can be therefore incomparable to each other. Because of this situation, there is a real need of correction and presentation of risks to professional public, creating of discussion and in the ideal case of establishing a unified system of measuring the attendance data. The first aim of this paper is to identify and introduce which basic methodological problems/risks exist.

The main aim of this study is to find, document and explain mutual deviations and deviations from reality and to highlight risks of interpretation by ignorance or lack of knowledge of data-producing methodology. This has been performed through different methods. The authors ask which influence can be caused by the diverse methods of counting the annual passes entries. Throughout tourist attractions with statistics of entries, zoos (zoological gardens and other animal-based attractions) have a very privileged position with one of the highest attendance worldwide (Dobroruka 1989; Baratay, Hardouin-Fugier 2004, Rees 2011; Fialová, Nekolný 2015). It is the first reason for using these institutions in the study. The second one is connected to zoos as appropriate institutions for explaining the methodological problems with annual passes.

2. Theoretical basis

Tourism generates significant economic benefits in many tourist destinations of different scale levels (locations, regions or countries), assuming proper management (Jakubíková 2012, Borská 2013). Some countries, typically small islands, are even dependent on the revenue from the tourism sector (UNWTO 2009; Fialová 2012). In contrast with other fields, in tourism there is a scarcity of high-quality data (Borská 2013). Not only has data knowledge crucial importance for several reasons, it also serves many purposes. For a destination of any scale level, attendance data are one of the basic performance indices (Vystoupil 2007). These data can be reflected in statistics of overnight stays in collective accommodation establishments (in Czechia, these statistics are secured by the Czech Statistical Office) and it can be said this information plays a prime role in academic literature (Vystoupil 2007). However, considering the current trends in tourism, individualisation and "authenticity" in private accommodation, which is not registered statistically (mainly due to a protection of individual data), is growing to be another significant variable in the overall rating. It is for this same reason that data responding to collective accommodation establishments are available only in municipalities with more than three pursued possibilities of accommodation (Fialová, Nekolný 2017).

Attendance can also be related to catering, cultural and sport establishments as well as to all sorts of tourist attractions (firstly Cohen in 1972; Zelenka, Pásková 2012). A correct understanding of the term "attendance" is absolutely essential for an accurate interpretation of data in tourism. According to Zelenka, Pásková (2012), attendance refers to the number of people who visit specific tourist attractions or destinations and who do so within some specific time range (typically one year). This is, nevertheless, a rather idealistic approach. It is the number of visits and not visitors that represents a more accurate perception of this concept (Smith 2013). This relates to the fact that the same visitor can return to the same attraction or destination repeatedly during the selected time (most commonly one calendar year). For the correct assessment of any tourism segment, it has to be taken into consideration that the same person can visit more facilities of the same kind, and therefore the total number of attendances does not equal to that specific share of human population. This "one person = one visit approach" is, unfortunately, used fairly often and contributes to incorrect statistics. Holtorf (2008), for example, says that zoos in *World Association of Zoos and Aquaria* (WAZA) are annually visited by over 600 million people, i.e. roughly 10% of the world's population (or more actually Gusset, Dick [2011] with 700 million people). This value is, nonetheless, only the sum of the entries in zoos under this organization and the fact that the same people can visit one specific zoo or more institutions more than once per year is not acknowledged. Consequently, under no circumstances the total number can be contrasted with the world's population.

The number of paid visits is essential for anyone engaged in tourism. It is, therefore, necessary to realize that one subject can manage more objects (tourist attractions) in more locations (e.g. National Museum). Statistics handled these criteria are valuable for aggregate management and economic view. On the other hand, attendance data for each tourist attraction separately hold significance mainly for social and geographical indicators. Collected attendance data about all branches of the National Museum (located in seven municipalities of four districts; annual report [AR] 2013) provide no information about their distribution and different position to other tourist attractions. The data about individual objects (tourist attractions) play a key role in proper destination management and also in economic factors such as different structure of visitors, specific needs of non-paying visitors (e.g. disabled people), who often greatly contribute to the overall attendance (e.g. zoos in Bratislava, Ostrava or Hluboká mention a 15% share [AR, emails]).

Attendance is one of the rudimentary scales of success/failure of tourist institutions, subjects, locations, and destinations (Vystoupil 2007; Zedková 2012). That is why each such subject aims to reach the highest attendance, highest number of visits, and if possible, the highest number of visitors who will return again (in this scenario the term *number of visits* instead of *visitors* is more easily justified). Studies emphasize that the returning visitors usually give positive feedback and recommendation of the destination to other potential clients (Jang, Feng 2007). These recommendations save money as finances spent on attracting new customers can be up to five times more expensive than retaining the existing ones (Rust, Zahorik 1993).

The high number of visits is an indicator of tourist destination attractivity and is reflected in their budgets. On the other hand, too high number of visits can have a negative impact in case of many tourist attractions (in particular historical sites or natural monuments) as the carrying capacity becomes exceeded. Zedková (2012) finds this as a significant factor in the decrease of the genius loci and the authenticity of the specific attraction. In addition, the concept of authenticity (firstly MacCannel in 1973; Wang 1999) is an unstable variable. The perception of authenticity is traditionally not applicable in case of zoos (Wang 1999). Pásková (2014) refers to the excess of carrying limits as the tourism trap effect, during which the tourism is depreciating itself by extending its own capacity for the sake of profit.

The attendance (and its seasonality) depends on many factors. The basic ones are geographical, like area and geographical position, site or localisation (Baratay, Hardouin-Fugier 2004). These are connected to settlement (residential and demographic factors and structure [Rees 2011]) and also to traffic infrastructure and the proximity of other tourist attractions (Mizicko, Bell 2001). In this context, Frost (2011) emphasizes that aquaria, for instance, are founded in places with the benefit of the existence of long-term, attractive destinations. Cultural predispositions (e.g. intercultural relationships, attitudes and connections with the environment) are also very important in different countries and regions (Davey 2007). For instance in the US and partly in Europe, there is a very high level of using zoos as the sites of recreation and family and social grouping (e.g. Reade, Waran 1996; Turley 2001; Woods 1998). Even weather plays its role (Rees 2011). In winter, the indoor attractions have bigger success than the outdoor ones, so there are some surveys exploring the parameters of weather and attendance as a complex matter (e.g. AR of Magdeburg Zoo). According to Rees (2011) the period from Easter to the end of summer holiday has a main influence on the balance of attendance in the temperate climate. The fact that different tourist attractions offer different activities makes it necessary to differentiate between more types of tourist attractions (in the case of Baratay, Hardouin-Fugier [2004] zoos) in a specific country or destination. This differentiation also plays a role in attendance levels. Tourist attractions can be very different (Kušen 2010) and they have distinct character, cause and the circumstances of their creation (Swarbrooke 2002). They have distinct area, capacity and attractivity (Vaníček 2012) and as a result also a distinct number of visits. Extraordinary events such as floods or disease also play an important role in the attendance levels.

High-quality statistics of the attendance have an application in the research, therefore in tourism exists the trend of decreasing average number of overnight stays (Franke 2012), shortening of journeys, and an increase in the importance of trips. The importance grows with the necessity of better knowledge about these shorter journeys without the use of accommodation. This obtained information also thanks to the attendance data throughout the visitor attractions.

These data are not compared only within a one-year period but mainly in trends. Namely, the important information about the position of any attraction on the market (and in the tourist-area life cycle) may be found based on trends (Davey 2007). Butler came with this concept (Lew et al. 2004) in 1980 and even though the reality is rather simplified by this approach and the model itself has a lot of problems (Palatková 2006), it is an appropriate model instrument for forecasting the future in any destination on any scale level.

Knowing the motivation and reasons behind visits of touristic attractions is crucial for the visitor management – therefore the visitor studies were founded (Mizicko, Bell 2001). For example, Ryan (2003) dealt with the typology of motivation. Motivation differs according to the type of attraction. Gelná and Fialová (2011) believe that experiences play bigger and more important role in leisure time. First visitor motivation studies were focused on museums. This approach could be also applied to zoos (Mizicko, Bell 2001). These particular studies also dealt with the issues of annual passes. Mizicko, Bell (2001) emphasize that visitor studies are becoming more and more useful in decision-making processes.

Aside from the aforementioned way of data usage, financial sources represent another important factor (Davey 2007). It can have a considerable impact on economic support, which goes hand in hand with high-quality marketing, promotion and a big role of media (Fialová, Nekolný 2017). If the attendance figures are high, it is much easier to find financial support from important sponsors, visitors, and public resources. The compilation of rankings (e.g. CzechTourism) can be perceived as a media support or a marketing tool. High rankings can put the spotlight on these attractions and ensure their brand being acknowledged (Kašková, Chromý 2015; Matlovičová 2015). On the other hand, it can be viewed as a natural process – benchmarking – comparing to competition (Zelenka, Pásková 2012) or interorganizational analysis (Holešinská 2010). The output of the comparison depends on the chosen methodology and hence cannot be realistic.

The attendance of many tourist attractions (e.g. monuments, natural attractions) can be perceived in connection with education or gaining some knowledge: especially informal, but in the last years also directly formal through tours, information leaflets etc. The attendance levels cannot be interpreted as a number of people who received some education or knowledge, since one can visit the same attraction more times. Consequently, the question of a precise interpretation is very problematic and overvalued (Smith 2013).

Accurate and complete data are necessary but, unfortunately, very difficult to obtain. The basic problem that statisticians in tourism face is the lack of data of other than paid services (Vystoupil 2007; Kruczek 2014). Except for this limitation, there are, furthermore, many methodological differences complicating any relevant data work. Some essential topics of methodological problems were defined thanks to the detailed and long-term study of research articles (e.g. Smith 2013), mainly annual reports and statistical reviews with methodological notes, and also through correspondence with the representatives of almost 40 zoos and the Hořovice Chateau (Nekolný 2016). A simple summary is stated below:

- Data per whole organisation/specific tourist attraction
- Data per whole attraction/specific tour routes, exhibitions
- Data per paid/all visitors/visits
- Data realistic per entries/throughout the coefficients
- Data per the different/the same time

These points are valid in general and they are very well applicable to the context of zoos. The topic of zoos from the view of tourism is not elaborated strongly enough in literature, as commented by Mason (2000) and Frost (2011). However, a very interesting article with methodological warnings was published by Davey (2007) in International Zoo Yearbook. This begins to emerge new term – zoo tourism (e.g. Mason 2000 in Journal of Sustainable Tourism). Hosey et al. (2013) say that zoos are full of animals as well as people. Poley (1993) highlighted the same, saying that there are more people than animals in zoos. Zoos belong to the most visited tourist attractions (e.g. Dobroruka 1989; Baratay, Hardoui-Fugier 2004; Fialová, Nekolný 2015). With the development of society, their function has changed, especially in the most popular ones. Originally, zoos were mostly places of entertainment (Carr, Cohen 2011). Turley (1999) talks about three basic roles of modern zoos. According to Poley (1993), Hediger talked about four aims of zoos. Very similar goals were mentioned by Carr, Cohen (2011), Rees (2011) and by Dobroruka (1989) in Czech literature. Therefore, the functions of zoos can be distinguished in these four points: entertainment, recreation, education, research and conservation of natural and cultural heritage.

If we talk about the topic of attendance, literature is mostly silent. This is a very problematic topic with differences in methodology. British zoo expert Anthony Sheridan (2016) is aware of this issue, too. In the zoo area there are many differences. The first constrain is connected to the fact that not every zoo belongs to paid attractions (e.g. wildlife parks [wildparks] in Germany) and in those cases, the statistics are very often missing. Some zoos have detached expositions and, strictly speaking, these are separate tourist attractions. As the most troubling theme the measuring of visits through the coefficients instead of real entries can be identified.

Annual (or seasonal) passes (in German *Jahreskarten*, in Czech *permanentky*) are offered only by some tourist and leisure attractions. The annual passes are very popular in areas which are worth visiting more times per year, typically in zoos (Vaníček 2012). Prototypical visitors of zoos as well as the most typical users of annual cards are families with children (Cain, Meritt 2007). Such visitors come mainly directly from the "zoo city" because the shorter proximity, the higher probability of more frequent visits. Annual passes can be very profitable for this kind of visitors. Nevertheless, the financial profit gained from holding an annual pass can vary significantly and the same applies for methodology. The main aim of this contribution, hence, aims to detect and identify the methodologically problematical procedures.

3. Methodology

The information and data come from an own long-term study of different attendance data sources, especially from hundreds of annual reports and statistical yearbooks. In some cases this information is used only as a source of numbers (e.g. attendance, attendance per annual passes, the number of annual passes) in selected years (e.g. AR Jihlava Zoo). In other reports, there are also very useful methodological explanatory notes or even longer texts with a more detailed evaluation, explanation and interpretation (e.g. AR Magdeburg Zoo). In some zoos, it was possible to compile long time series, which had been used in this contribution and parts of them are shown graphically (e.g. fig. 3 and 6). The profitability of annual passes can be calculated through the websites of tourist attractions. These data were used for critical comparative analysis of methodological approaches. The email correspondence with the representatives of nearly forty European zoos served as a very substantial source of information (Nekolný 2016). The emails were sent to 133 zoos in 17 countries, so the data and other information provided approximately one third of the accosted. The emphasis has been put on Czechia and German-speaking countries that reflect different approaches to the methodology in Europe. The amount and the details of information depend on the situation in the institutions (in smaller zoos there is often no detailed data basement, at least in the long-term view.). The communication has been under way in Czech, German and English since 2015. Addressed people work as directors, business managers and senior officers at departments of public contact, marketing and public relations, press officers or at the secretariat.

4. Significance of annual passes

The share of the annual pass entries in zoo attendance can reach several tens of percent a year (e.g. AR Cologne Zoo; AR Münster Zoo). Although the interest in this type of tickets is growing, it is not a new trend. Notwithstanding, Mulhouse Zoo, France reported in 1974 that about 25% of all visits were accomplished via annual cards (Baratay and Hadouin-Fugier 2004), this number can´t be generalized. The importance of annual passes in different regions, countries, and even for various tourist attractions and in different time periods differs depending on many factors. The number of inhabitants around an attraction (mostly in cities) and the city's connection with the transportation network play the main role. These factors are mainly geographical. Many zoos are visited by the same people multiple times a year, such as Zlín Zoo, which is located in the suburbs of Zlín (it is not a typical city zoo). In 2009, a survey was conducted in the zoo, finding that repeated visits within the same year were reported by 14% of respondents, even though their share of the attendance, naturally, had to be noticeably higher (multiplied per number of visits). The motivation hidden in repeated visits can be found in the zoo's attractive facilities. The annual cards are usually of less importance in smaller zoos and of larger importance in bigger ones. Another important factor is the size of the city, where the zoo is situated. Furthermore, according to Floriánová (2017) from Děčín Zoo, the wide range of annual passes prevents conflict situations at ticket offices.

Moreover, the role of annual cards most likely depends on the political, historical and economic situation. Identifying the share and the importance of these long-term passes is dependent on the use of the proper methodology. Since there is no unified methodology for arriving at results, this fact becomes the fundamental problem which hinders the comparability of published data. According to recent information (e.g. Goldner 2014), we can talk about a problem that has existed for many years and has been gradually becoming more significant, especially in the last few years. Therefore, this topic deserves a professional research.

5. The same coefficient in different situations

The most important zoos in German-speaking countries are associated in an organisation called Verband der Zoologischen Gärten (VdZ). This professional association implemented a specific methodology of including the annual passes through coefficients many decades ago. The aim of this approach was to achieve high level of comparability across these zoos (Dommes 2015; Kanton Basel-Stadt 2015). According to this VdZ-coefficient (VdZ-Schlüssel in German), the annual card for one person is counted automatically as 20 entries in the attendance number. The passes for up to four people (such as families and sponsors) are counted as 80 entries (e.g. Basel Zoo 2015a). In the last years the methodology was criticised due to the overestimation of attendance (Goldner 2014). Although the average number of twenty entries per one person is shown to be usually too high, it is difficult to estimate the right number. The particular situation is influenced by many factors (see part 2 and 4) including the profitability of the pass – by what number of it is cheaper than by buying per partes/individual entries. This essential difference was acknowledged by VdZ (VdZ 2014). For basic orientation in different approaches/attitude of zoos see table 1.

For instance, the Apenheul Primate Park in the Netherlands offers a season pass, pays for itself by the second entry (Apenheul website 2016), whereas the zoo in Kraków, Poland, sells annual cards with a financial advantage for every twelfth visit (Kraków Zoo website 2016). For the majority of cases, an annual card makes sense after between the fourth and fifth, eventually sixth entry (for German-speaking countries offers this comparison Schüling, Altefrohne [2018], in Czechia e. g. Ostrava Zoo, Pilsen Zoo, Zlín Zoo). We have to also know that the particular type of cards can become favourable at different number of entries. For example, in Nordhorn Zoo (Germany), in case of an adult card it is worth buying by the fourth entry, even though for children it is by the fifth one and for families already at the third visit (Nordhorn Zoo website 2016). This topic of different favourability in selected European zoos is analysed in detail with more examples in table 1. Also the offers vary – there are transferable annual tickets (e.g. Liberec Zoo – see in the table 1), but the majority of the cards are passes non-transferable from person to person (linked to a specific name, often with a photo of the holder). Annual cards with unlimited number of entries are typical in Germany. In contrast, limited passes are quite common in Czechia (e.g. Jihlava Zoo, for other examples see in the table 1).

On the basis of these data we can say the share of the annual passes in the annual attendance depends on the favourability/profitability, geographical location as well as on the calculation methodology. In case of zoos in central Europe with accurate inclusion of the visitors it is possible to talk about a share of up to 10% of the total number of visits. In Western Europe the numbers are higher. It is probably a result of historical, cultural and political development along with the settlement structure. However, it could be a very useful and interesting topic of future research. In many Czech zoos the share is even lower than the aforementioned 10% – e.g. only ten passes in Hodonín Zoo in 2016 (Uhrová 2017). But it seems to be an extreme case. More often we can find the portion of about four percent (e.g. Ústí nad Labem Zoo [Balejová 2016], Brno Zoo [Vavřinová 2016], Jihlava Zoo app. 3% [Mrázková 2016]). The institutions using the VdZ-coefficient have a higher share of annual card

Tab. 1 Comparison of profitability of annual cards for adults in selected European zoos, March 2017.

Zoo	Country	Currency	Price of day-ticket	Price of annual card	Price ratio	Economical from entry no.	Note
Basel Zoo	Switzerland	CHF	21	90	4.29	5	non-transferable
Zürich Zoo	Switzerland	CHF	26	130	5.00	6	non-transferable
Berlin Zoo	Germany	EUR	14.5	49	3.38	4	non-transferable
Cologne Zoo	Germany	EUR	19.5	85	4.36	5	non-transferable
Leipzig Zoo	Germany	EUR	17	76	4.47	5	non-transferable
Magdeburg Zoo	Germany	EUR	13	55	4.23	5	non-transferable
Hellabrunn Zoo, Munich	Germany	EUR	15	49	3.27	4	non-transferable
Dvůr Králové Zoo	Czechia	CZK	195	590	3.03	4	non-transferable
				1600	8.21	9	transferable, 10 entries
Liberec Zoo	Czechia	CZK	120	850	7.08	8	transferable, 15 entries
Ostrava Zoo	Czechia	CZK	110	550	5.00	6	non-transferable
Pilsen Zoo	Czechia	CZK	150	700	4.67	5	non-transferable
Prague Zoo	Czechia	CZK	200	700	3.50	4	non-transferable, 12 entries
				1350	6.75	7	non-transferable, "No limit" – unlimited
Zlín Zoo	Czechia	CZK	130	600	4.62	5	non-transferable, additional benefits
Copenhagen Zoo	Denmark	DKK	180	440	2.44	3	non-transferable, additional benefits
Kraków Zoo	Poland	PLN	18	200	11.11	12	–
Vienna Zoo	Austria	EUR	18.5	44	2.38	3	non-transferable

Source: Own processing based on the zoo websites.
Note: This table shows annual cards without discounts (for adults). Some zoos offer two types of annual passes (e. g. Prague Zoo, Dvůr Králové Zoo).

entries, in some cases even higher by an order. E.g. Halle Zoo (ca. 30%; Bernheim 2016), Hellabrunn Zoo in Munich (2014: 43%; AR 2014), Dresden Zoo (2014: 56%; Marx 2015; 2013: 65%; AR 2013) or Basel Zoo (2014: 71%; AR 2014). The last mentioned institution reported a decrease of this share after the change of methodology to "mere" but more realistic nearly 45%.

Likewise, some smaller German zoos which are not a member of the VdZ use the coefficient – e.g. Zittau Zoo (2014: 19%; Großer 2016). The zoos offering the family annual cards (automatic coefficient: 4 × 20 = 80 entries) tend to achieve the highest share (e.g. Basel Zoo). Therefore it is possible, due to family cards, to err towards an even higher overestimation than due to passes for one person because the whole four-part family does not have to be complete in all visits. Consequently, we wanted to find out the average number of entries made through annual cards in surveyed zoos. Here are the basic findings:

1) The average number of entries per one annual pass does not usually exceed the value ten. This number is often only slightly lower than this value (2014: e.g. Jihlava Zoo 9.46 entries; Brno Zoo for unlimited passes 9.01 or 2015: Ústí Zoo 9.26). However, these data cannot be generalized. In Marlow Bird Park in Germany the season ticket is favourable to buy for at least three visits. Therefore there are only 3.1 entries per annual card on average (2015; Gereit 2016).

2) Some zoos are unable to say how many entries were done through the annual cards. These are not only zoos using coefficients but also other institutions – e.g. zoos in Hluboká, Liberec, Olomouc, Ostrava (Czechia) or Delitzsch (Germany). Břečková (2016), the press officer in Olomouc Zoo, says that this information is not important for this zoo. A different approach can be seen in Ústí nad Labem Zoo, where they write down lines by hand for the accurate numbers (Balejová 2016). A very unique situation is to be found in Vyškov Zoo. The deputy director Nepeřená (2017) talks about the annual cards as something outside the statistics – people with these passes are not included in the number of visits. The zoo only has internal statistics about the number of sold passes. Therefore, it can be said that virtually every zoo uses a different approach to this topic and it is an important question in which cases we can find significant or insignificant differences.

6. Recent changes in methodology

In 2013, according to the recommendation brought by the criticism of coefficients used in German-speaking countries, some zoos have changed the coefficient from 20 to 10 entries, starting from 2014 (e.g. Heidelberg Zoo; Heck 2016). Although the attendance number in 2014 was lower than in 2013, the media talked about a record number of visits (Knopik 2015). But in 2014 the attendance was actually higher, only the change of the methodology caused the smaller number. Similar situation happened in Münster Zoo (Germany) – its director said that he was pleased more with the number 618 thousand than with the 915 thousand (VdZ 2016) – it was caused throughout the methodology. In summer 2015 VdZ decided to retire the united VdZ-coefficient, which had been used for decades. Each zoo has to decide, how to approach this change (Dommes 2015). So now we can encounter more methodologies than before. However, each zoo has to take different factors into account. E.g. Neunkirchen Zoo did not use the coefficients and therefore there was not a need for change in methodology (Andres 2015). There is a very paradoxical situation in Austria since Salzburg Zoo and Innsbruck Alpenzoo publish two different attendance numbers. The first number can be found in Austrian tourist destination rankings (AustriaTourism 2015) and the second one – which uses the coefficients – in VdZ website and tables (VdZ 2016). Thus data are dependent on their source and one who is not familiar with the situation cannot grasp the reason of these dissimilar values.

Some zoos are turning to new ways of data acquisition – often via use of turnstiles. The turnstiles are connected with the methodology of "person-entries", which means each pass through the entrance turnstiles is counted as an entrance of one person. This principle is used in Prague Zoo or Artis Royal Zoo in Amsterdam (Macháčková 2015; Sloet 2016). These data are comparable with the majority of Czech zoos and relatively recently with the German zoos which gave up using the coefficients. And this marks a positive trend in comparability with other zoos and tourist attractions. In tourism, the knowledge of the development is the most important aspect and as a result of that we also need to know the attendance calculated via the methodology used long-term (albeit it could be inappropriate). Other way of solving this problem can be found in Basel Zoo where the number for 2015 (according to the new methodology) was compared to the attendance for 2005 which was recalculated using the annual cards average of the year 2015 (AR 2015).

The figures 1 and 2 show that the coefficient causes noticeable deviations from the reality in case of selected zoos. These are located in bigger cities or in areas with the higher population density. Such zoos usually have a high level of attendance and also a considerable number of people buying the annual passes. Therefore the number of visits grows faster in such cities and mainly in Germany, Austria and Switzerland this development was multiplied by methodology – throughout the overstated coefficients.

7. The annual passes and their trends

The increasing attendance in some zoos can be assigned to inaccurate and unrealistic

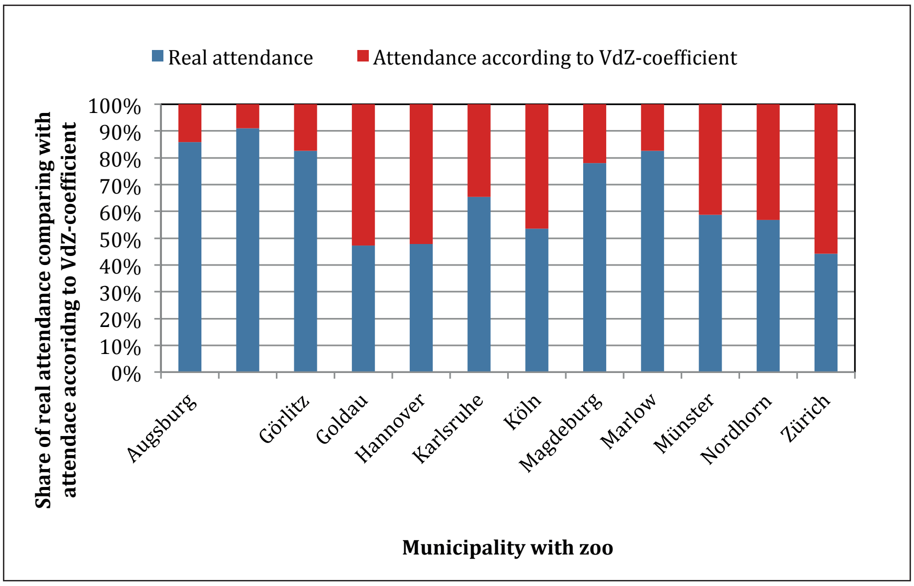

Fig. 1 Comparison of relative attendance according to different methodologies, selected member zoos of VdZ 2014.
Source: Own processing based on VdZ 2016.

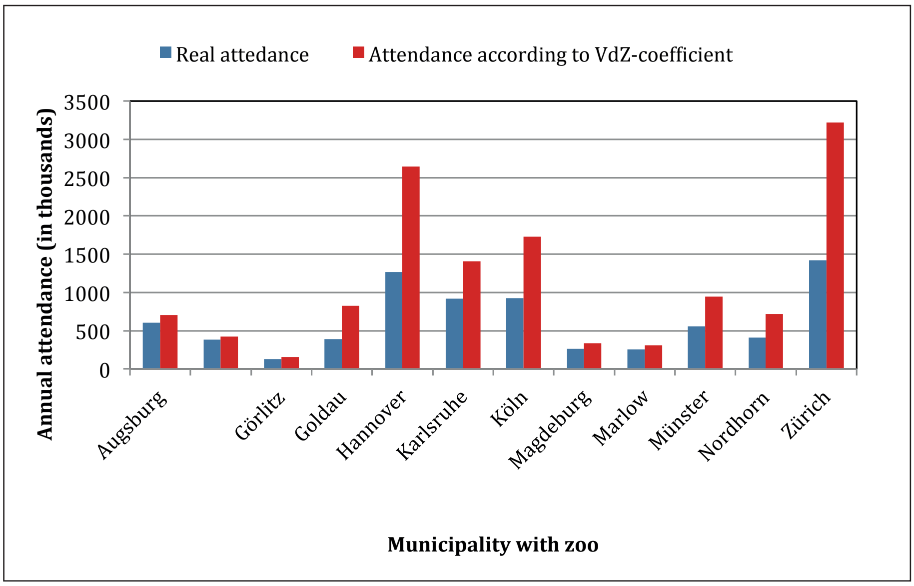

Fig. 2 Comparison of absolute attendance according to different methodologies, selected member zoos of VdZ 2014.
Source: Own processing based on VdZ 2016.

methodological procedure. The interest of visitors in the annual passes is very different and it also depends on the selected region or country. In Czechia, the sales of this kind of tickets are lower compared to Germany and the Western Europe. Hence we suggest more detailed analysis aimed at finding the cause and the development of this phenomenon. In this article we can only mention a couple of important examples, especially from Switzerland. In the late ’80s and ’90s of 20th century some zoos saw a large increase in sales of annual cards but others did not (adhere to this trend). This was, for instance, the case of Hellabrunn Zoo in Munich (1987: 1492 pc, 1992: 1612 pc, 1997: 1521 pc). This zoo of worldwide importance has experienced extreme surge of interest since 2009. From 2009 to 2014 the number of annual pass owners grew tenfold (from 4310 pc in 2009 to 44 697 pc in 2014; fig. 3). It is likely that the knowledge about this offer is increasing.

In other zoological gardens the increasing trend are not that noticeable, but in the majority of examples positive development can be seen (fig. 4). For example, the number of season passes sold increased 2.5 times in Zoo Magdeburg during the years 2009–2014.

Established system of coefficients has significant disadvantages also due to the reason that this system

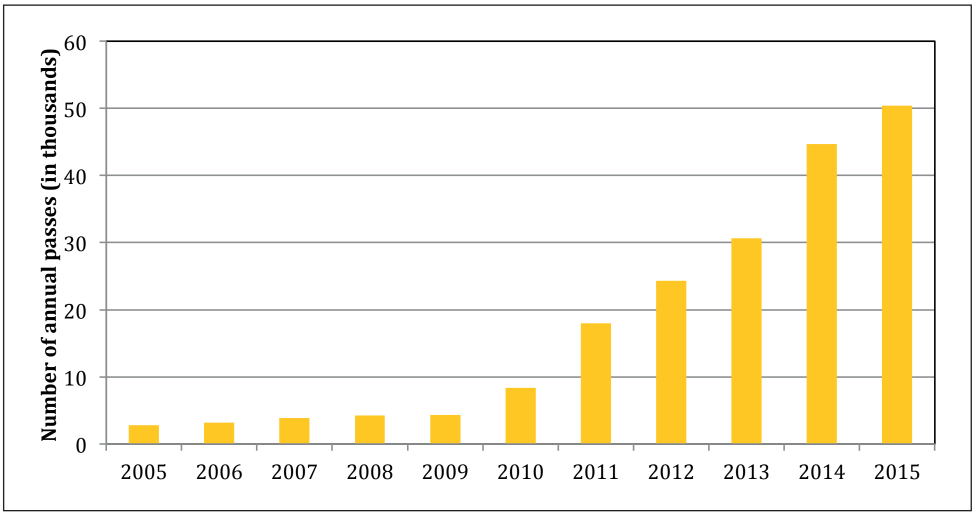

Fig. 3 The development of annual cards in Hellabrunn Zoo, Munich (D), 2005–2015.
Source: Own processing based on AR of Hellabrunn Zoo in 2005–2015.

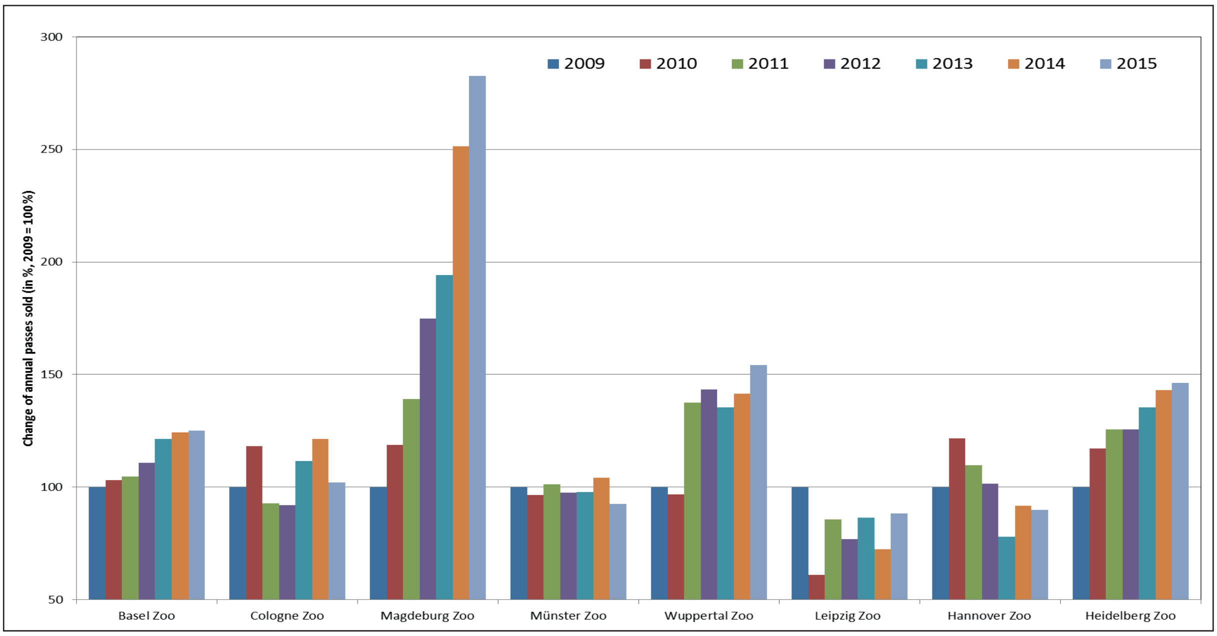

Fig. 4 The development of sales of annual cards in German and Swiss zoos with available data, 2009–2015.
Source: Own processing based on AR of zoos and email correspondence with representatives of zoos.

is not able to answer the basic questions about visitors. These questions represent the key for visitor management. Knitter (2016) from the zoological and botanical garden Wilhelma Stuttgart especially emphasizes these questions:

– How often the owners of the annual cards go to the zoo?
– How many children go to the zoo as a part of a family (per family tickets)?
– How many children under the age of six (entrance free) go to the zoo?
– How many visitors use the free tickets?
– How many students take part in the educational programs?
– How many visitors visit free special events?

So the coefficient calculation causes many difficulties, including the unawareness of the daily attendance. In this case daily data include visitors buying one day tickets – non-paying visitors are then calculated as 5% of the paying visitors. And the number of entries via annual cards is calculated only in the annual attendance.

The other methodical problem is differing lengths of validity. Most often it is one year (12 months) after the purchase of the pass (e.g. Münster Zoo). In Ostrava Zoo can be seen similar situation, but as Šoupalová (2016) emphasizes, the annual pass purchased before Christmas can be activated in June of the next year

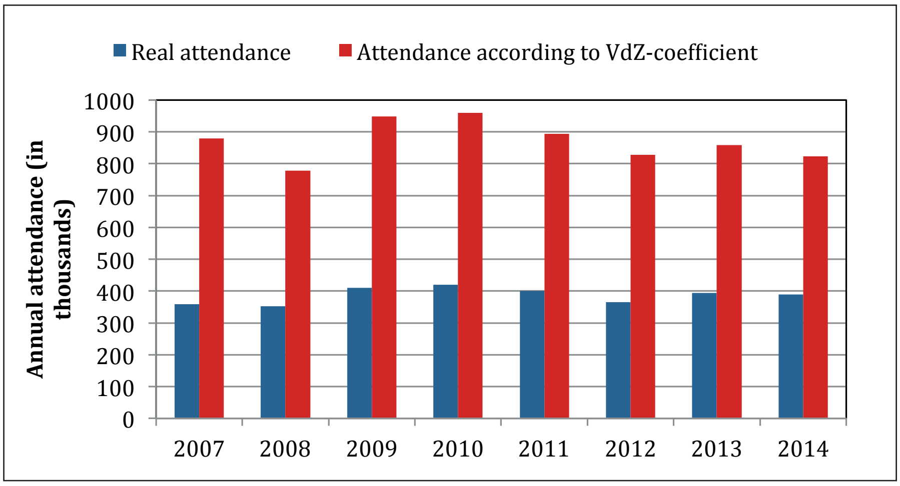

Fig. 5 Trend of attendance according to different methods, Goldau Zoo (CH), 2007–2014.
Source: Own processing based on Hürlimann (2010), AR of Goldau Zoo in 2009/10–2014/15.
Note: Methodological problem pertaining to Goldau Zoo: data from 1 April to 31 March of the next year – other zoos have data per calendar year.

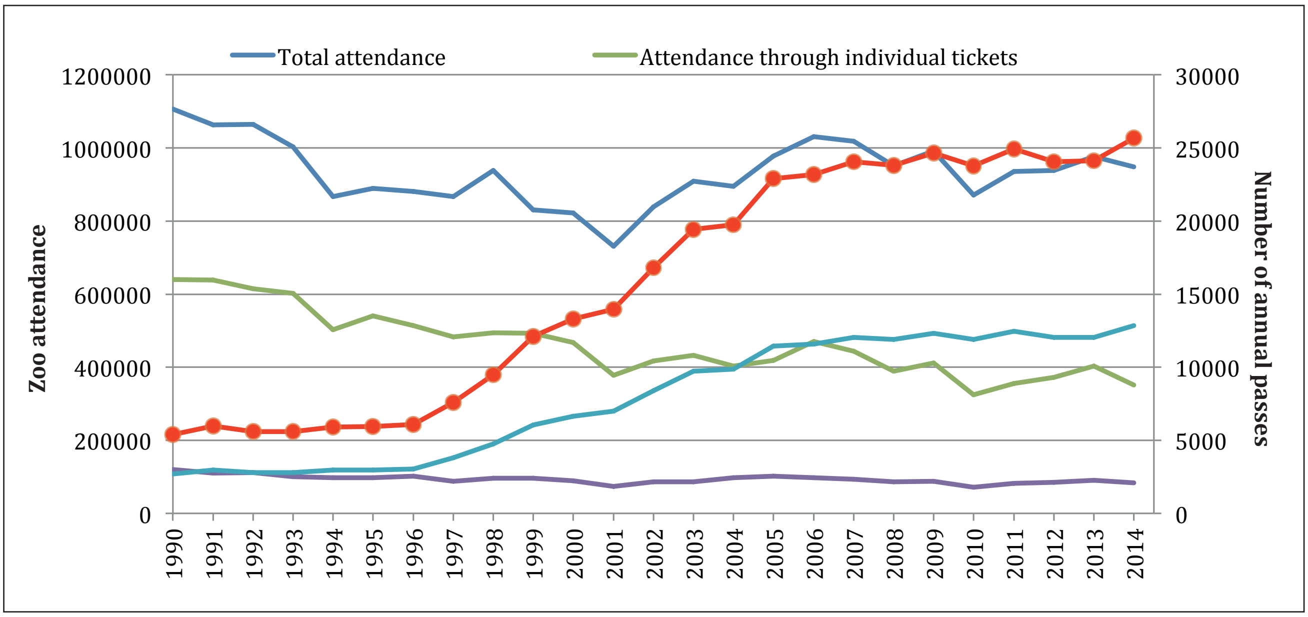

Fig. 6 Development of attendance by type of ticket, Münster Zoo (D), 1990–2014.
Source: Own processing based on data from Jahresstatistik Stadt Münster 2014.
Note: Decrease in 2001 below 800 thousand visits was caused by an outbreak of foot and mouth disease.
The zoo had to be closed for 20 days in spring
Methodological note: The number of visits with the annual passes is calculated for families as four people × 20 = 80 entries. Thanks to this procedure we can compare all the types of annual passes (for families and individuals) among each other. Hence the number of sold annual tickets differs from the number in the fig. 6. Members of zoo society (Zoo-Verein) are also included in the annual passes chapter (Münster Zoo website 2016; AR 2000).

and it is the start of the 12 months period. If we calculate the annual cards via coefficients, a significant problem arises because both new and old ones (from the last year) are valid. In other cases the validity of the passes is constrained by calendar year or season (e.g. Apenheul Primate Park in the Netherlands, Pairi Daiza in Belgium). In Basel Zoo the long-term ticket is valid to 15 March of the next calendar year (Basel Zoo website).

The introduced problems are possible to demonstrate also in the case of famous Swiss zoos. At first, zoo in Zürich established a new electronic system of evidence in 2013. The resulting number of visits in this year exceeded one million (precisely 1,079,919). However, using the previous method (via coefficients) yielded a number that was nearly twice the result of the usage of the electronic system – 2,003,043 entries (AR Zürich Zoo 2013; Goldner 2015). Nearly the same situation happened in Basel Zoo. The result of these radical changes has to be reflected in the interpretation of the results and comparison with other European zoos.

Goldau Zoo encountered similar problems and changes. This zoo exceeded the level of 300 thousand visits immediately after including children under six years of age in 1994 (Hürlimann 2010). The VdZ-coefficient was first used there in 2006 and after this change the published attendance surpassed the level of 800 thousands entries annually. Since financial year 2014/2015, when the former methodology was re-adopted the attendance decreased below 400 thousands entries. For the differences between the results of the methodology see fig. 5.

8. Coefficient – risk of influencing a trend

If we talk about the annual passes, we have to think about the trend of this segment as well as about the trends of the attendance of the other visitor groups in context of the total numbers of visits. The question is whether the used coefficient can affect the attendance trends? On the basis of the example of Münster Zoo (Germany) it is clear that it could be possible. Fig. 6 shows the total attendance fluctuated mostly around 900 thousand or one million entries during the 25 years 1990–2014. From the long term view the numbers are relatively stable. However, the term stability is not applicable to the cases of different visitor groups – there were plenty of changes. The number of visits made with traditional one-day tickets declined from 650 thousand to 350 thousand (Stadt Münster 2015). The sum of group tickets decreased as well, but not that dramatically. The most crucial change came in case of annual passes – their number and officially also the number of entries made with them increased during these 25 years fivefold (fig. 6). Similar development was found in other German or Swiss zoos (see previous chapters). And where is the problem hiding? In 2014 Münster Zoo changed the methodology. Using the former VdZ coefficient the attendance reached nearly 948 thousand entries, but the electronic calculation yielded only around 556 thousand of visits (VdZ 2016). Such big difference is important for the interpretation, just as in aforementioned Swiss zoos. Thus in Münster there really was a significant decrease of attendance – no stability to speak of. At the beginning of the 1990s the number of one-day tickets exceeded the number of 700 thousand entries which means approximately about 150 thousand more than was the total attendance in 2014. Therefore the analysis and evaluation of these data and the subsequent practical response should be one of the most important challenges for the destination management.

9. Conclusion

The goal of this paper is to highlight and evaluate the very current topic of methodological differences in the data collection of tourist attractions attendance. The work focuses especially on the subtopic of annual passes. This theme is exemplarily shown in detail on the example of zoos, where the sale of annual cards is a typical service. It is also a possibility how the interest in repeated visits as well as in the whole zoo can be increased and how the attendance can be enhanced. Nevertheless, this article proves on the example of the zoos in German-speaking countries that the rise of attendance does not only have to show the reality and the visitor financial profitability (in the situation of repeated visits). This positive development is in some zoos connected to the methodological "conception". It could be true that both trends operate at the same time but it is not clear if both influences are convergent. The detailed example of the attendance development in Münster Zoo shows that use of the coefficients can modify the trend – whether it is growing or declining. Publisher data aren't comparable automatically, it is necessary to know their methodology and to assess if and how we can work with them. The possibilities of methodological purifying are sometimes limited but only with the emphasis on the data purity and reality relevant results can be reached. This methodological approach could also be used for other tourist destinations and cultural institutions, especially in the boundary areas with entry fees and people motivation for repeated visits. If the explorer does not know (at least) any part of the methodology or anything that demonstrates the knowledge of the context, he exposes himself to the risk of an inappropriate interpretation. Unfortunately, this situation is known from the media as well as from important statistical institutions and from time to time also directly from explorers.

The annual cards coefficients have been brought many years ago, in the time when there was lack of technical monitoring equipment, a specific comparability. However, these coefficients have ignored the geographical, time and other differences among destinations and attractions. This is exactly something that influences the attendance as well as the importance of the utilization of annual passes. Additionally, the coefficients are in most of the examples unsuitable – they usually overvalue the real situation. For these reasons the basic recommendation is always to collect data more accurately, to check the methodology, to use only the real(istic) data and not to use the coefficients if it is possible and not necessary.

References

Baratay, E., Hardouin-Fugier, E. [translated by Oliver Welsh] (2004): ZOO: a history of zoological gardens in the West. Pbk. ed. London: Reaktion books.

Besucherzahlen von Sehenswürdigkeiten 2013 (2015). In: AustriaTourism. http://www.austriatourism.com/wp-content/uploads/2014/07/2014e_besucherzahlen-von-sehenswuerdigkeiten-2013-mit-wien.pdf (15 March 2015).

Borská, K. (2013). Marketing státních hradů a zámků. Master thesis. Praha, Vysoká škola ekonomická v Praze. Fakulta podnikohospodářská.

Cain, L., Meritt, D. (2007): The Demand for Zoos and Aquariums. Tourism Review International [online] 11(3), 295–306. http://openurl.ingenta.com/content/xref?genre=article,issn=1544-2721,volume=11,issue=3,spage=295 (10 March 2017).

Carr, N., Cohen, S. (2011): The Public Face of Zoos: Images of Entertainment, Education and Conservation. Anthrozoös 24(2), 175–189, https://doi.org/10.2752/175303711X12998632257620.

Davey, G. (2007): An analysis of country, socio-economic and time factors on worldwide zoo attendance during a 40 year period. International Zoo Yearbook 41, 217–225, https://doi.org/10.1111/j.1748-1090.2007.00007.x.

Dobroruka, L. (1989): Zoologické zahrady. Státní pedagogické nakladatelství, Praha.

Fialová, D. (2012): Cena za cestovní ruch: přínosy versus ztráty: vzdělávací modul geografie: výukový a metodický text: Přírodní vědy a matematika na středních školách v Praze: aktivně, aktuálně a s aplikacemi. P3K, Praha.

Fialová, D., Nekolný, L. (2015): Rub a líc turistické přitažlivosti. Geografické rozhledy 24(4), 8–9.

Fialová, D., Nekolný, L. (2017): Řečí čísel cestovního ruchu. Geografické rozhledy 26(3), 22–23.

Franke, A. (2012): Statistiky cestovního ruchu. Wolters Kluwer Česká republika, Praha.

Frost, W. (2011): Zoos and Tourism: Conservation, Education, Entertainment?. Channel View Publications, Buffalo.

Gelná, T., Fialová, D. (2011): Changes in the current Czech and German society and their effects on free time spending. AUC Geographica 46(2), 55–66, https://doi.org/10.14712/23361980.2015.31.

Goldner, C. (2014): Zoos manipulieren Besucherzahlen. Humanistischer Pressedienst. http://hpd.de/artikel/10124 (20 January 2015).

Goldner, C. (2015): Eingesperrte Tiere angaffen? Nein danke!. 6. Norderstedt: BoD - Books on Demand.

Gusset, M., Dick, G. (2011): The global reach of zoos and aquariums in visitor numbers and conservation expenditures. Zoo Biology 30(5), 566–569, https://doi.org/10.1002/zoo.20369.

Holešinská, A. (2010): Destinační management jako nástroj regionální politiky cestovního ruchu. Dissertation thesis. Brno, Masarykova univerzita, Ekonomicko-správní fakulta.

Holtorf, C. (2008): Zoos as Heritage: An Archaeological Perspective. International Journal of Heritage Studies 14(1), 3–9, https://doi.org/10.1080/13527250701711994.

Hosey, G. R., Melfi, V., Pankhurst, S. (2013): Zoo animals: behaviour, management, and welfare. Oxford University Press, Oxford.

Hürlimann, M. (2010): Der Natur- und Tierpark Goldau, Die Entstehung und Entwicklung zum modernen Zoo. Schwyzer Hefte 95, Verl. Schwyzer Hefte, Schwyz.

Jakubíková, D. (2012): Marketing v cestovním ruchu: jak uspět v domácí i světové konkurenci. Grada, Praha.

Jang, S., Feng, R. (2007): Temporal destination revisit intention: The effects of novelty seeking and satisfaction. Tourism Management 28(2), 580–590, https://doi.org/10.1016/j.tourman.2006.04.024.

Kašková, M. Chromý, P. (2015): Přitažlivost „značkových" míst. Geografické rozhledy 24(4), 8–9.

Knopik, M. (2015): Ein tierisch gutes Jahr für den Heidelberger Zoo. Heidelberg24. http://www.heidelberg24.de/heidelberg/heidelberg-rekord-heidelberger-zoo-viele-besucher-zuvor-4610766.html (20 January 2016).

Kruczek, Z. (2014): Frekwencja w atrakcjach turystycznych. Polska Organizacja Turystyczna, Kraków – Warszawa.

Kultur, Sport, Freizeit (2015). Kanton Basel-Stadt. http://www.statistik.bs.ch/zahlen/tabellen/16-kultur-sport-freizeit.html (19 February 2016).

Kušen, E. (2010): A system of tourism attractions. Tourism – An international Interdisciplinary Journal. 58(4), 409–424.

Lew, A. (ed.), Hall, C. M. (ed.), Williams, A. M. (ed.) (2004): A companion to tourism. MA: Blackwell, Blackwell companions to geography, Malden.

Mason, P. (2000): Zoo Tourism: The Need for More Research. Journal of Sustainable Tourism, 8(4), 333–339, https://doi.org/10.1080/09669580008667368.

Matlovičová, K. (2015): Značka územia. Prešov: Prešovská univerzita, Grafotlač. Mizicko, L., Bell, C. E. (eds.) (2001): Encyclopedia of the world's Zoos. Fitzroy Dearborn Publishers, Chicago.

Nekolný, L (2016): Problémy metodiky a interpretace dat návštěvnosti turistických cílů. Diploma thesis. Praha, Univerzita Karlova. Přírodovědecká fakulta.

Palatková, M. (2006): Marketingová strategie destinace cestovního ruchu: jak získat více příjmů z cestovního ruchu. Grada, Praha.

Pásková, M. (2014): Udržitelnost cestovního ruchu. Gaudeamus, Hradec Králové.

Poley, D. (1993): Berichte aus der Arche: Nachzucht statt Wildfang.: TRIAS, Stuttgart.

Reade, L. S., Waran, N. K. (1996): The modern zoo: How do people perceive zoo animals? Applied Animal Behaviour Science 47(1), 109–118. https://doi.org/10.1016/0168-1591(95)01014-9.

Rees, P. A. (2011): An introduction to zoo biology and management. Wiley-Blackwell, Hoboken, NJ, https://doi.org/10.1002/9781444397840.

Rust, R. T., Zahorik, A. J. (1993): Customer Satisfaction, Customer Retention, and Market Share. Journal of Retailing 69(2), 193–215, https://doi.org/10.1016/0022-4359(93)90003-2.

Ryan, C. (2003): Recreational tourism: demand and impacts. Channel View Publications, Aspects of tourism, 11, Clevedon.

Sheridan, A. (2016): Europas Zoo unter der Lupe: Sheridans Handbuch der Zoos in Europa. Schüling, K. (Verlag), Münster.

Schüling, K., Altefrohne, T. (2018): Eintrittspreise. Tiergarten (2), 62–70.
Smith, L. (2013): Visitors or Visits? An Examination of Zoo Visitor Numbers Using the Case Study of Australia. Zoo Biology 32(1), 37–44, https://doi.org/10.1002/zoo.21013.
Stadt Münster. Jahres-Statistik 2014 (2015). http://www.stadt-muenster.de/stadtentwicklung/publikationen.html (25 February 2016).
Swarbrooke, J. (1998): Sustainable tourism management. CABI publishing, Wallingford.
Swarbrooke, J. (2002): The development and management of visitor attractions. Butterworth-Heinemann, Boston.
Turley, S. K. (1999): Conservation and Tourism in the Traditional UK Zoo. The Journal of Tourism Studies 10(2), 2–13.
Turley, S. K. (2001): Children and the demand for recreational experiences: the case of zoos. Leisure Studies 20(1), 1–18, https://doi.org/10.1080/02614360122877.
UNWTO (2009): From Davos to Copenhagen and Beyond: Advancing Tourism's Response to Climate Change: UNWTO Background Paper.
Vaníček, J. (2012): Využití marketingového výzkumu pro měření atraktivity turistického cíle. In: 3. mezinárodní kolokvium o cestovním ruchu: Sborník příspěvků. Brno, Masarykova univerzita, 73–80.
VdZ (2016): Besucher 2014 und Eintrittspreise 2015 (internal document).
Vystoupil, J. (2007): Metody pro tvorbu strategických a programových dokumentů cestovního ruchu. Brno, Masarykova univerzita.
Wang, N. (1999): Rethinking authenticity in tourism experience. Annals of Tourism Research, 26(2), 349–370, https://doi.org/10.1016/S0160-7383(98)00103-0.
Woods, B. (1998): Animals on Display: Principles for interpreting captive wildlife. The Journal of Tourism Studies 9(1), 28–39.
Zedková, A. (2012): Analýza vlivu vybraných faktorů na návštěvnost památek a kulturních institucí v České republice. Sborník příspěvků V. Mezinárodní vědecké konference doktorandů a mladých vědeckých pracovníků, Obchodně podnikatelská fakulta Slezské university, Karviná.
Zelenka, J., Pásková M. (2012): Výkladový slovník cestovního ruchu. Linde Praha, Praha.
Annual reports (AR), statistical yearbooks, email correspondations with the representatives of European zoos (Andres 2015; Balejová 2016; Bernheim 2016; Břečková 2016; Dommes 2015; Floriánová 2017; Gereit 2016; Großer 2016; Heck 2016; Knitter 2016; Macháčková 2015; Marx 2015; Mrázková 2016; Nepeřená 2017; Uhrová 2017; Sloet 2016; Šoupalová 2016; Vavřinová 2016), websites of mentioned zoos.

Interregional inequalities at NUTS level 3 in Macedonia

Dejan Iliev*

Ss. Cyril and Methodius University, Faculty of Natural Sciences and Mathematics, Institute of geography, Macedonia
* Corresponding author: d.iliev@hotmail.com; diliev@pmf.ukim.mk

ABSTRACT
This article evaluates interregional inequality among Macedonian regions in the period 2008–2014 by employing statistical measures of regional inequality, including the coefficient of variation, Gini coefficient of concentration and Theil index. Regional inequality is analyzed by using socio-economic indicators, namely the GDP per capita, registered unemployment rate, gross fixed capital formation and entrepreneurial activity. The results show the changing interregional inequality at the level of NUTS 3 regions. The strong polarization between Skopje Region and the rest of the country is the most remarkable feature of regional development in Macedonia between 2008 and 2014.

KEYWORDS
regional inequalities; socio-economic indicators; coefficient of variation; Gini coefficient of concentration; Theil index; Macedonia

Received: 26 December 2017
Accepted: 1 August 2018
Published online: 7 September 2018

Iliev, D. (2018): Interregional inequalities at nuts level 3 in Macedonia.
AUC Geographica 53(2), 180–192
https://doi.org/10.14712/23361980.2018.18

1. Introduction

In the Republic of Macedonia, the problem of regional inequalities is a problem of several decades now and it derives by the political, social and economic factors and changes that have resulted with an inadequate regional distribution of the demographic trends, urbanization, industry and other economic activities. These factors have enabled Macedonia to be a country of two different spatial worlds, one of the dynamic demographic and economic concentration, and the other of an unused space, that is demographically and economically forgotten.

Reducing regional inequalities has been an essential part of the integration and cohesion process of the European Union. The research and analysis of the regions in the Republic of Macedonia and obtaining the concrete results about the problems and disparities in the socio-economic development are of particular importance for the country in the pre-accession period in which the country will have to prepare for the future implementation of structural funds of the European Union and cut the regional differences.

In this study, the subject of its interest is interregional inequalities. The aim of the paper is to measure the interregional inequalities in the Republic of Macedonia after 2008 through an analysis of select socio-economic indicators of Macedonian NUTS 3 regions, and scientifically and empirical to show that the interregional differences exist. Also, the purpose of this article is to find and the causes of differences and to give some recommendations. The whole research was based by theoretical attempts and practices according to the relevant international literature, from which the conceptual framework of the study developed.

In the study, the author has strived to emphasize the gap between the Skopje region and other regions in Macedonia. Hence, the author created the following two hypotheses: H1: There are significant interregional differences in the socio-economic development between the regions, especially between the Skopje region and the other regions; H2: The differences between regions would be decrease in the period 2008–2014. To find whether or not the first hypotheses and assumptions are valid, it was necessary to do several research tasks: (1) to collect statistical data on the value of the GDP per capita, the registered unemployment rate, the gross fixed capital formation and the entrepreneurial activity in 2008, 2011 and 2014; (2) to apply of an adequate statistical method (coefficient of variation, Gini coefficient of concentration and Theil index) to testing series of data and to analyze and quantify the interregional inequalities and their results to be compared; (3) to explore the political context and the government's role in policy-making of reducing the interregional inequalities and implementation at regional development.

The paper is organized into six sections. The next section explores the concept of the theoretical framework of regional inequalities and the theoretical contribution of the study. The third section discusses the research problem and method. The fourth section analyses the development of selected indicators in Macedonian NUTS 3 regions and measuring regional inequality by statistical measures. The fifth section discusses the government's role in regional development. The sixth section summarizes the study with certain conclusions and implications.

2. Theoretical background and contribution of the study

In literature, the best practices for social inequalities (Benyaklef 1997; Michálek and Podolák 2010) and economic inequalities in regional development (Arndt and Sundrum 1975; Banerjee and Jesenko 2015; Barjak 2001; Buckwalter 1995; Clercq and Naert 1985; Ezcurra and Rodríguez-Pose 2013; Galbraith and Garcilazo 2010; Huang, Kuo and Kao 2003; Pérez-Moreno and Angulo-Guerrero 2016) are scientifically elaborated. Trends in regional polarization and econometric model of regional disparities are successfully described in the article on Baláž (2007), and regional inequalities in the articles (Cardoso 1993; Casellas and Galley 1999; Dunford, Key Laboratory of Regional Sustainable Development Modeling (IGSNRR) and Perrons 2012; Ezcurra and Pascual 2007; Kluvánková-Oravská 2004; Lessmann 2014; Ozornoy 1991; Sawers 2006; Smętkowski 2013; Torrisi et al. 2015).

In studying regional differences, inequalities between regions are based on the values of selected quantifiable indicators. Many empirical studies of regional disparities have used a single economic indicator such as the per capita income; however, this approach is not quite objective, because results are dependent on the selected indicator (Quadrado et al. 2001, as cited in Baštová, Hubáčková and Frantál 2011).

There are different approaches to the choice of indicators that are measuring the interregional inequality. For example the EU uses major indicators reflecting economic and social development such as GDP per capita and its sectoral distribution, demographic structure, population density, unemployment rate, etc. (Özaslan, Dincer and Ozgur 2006). The United Nations (UN) evaluates, using the human development index (HDI), mainly the level of development in observed regions, which includes GDP per capita and the level of education and health (average life expectancy) (Baštová et al. 2011: 3).

There are many disparities that can be measured by demographic, social, economic and other indicators, such as: literacy and education, health indicators, gross national product (GNP) per capita, unemployment, foreign trade (Benyaklef 1997). Martin (1997) in showing regional differences in UK, as the main indicator takes into unemployment. In his study, he concludes that the regional unemployment differences are a long-standing feature of the British economic landscape. Barjak (2001) takes into account the following indicators: income indicators, employment opportunities, quality of the environment, training and educational opportunities, the satisfaction of the need to work etc. Berentsen (2006), show the results, based on analysis of changing distributional characteristics of GDP per capita, infant mortality, and unemployment. Gezici and Hewings (2007) in their study to determine regional differences as the main indicator take the following: GDP per capita, as well as Huang, Kuo and Kao (2003). It identified unemployment rates, urbanization rates, shares of the population with a university degree and share of foreign investors in total enterprises as major factors behind regional divergence (Baláž 2007). Goschin, Constantin, Roman, and Ileanu (2008) select just three main economic indicators such as GDP per capita, unemployment rate, and average monthly net earnings. Baštová et al. (2011) in their study used four main socioeconomic indicators (GDP per capita, registered unemployment rate, average monthly gross wage and entrepreneurial activity), to find the differences in the regional development of the Czech Republic.

Therefore, it is necessary to combine different variables to explain socioeconomic regional differences, but it is not always possible to collect selected indicators for certain period's combination with territorial units.

As in many other studies, this paper used a set of socioeconomic indicators and statistical measures as a framework of analysis to computed a interregional inequalities. However, it also pays particular attention to the role of government and politics influence regional development in Macedonia, and in so doing adds another dimension to the current research on regional development in developing countries. Its major theoretical contribution lies in its highlighting of the linkage between political context and government policy, on the one hand, and measuring regional inequality by selected indicators and inequality measures, on the other to understand better how complex are the regional development and the inequalities. The research assessing the interregional inequalities from this perspective has thus far been limited. Most studies focus on the traditional descriptive approach for measuring regional disparities according to various variables, and give little or no attention to the role of government policy.

This article is consistent with many other articles which found the different statistical methods and measures to be fully applicable for assessing the size of regional inequalities, such as the statistical method the coefficient of variation (Kyriacou and Roca-Sagalés 2014; Lessmann 2014; Sacchi and Salotti 2014), then the Gini coefficient of concentration (Mussini 2017; Zubarevich and Safronov 2011, 2014), and the Theil index (Doran and Jordan 2013; Ezcurra and Rodríguez-Pose 2014; Martínez-Galarraga, Rosés and Tirado 2015).

3. Methodology

The main aim of this paper is to find interregional inequalities in the Republic of Macedonia in the period 2008–2014. For this analysis, years 2008, 2011, and 2014 were selected as reference years, which involved significant socioeconomic changes in the Macedonian society, after granting candidate status for membership in the European Union, during the beginning of the global economic crisis and after that.

To prepare the scientific work, the author used secondary data from the State Statistical Office of Macedonia. The primary data is obtained from the researches made by the author. Figure and tabular overviews are used to present regional inequalities. Due to the highly limited availability of data for smaller administrative units[1] (in terms of their existence, structure and timeliness), NTES level 3 which correspond to NUTS level 3 regions were chosen as territorial units for this analysis. The Nomenclature of Territorial Units for Statistics – NTES was adopted by the Government of the Republic of Macedonia in December 2007 (Official Gazette no. 158 of 28.12.2007); the last changes are prepared in January 2014 (Official Gazette no. 10 of 20.01.2014).

The Republic of Macedonia (25,713 km^2) consists of eight NUTS 3 regions:

- *East Region* (14.2% of the territory of the Republic of Macedonia, 8.6% of the total population, a population density of 50.4 inhabitants per km^2).
- *Northeast Region* (9.3% of the territory of the Republic of Macedonia, 8.5% of the total population, a population density of 76.1 inhabitants per km^2).
- *Pelagonia Region* (18.9% of the territory of the Republic of Macedonia, 11.3% of the total population, a population density of 49.3 inhabitants per km^2).
- *Polog Region* (9.7% of the territory of the Republic of Macedonia, 15.4% of the total population, a population density of 131.6 inhabitants per km^2).
- *Skopje Region* (7.3% of the territory of the Republic of Macedonia, 29.7% of the total population, a population density of 337.8 inhabitants per km^2).
- *Southeast Region* (10.9% of the territory of the Republic of Macedonia, 8.4% of the total population, a population density of 63.3 inhabitants per km^2).

1 NTES level 4 there are municipalities (80 units) – correspond to LAU 1 and NTES level 5 there are settlements (1776 units) – correspond to LAU 2.

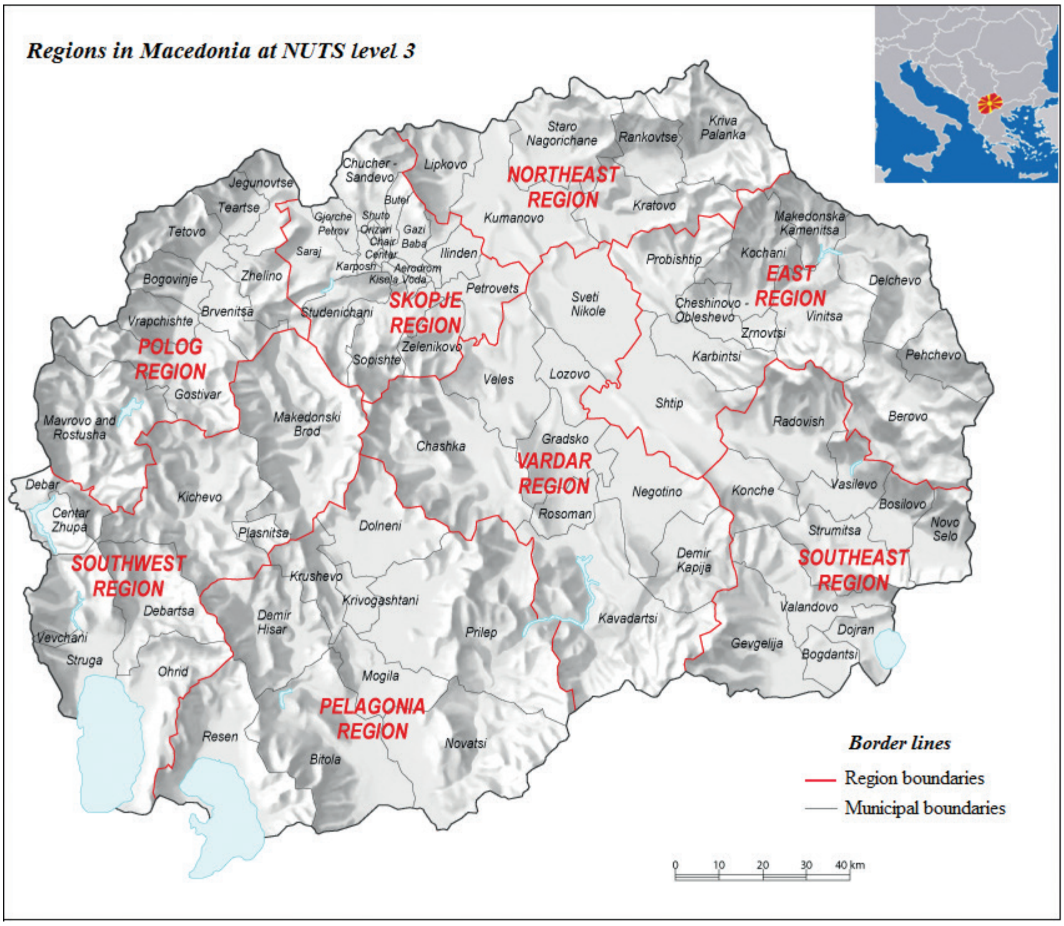

Fig. 1 Regions in Macedonia at NUTS level 3.

- *Southwest Region* (13.4% of the territory of the Republic of Macedonia, 10.7% of the total population, a population density of 66.0 inhabitants per km^2).
- *Vardar Region* (16.2% of the territory of the Republic of Macedonia, 7.4% of the total population, a population density of 38 inhabitants per km^2).

Due to the specific position of the Skopje Region, the regional inequalities will be analyzed twice: firstly in eight administrative regions including Skopje Region, and secondly in seven regions excluding Skopje Region. By applying this analysis, the distortions that are caused by extreme values of selected indicators of Skopje Region will be eliminated. In this way, a more objective evaluation of regional differences in the other seven regions will be allowed.

A key part of the first phase of the empirical analysis is the selection of proper indicators. It is very difficult to choose the set of indicators that will identify different areas and developed regions. Thus, it is necessary to combine various indicators to express overall socioeconomic regional inequality. Based on the literature review and data availability four indicators were selected: GDP per capita, registered unemployment rate, gross fixed capital formation and entrepreneurial activity. One of the main reasons for the selection of these four socio-economic indicators in the analysis of regional inequalities was their availability and compatibility for the selected years of the monitored period, which was not the case with other indicators which can be used in analyzes.

GDP per capita is a widely used indicator mainly characterizing the economic performance of the region. Gross domestic product (GDP) at market prices is the final result of the production activity of the resident producer units and it is the sum of gross value added of the various institutional sectors or the various activities at basic prices plus value added tax and import duties less subsidies on products (which are not allocated by activities) (SSORM 2017: 358). Data for the calculation of GDP are determined by the Eurostat methodology of ESA 2010. But one should have in mind that the data on GDP per capita in certain regions is largely under the influence of the migration of the labor force, and it contributes in a given region to create GDP which is not a result of the resident population. That is one of the shortcomings of this indicator. According to own insights, the data on GDP per capita can be overestimated in the region where it is produced (e.g, the Skopje Region), and on the other hand, it can be underestimated in regions of the large outflow of workforce (for example, many people from the other regions are working in the Skopje Region and there they have registered their revenues).

The unemployment rate is a standard socio-economic indicator presenting the quality of life of the population and economic situation in the region. The unemployment rate as the participation of the number of unemployed in the total labour force is used to find interregional inequalities. The unemployment rate is calculated using the following equation:

$$UNr = \frac{NUp}{TWf} \cdot 100\%$$

where, UNr is the unemployment rate, NUp is the number of unemployed persons, and TWf is the total workforce.

The third indicator 'gross fixed capital formation' is often used for comparing the socio-economic level of regions. The data on Gross Fixed Capital Formation a result of the annual survey on Gross Fixed Capital Formation (INV.01) of private, cooperative, mixed and state enterprises and organizations, the commodity flow method and the calculation of investments for the entities which do not have the status of a legal person. The method used in the calculation of gross fixed capital formation is in line with the recommendations of SNA 2008 and ESA 2010 (SSORM).

The fourth indicator represents the rate of entrepreneurial activity. The calculation of entrepreneurial activity is based on data of the number of active businesses per 1000 inhabitants, following equation:

$$Ea = \frac{NAb}{TNp} \cdot 1000\%$$

where, Ea is entrepreneurial activity, NAb is the number of active business entities, and TNp is the total number of population.

Entrepreneurial activity was calculated from records of the number of active business entities listed in the Statistical Business Register. Active business entities are all business entities that contribute to the gross domestic product, and the basic criterion for determining the activity of the entity is the data on income and/or employees (SSORM 2017: 529).

For assessment of the size of regional inequalities in the selected years, it was also necessary to choose different methods. The author selects three statistical measures of regional inequalities such as the coefficient of variation, Gini coefficient of concentration, and Theil index. With the application of the three measures will be determined the variability measured by chosen indicators. These measures can be monitored through a time series and in this way it can be seen whether the inequalities are increasing or decreasing. Thus, only these three indices will be used in what follows in the text.

The most widely used measure of differences is the coefficient of variation (*Vx*). This measure based on mean logarithmic deviation, which is related to its average and is set in a percentage. The coefficient of variation measuring regional inequality has the form:

$$Vx = \frac{\sigma}{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \frac{1}{\bar{x}}$$

where *n* is the number of observational units, x_i is the value of indicator *x* for the *i-th* region and $\bar{x}$ is the arithmetic mean of values of indicator *x*. The higher the value is, the greater the inequality of the distribution research phenomena.

According to Netrdová and Nosek (2009), this measure has a positive and negative side. The positive side of this measure is its non-dimensionality, through which it is possible to compare the variables in different measurement units. From the other side, the dependence of this measure on the average of the distribution is a negative feature, because this distribution is inappropriate within the scope of reality of very asymmetrical distribution of socioeconomic phenomena. So, the analysis of regional inequalities within a set of regions (e.g. 7 regions in Macedonia, excluding Skopje Region), which have a more symmetrical distribution of selected indicators, will refine the results.

Perhaps, the most used measure of determining inequality, especially in the measurement of income inequality, is the Gini coefficient of concentration (*Gi*). The Gini coefficient, proposed by Corrado Gini in 1914, has been the focus of many theoretical and empirical studies. The Gini coefficient of concentration can be respectively written as:

$$Gi = \frac{1}{2n^2\bar{x}} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|$$

where *n* is the number of observational units, x_i is the value of the indicator *x* for the *i*-th region, x_j is the value of the indicator *x* for *j*-th region and $\bar{x}$ is the arithmetic mean of the indicator x values. The values range from $0 \leq Gi \leq 1$, where 0 represents absolute equality and 1 represents an absolute inequality of the distribution of the phenomenon.

The third method 'Theil index' is defined as a weighted geometric mean of a relative phenomenon and it is mainly used to measure economic inequality. In this article, a non-weighted form of the index is used because the selected indicators have already been related to the population of the region. The Theil index (*Th*) is calculated with the equation:

$$Th = \frac{1}{n} \sum_{j=1}^{n} \frac{y_j}{\bar{y}} \ln\left(\frac{y_j}{\bar{y}}\right)$$

where *n* is the number of regions, $\bar{y}$ expresses the arithmetic mean of the phenomenon, y_j is the value of the phenomenon in the region *j*. The values range from 0 which corresponds to absolute equality of distribution of the phenomenon, to 1 which corresponds to absolute inequality of distribution of the phenomenon.

Many academics scholars used the decomposition of inequality measures. Thus, the most commonly used is the Theil index decomposition into (the within-group) and between (group components), where the total inequality is equal to the sum of these groups. Without the knowledge of data for the lower territorial levels (e.g. municipalities), it is not possible to undertake the decomposition. Because, the selected indicators in this paper were available only for NUTS 3 regions, the Theil index decomposition is not calculated.

4. Measuring regional inequality by selected indicators and inequality measures

In this part of the paper, a comparison of selected indicators' development in 2008, 2011, and 2014 as well as a comparison of inequality measures for selected variables are presented.

4.1 GDP per capita

GDP per capita increased on average by 32.7% in all regions during the analyzed period. The growth rate of by GDP per capita was higher in the period 2008–2011 compared with the period 2011–2014 in most regions. The most progressive increase of GDP per capita (80.8%) was recorded in the Southeast Region as the fastest growing region, and the smallest growth rate of by GDP per capita was recorded in Skopje Region (16.5%). Above national average GDP per capita was recorded in the Skopje, Pelagonia, East, Southeast, and Vardar Region over the tracked period (2008–2014). On the other hand, regions with the lowest GDP per capita not varied. In 2008, the Northeast Region and Polog Region achieved the lowest level of GDP per capita. Also, in 2014 the lowest GDP per capita was recorded in the Northeast Region and Polog Region which underwent minimal increases of the indicator over the eight year period (Table 1). In Polog Region the gray economy has bigger dimensions compared to other regions in the country, so the data that shows a lower GDP per capita because of the unregistered economy and it should be taken with caution. Also, the foreign currencies that are sent from abroad mostly are remittances for people who live in the Polog Region, which only indicates that the population of this region had a higher standard than other less developed regions (e.g. the Northeast Region is traditionally poorer).

There are several reasons that contribute to the development of GDP per capita in particular regions. Thus, one of the main causes of regional inequalities is the sector structure of regional economies. For example, the economic performance of Skopje Region is driven by commercial services, trade, financial and insurance activities, manufacturing, construction, information and communication etc.; the Southeast Region is significantly more involved in agriculture and service activities compared with the others region. Also, in the Northeast Region, an unfavourable sector structure of manufacturing might be the cause of the low value of GDP per capita.

Tab. 1 Regional differences in GDP per capita and registered unemployment rate in Macedonian regions.

Region	GDP per capita (MKD)			Registered unemployment rate (%)		
	2008	2011	2014	2008	2011	2014
East	173.815	224.455	244.272	20.0	16.4	20.1
Northeast	122.014	146.047	148.745	58.0	59.6	44.0
Pelagonia	208.990	224.485	251.988	34.5	31.4	18.7
Polog	95.277	114.113	117.284	26.4	31.8	30.7
Skopje	314.531	319.717	366.482	37.3	30.7	29.0
Southeast	168.211	251.471	304.140	11.7	9.3	20.8
Southwest	150.771	174.509	189.109	39.3	42.8	36.4
Vardar	196.028	220.590	274.404	43.6	36.4	27.6
Dataset including Skopje Region						
Vx	0.3470	0.2866	0.3246	0.3989	0.4457	0.2877
Gi	0.1846	0.1582	0.1836	0.2232	0.2426	0.1593
Th	0.0571	0.0419	0.0547	0.0845	0.1055	0.0403
Dataset excluding Skopje Region						
Vx	0.2333	0.2384	0.2905	0.4307	0.4727	0.3083
Gi	0.1312	0.1304	0.1634	0.2435	0.2619	0.1706
Th	0.0286	0.0302	0.0447	0.0972	0.1195	0.0462

Source: Computed by author based on data from Stat Statistical Office of the Republic of Macedonia.

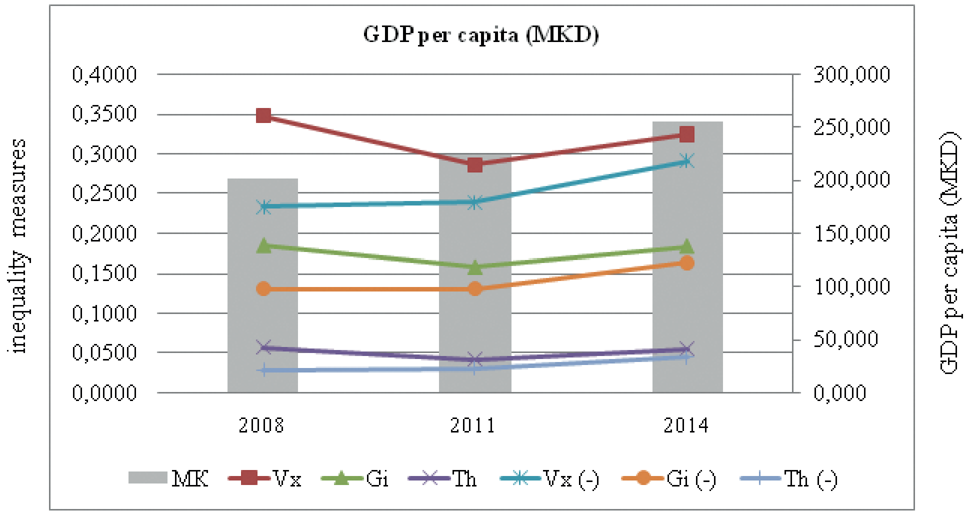

Fig. 2 GDP per capita in Macedonian regions (including Skopje Region and excluding Skopje Region) in 2008, 2011, and 2014 (coefficient of variation, Gini coefficient of concentration, Theil index).

Source: Author chart based on data from Stat Statistical Office of the Republic of Macedonia. Notes: MK = national average; Vx and Vx (–) = coefficient of variation calculated from the dataset including and excluding Skopje Region; Gi and Gi (–) = Gini coefficient of concentration calculated from the dataset including and excluding Skopje Region; Th and Th (–) = Theil index calculated from the dataset including and excluding Skopje Region.

The indicator of GDP per capita is influenced by the high value of GDP per capita in Skopje Region in the long-term. The increasing economic performance of Skopje Region can be explained by a comparison of the inequality measures computed with Skopje Region and without it. The inequality measures of GDP per capita computed with Skopje Region recorded 1.5 times (2008), 1.2 times (2011) and 1.1 times (2014) higher values of the coefficient of variation; 1.4 times (2008), 1.2 times (2011), and 1.1 times (2014) higher values of the Gini coefficient of concentration; 2.0 times (2008), 1.4 times (2011) and 1.2 times (2014) higher values in the case of the Theil index in comparison with values of these inequality measures based on dataset excluding Skopje Region (Table 1).

The comparison of values of inequality measures for GDP per capita in selected years shows that there are no statistically significant trends in the development of these differences. In the period 2008-2014, the values of inequality measures based on a dataset including Skopje Region, are reduced. On the other hand, a very slight growth was recorded for the values of inequality measures based on dataset excluding Skopje Region, but these differences can be considered as flat (Figure 2). These small deviations are likely to have been caused by the increase of economic performance in East Region, Southeast Region and Vardar Region compared to other regions (increase by 40.5%, 80.8% and 40.0% between 2008 and 2014). In contrast, the Skopje Region in this period registered a smaller increase of 16.5% of GDP per capita.

4.2 Registered unemployment rate

The unemployment rate of the population at a regional level shows a difference compared to the overall rate level in Macedonia. With the highest the unemployment rate above the national average is the Northeast, Pelagonia, Skopje, Southwest and Vardar Region in 2008 and Northeast, Polog, Skopje and Southwest Region in 2014 (Table 1). As long as the unemployment rate in a given region is above average, a tendency to higher emigration and lower immigration is expected (Michálek and Podolák 2010: 41).

During the reference period, in terms of average registered unemployment rate, there was a change in the values of interregional inequalities, but this change was being not statistically significant. Although computed inequality measures (coefficient of variation, Gini coefficient of concentration, Theil index) for registered unemployment rate show a slight decrease in the monitored period, the results are statistically insignificant (Figure 3).

The above-mentioned slight decline in the size of interregional differences was consequence mainly by the decline of registered unemployment rates in Northeast, Pelagonia, Skopje, Southwest and Vardar Region. This can be seen especially in 2014 when the Vardar Region recorded the largest decrease of unemployment in the Republic of Macedonia. One of the reasons for this decrease could have been the new work positions, which opened with the construction of the Dräxlmaier factory in Kavadarci Municipality. The highest unemployment rate of 44% in the Northeast Region is a consequence of the emigration of many people outside the country, who are registered as residents in Macedonia (mostly unemployed for using free health insurance), as well as the considerable number of engaged workers in agriculture and the informal sector who are registered as unemployed.

The problems with the increased level of unemployment are more than obvious in less developed regions sparking the labour migration flows towards

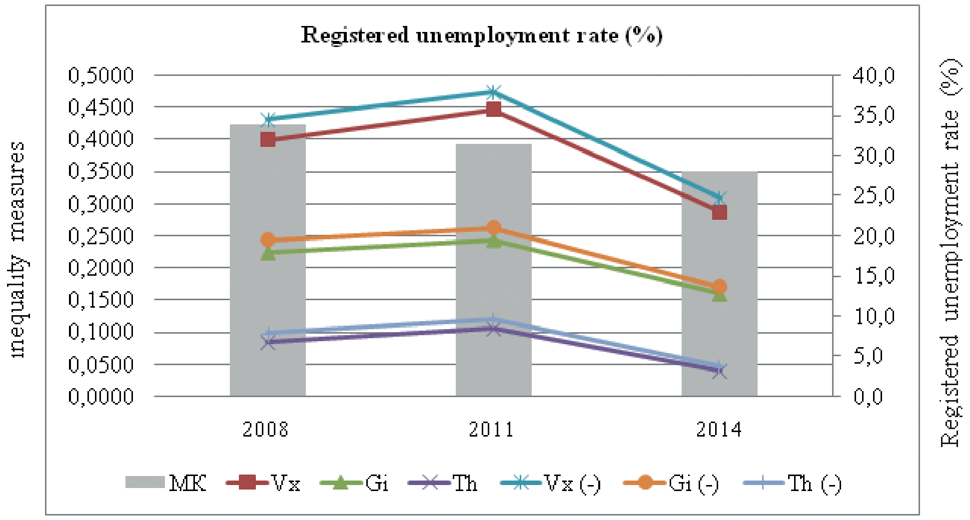

Fig. 3 Registered unemployment rate in Macedonian regions (including Skopje Region and excluding Skopje Region) in 2008, 2011, and 2014 (coefficient of variation, Gini coefficient of concentration, Theil index).

Source: Author chart based on data from Stat Statistical Office of the Republic of Macedonia.
Notes: MK = national average; Vx and Vx (–) = coefficient of variation calculated from the dataset including and excluding Skopje Region; Gi and Gi (–) = Gini coefficient of concentration calculated from the dataset including and excluding Skopje Region; Th and Th (–) = Theil index calculated from the dataset including and excluding Skopje Region.

more developed ones and, consequently, reducing the regional perspectives for more dynamic economic growth. At the same time, one may note a higher concentration of poor population in less developed regions with certain implications in terms of social exclusion (Rocheska, Angeleski, Milevska, Kostoska, 2014: 71–72). Also, some unemployed move out of the region, mostly the young and educated labour force. Emigration of the educated young labour force causes a structural lack of qualified labour. Thus, the region can attract only enterprises with low innovation levels.

Generally, the indicator of registered unemployment rate recorded minimal differences among inequality measures computed in two separate ways (including and excluding data of the Skopje Region). This means that the very high values of the registered unemployment rate in Skopje Region have no impact upon overall regional differentiation. Unlike, with measuring regional inequality by selected inequality measures based on the data excluding Skopje Region, it can be noted that regional inequalities are larger but insignificant (Table 1).

4.3 Gross fixed capital formation

Referring to the values of the inequality measures used, the highest interregional differences are in gross fixed capital formation. This fact is probably the result of the opposite values of gross fixed capital formation: very high values in the Skopje Region, and very low values in the other 7 regions (Table 2). Average by gross fixed capital formation constantly increased between the monitored years in the Republic of Macedonia (national average increased by 26.4% between 2008 and 2011, and by 13.1% between 2011 and 2014).

The highest value of gross fixed capital formation during this period is present in the Skopje Region, which has 4.6 times higher value than the national average. The other regions have a value below the national average. In the monitored period, the highest growth rate of gross fixed capital formation was recorded in the Southwest Region (145.5%). Also, the high growth rate of this indicator was registered in Southeast Region (87.2%), East Region (83.1%), Vardar Region (65.3%), Pelagonia Region (40.0%) and Skopje Region (32.5%). The smallest growth rate was recorded in the Polog Region (6.3%) and Northeast Region (17.6%).

One of the main reasons of regional differences is the sector structure of the gross fixed capital formation. Observed by sectors of activity by regions, in all sectors, except Agriculture, the Skopje Region had the highest share. For example, the sector Construction had the highest share (35.4%) in total gross fixed capital formation or 51.1% share on national level. In the monitored period, the sector Construction is the driving force of economic growth in the Skopje Region. This was a consequence of a stable real estate market, as well as projects for infrastructure investments, especially in the urban project ‘Skopje 2014’, which had a significant contribution to increasing the capacities in the construction sector.

Table 2 illustrates the regional differentiation of gross fixed capital formation in the Republic of Macedonia. This differentiation is generally great, as shown by the high values of inequality measures. The size of interregional differences in gross fixed capital formation was nearly the same in all monitored years, but with a minimum reduction in the last analyzed year as confirmed by individual inequality measures

Tab. 2 Regional differences in gross fixed capital formation and entrepreneurial activity in Macedonian regions.

Region	Gross fixed capital formation (MKD)			Entrepreneurial activity (%)		
	2008	2011	2014	2008	2011	2014
East	5.558	5.070	10.175	29.40	32.58	31.94
Northeast	2.163	1.353	2.543	21.18	24.41	23.63
Pelagonia	5.950	5.765	8.329	32.03	35.56	35.06
Polog	6.637	7.708	7.058	19.35	22.47	22.96
Skopje	54.082	74.086	71.681	39.93	46.37	43.59
Southeast	3.708	5.355	6.940	31.93	36.10	34.40
Southwest	3.859	4.241	9.473	29.19	33.34	32.33
Vardar	4.445	5.642	7.350	31.37	38.08	35.47
Dataset including Skopje Region						
Vx	1.5197	1.6775	1.3833	0.2078	0.2111	0.1926
Gi	0.5577	0.6053	0.5143	0.1112	0.1155	0.1036
Th	0.7094	0.8499	0.6062	0.0222	0.0227	0.0189
Dataset excluding Skopje Region						
Vx	0.3084	0.3559	0.3094	0.1760	0.1750	0.1596
Gi	0.1733	0.1846	0.1610	0.0907	0.0940	0.0832
Th	0.0506	0.0761	0.0569	0.0164	0.0161	0.0134

Source: Computed by author based on data from Stat Statistical Office of the Republic of Macedonia.

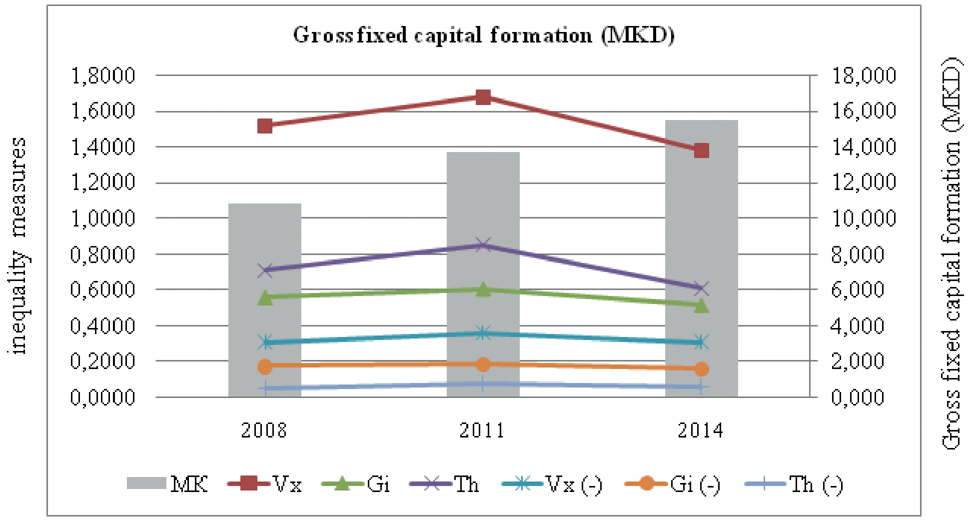

Fig. 4 Gross fixed capital formation in Macedonian regions (including Skopje Region and excluding Skopje Region) in 2008, 2011, and 2014 (coefficient of variation, Gini coefficient of concentration, Theil index).

Source: Author chart based on data from Stat Statistical Office of the Republic of Macedonia

Notes: MK = national average; Vx and Vx (–) = coefficient of variation calculated from the dataset including and excluding Skopje Region; Gi and Gi (–) = Gini coefficient of concentration calculated from the dataset including and excluding Skopje Region; Th and Th (–) = Theil index calculated from the dataset including and excluding Skopje Region.

(Figure 4). On the other hand, there is a considerable difference between the inequality measures of gross fixed capital formation computed with and without the Skopje Region. The high value of gross fixed capital formation in Skopje Region distorts the results. Thus the inequality measures of Macedonian regions without Skopje Region are actually by 4.9 times (2008), 4.7 times (2011) and 4.5 times (2014) lower values of the coefficient of variation; 3.2 times (2008), 3.3 times (2011) and 3.2 times (2014) lower values of the Gini coefficient of concentration; 14.0 times (2008), 11.2 times (2011) and 10.7 times (2014) lower values of the theil index in comparison with values of these inequality measures based on dataset including Skopje Region (Table 2).

4.4 Entrepreneurial activity

The indicator of entrepreneurial activity, demonstrated an increase in the monitored period (10.7%). The highest increase of entrepreneurial activity

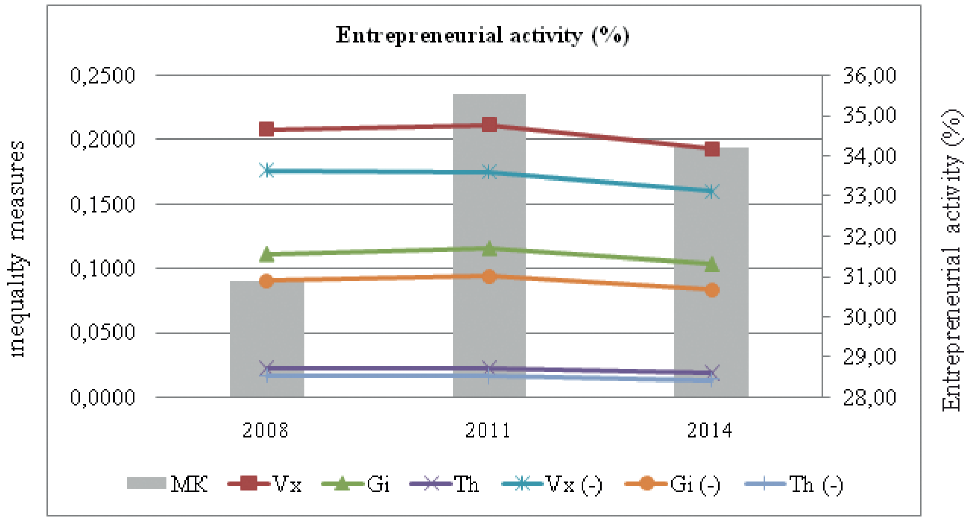

Fig. 5 Entrepreneurial activity in Macedonian regions (including Skopje Region and excluding Skopje Region) in 2008, 2011, and 2014 (coefficient of variation, Gini coefficient of concentration, Theil index).

Source: Author chart based on data from Stat Statistical Office of the Republic of Macedonia.
Notes: MK = national average; Vx and Vx (–) = coefficient of variation calculated from the dataset including and excluding Skopje Region; Gi and Gi (–) = Gini coefficient of concentration calculated from the dataset including and excluding Skopje Region; Th and Th (–) = Theil index calculated from the dataset including and excluding Skopje Region.

was reported in the Polog Region (18.7%), followed by Vardar Region (13.1%). Conversely, the lowest increase of entrepreneurial activity was experienced in the Southeast Region (7.7%), followed by East Region (8.6%).

In all monitored years, entrepreneurial activity reached the highest values in Skopje Region. This region is traditionally most attractive in terms of businesses. For example, in 2014, the data on the structure of enterprise births by regions show that the highest share of 37.1% belongs to the Skopje Region, while the Northeast Region had the lowest share of 5.6%. Also, the data on the structure of active business entities by regions show that the highest share of 38.0% belongs to the Skopje Region, while the Northeast Region had the lowest share of 5.9%. Simultaneously, 1/3 of the population of the country lives in Skopje. On the high values of entrepreneurial activity in Skopje may be influenced by the fact that a considerable part of companies operating in other regions in the country have their headquarters registered in Skopje.

The Polog Region reached the highest growth rate of entrepreneurial activity in the monitored period. This fact is quite surprising with regard to the lowest values of GDP per capita in this region. High values of entrepreneurial activity can be caused by developed trade enterprises, catering business entities, construction business and real estate activities. For example, in 2014, the data on the structure of enterprise births by regions show that the second highest share of 12.9% belongs to the Polog Region. Conversely, the low value of entrepreneurial activity is found in the Northeast Region, which corresponds to the low GDP in the region.

The impact of the development of entrepreneurial activity on the level of regional differences is presented by the values of inequality measures (Table 2). It can be observed that these measures achieved relatively the lowest level of interregional differences within the monitored indicators. However, the size of interregional inequalities of entrepreneurial activity is statistically insignificant especially with regard to the low values of the Gini coefficient of concentration and the Theil index. Generally, there are no important trends indicated in the development of these differences, except in the insignificant reduced values between 2011 and 2014 (Figure 5).

The comparison of values of individual inequality measures computed by including and excluding data for the Skopje Region, shows the impact of entrepreneurial activity in the Skopje Region on the overall size of regional differentiation. It should be noted that in the monitored period (2008–2014), the values of inequality measures based on the dataset including and excluding Skopje Region, are reduced. However, these inequalities can be considered as flat (Figure 5).

5. The government's role in regional development

As a result of the problems arising from unequal regional development in Macedonia, it started legally and institutionally to regulated by the Law on Balanced Regional Development (Official Gazette no. 63/2007) and the Strategy for Regional Development of the Republic of Macedonia 2009–2019. The Law stipulates the establishment of the planning regions which corresponds to the regionite of the NUTS 3 nomenclature and the realization of the measures and instruments for the improvement of the development. In accordance with the Law, the planning regions are the basic territorial units for development in the Strategy for Regional Development. The Law on Regional Development stipulates the objectives of the regional development policy in the Republic of Macedonia, such as: balanced and sustainable development based on the model of polycentric development; reducing disparities between regions; increasing competitiveness and developing the special identity of the regions; revitalization of villages and development of areas with specific development needs; support to inter-municipal and cross-border cooperation. Based on these purposes, in the Strategy are defines the strategic purposes and priorities of the policy for stimulating the balanced regional development of the Republic of Macedonia for the period 2009–2019. Also, an attempt has been made to closely align the Strategic objectives of the Strategy for Regional Development with the priorities of the EU's economic and social cohesion policy. The Strategy is a national strategic document that is primarily intended for the Government and the Ministry of Local Self-Government, which is responsible for the management and implementation of the regional development policy. As a separate entity within the Ministry of Local Self-Government, the Bureau for Regional Development has the role of the main operational unit in the policy implementation.

The realization of the purposes provided in the Strategy for Regional Development of the Republic of Macedonia is related to significant financial resources. In accordance with the Law on Balanced Regional Development, as sources of financing of regional development are: the Budget of the Republic of Macedonia; the budgets of the units of local self-government; available EU funds; other international sources; donations and sponsorships from individuals and legal entities and other means determined by Law.

One of the Government's role is to divide funds for financing, which is an important precondition for implementing the model of polycentric regional development. Thus, almost all regions are envisaged to receive twice as many funds compared to the Skopje Region for which 6.4% of the total available funds will be allocated. These funds are provided by the Government from the Budget of the Republic of Macedonia (Assembly of the Republic of Macedonia, 2009). Thus, from the aspect of regional development, it is very important to ensure greater acceleration of economic growth in the less developed planning regions, which is directly conditioned by the dispersion of investments by regions.

The Law stipulates for the stimulation of balanced regional development from the Budget of the Republic of Macedonia annually allocating funds for at least 1% of GDP. These funds further need to be allocated to 70% financing of projects about the development of the planned regions, 20% for financing projects for the development of areas with specific developmental needs and 10% for financing projects for the development of villages. For example, in 2009, in the Budget of the Republic of Macedonia, 4,568 million MKD was planned to support regional development, which represents 1.15% of the GDP. Later, in 2010, 6,308 million MKD were provided for regional development, which represents 1.53% of the GDP.

Finally, the monitoring of the implementation of the Strategy for Regional Development is under the authority of the Government of the Republic of Macedonia through an annual report on the implementation of the Action Plan of the Strategy.

6. Conclusion, recommendations, limitations and future research

The paper studies the size and the development of interregional inequalities in the Republic of Macedonia by selected socioeconomic indicators in 2008, 2011 and 2014. Because of the highly limited availability of data, NUTS 3 regions were chosen as territorial units for this analysis. The interregional inequalities were computed from the dataset including the data of the Skopje Region and the dataset excluding Skopje Region. That allowed to show the expected discrepancy caused by the high values of selected indicators of Skopje Region. Three different statistical measures of variability were used for the assessment of interregional inequalities (the coefficient of variation, Gini coefficient of concentration, and Theil index).

The results show that the interregional inequalities in monitored indicators are small, especially in the set of regions excluding Skopje Region. In this case, the highest values of inequality measures were recognized in the indicator of registered unemployment rate computed by the coefficient of variation (0.4457 in 2011), Gini coefficient of concentration (0.2426 in 2011) and Theil index (0.1055 in 2011). The interregional inequalities based on the dataset including Skopje Region achieved values several times higher, however, the values were still relatively low in some inequality measures. In this case, the highest values of the coefficient of variation were recognized in the indicator of the gross fixed capital formation (1.6775 in 2011), the Gini coefficient of concentration (0.6053 in 2011) and Theil index (0.8499 in 2011). It should be noted that the values of the Theil index were lower than 0.1060 for all indicators in all selected years, except for the indicator of the gross fixed capital formation based on the set of regions including Skopje Region. These results indicate show that significant cause of inequalities at the level of NUTS 3 regions are the high values of the Skopje Region in all monitored indicators as expected.

The comparison of inequality measures between the individual indicators reveals some inequalities where the relatively highest values were found in the gross fixed capital formation and GDP per capita. The other indicators (registered unemployment rate and entrepreneurial activity) shows smaller interregional inequalities, which are however statistically insignificant. Furthermore, it was revealed that the Theil index reports the lower level of inequality for all selected indicators in comparison with the other measures, especially in dataset excluding Skopje Region. On the other hand, it was found that the coefficient of variation reports the higher level of inequality for all selected indicators in comparison with the other measures.

Generally, the size of interregional inequalities in the monitored socioeconomic indicators did not change significantly in the selected years, ie there is no clear trend in the development of interregional inequalities. Thus, the results did not confirm the expected impacts in the Republic of Macedonia after granting candidate status for membership in the European Union, during the beginning of the global economic crisis and period after that, in terms of greater reduction of regional differences.

Supporting the current regional development requires government intervention in a number of ways, particularly by facilitating systematic change and by fostering the less developed regions. As priority measures to cut inequalities at the regional level, I emphasize: demographic revitalization of the less developed regions, raising the level of social development, enhancing the economic growth and more uniform dispersion of investments and employees between regions, the development of the modern infrastructure, an improvement on the technical and technological basis for supporting the industries and utilizing the creative potential, valorization of the level of human assets, the creation of competitive advantages of the regions, an responsible use and valorization of the natural resources and the energetic potentials, the development and support of underdeveloped areas and areas with specific developmental needs, develop the cross-border cooperation of regions and so on.

Due to the limited size of this paper, research was limited to four socioeconomic indicators that can also serve as productive starting points for future research. Thus, the future research could be enhanced by the addition of more socioeconomic indicators to better assess and further identify regional inequalities. Also, future research could employ multiple inequality measures and theories relevant to find the interregional inequalities. It would be more effective to analyze regional inequalities at the level of lower administrative units, where higher regional inequalities can be expected. However, in these conditions, this is limited by the non-availability of data for lower administrative units. Thus, the future research could focus on increasing the number of administrative units (NUTS level 4), where higher regional inequalities can be expected. Overall, the research paper generated useful findings and points to valuable directions for further research.

References

Arndt, W. H., Sundrum, M. R. (1975): Regional Price Disparities. Bulletin of Indonesian Economic Studies 11(2), 30–68, https://doi.org/10.1080/000749175123 31332702.

Assembly of the Republic of Macedonia (2009): Strategy for Regional Development of the Republic of Macedonia 2009–2019. Official Gazette of the Republic of Macedonia, no. 119.

Baláž, V. (2007): Regional Polarization under Transition: The Case of Slovakia. European Planning Studies 15(5), 588–602, https://doi.org/10.1080 =09654310600852639.

Banerjee, B., Jesenko, M. (2015): Economic Growth and Regional Disparities in Slovenia. Regional Studies 49(10), 1722–1745, https://doi.org/10.1080/00343404 .2013.879981.

Barjak, F. (2001): Regional Disparities in Transition Economies: A Typology for East Germany and Poland. Post-Communist Economies 13(3), 289–311, https:// doi.org/10.1080/14631370120074849.

Baštová, M., Hubáčková,V., Frantál, B. (2011): Interregional differences in the Czech Republic, 2000–2008. Moravian Geographical Reports 19(1), 2–16.

Benyaklef, M. (1997): Socio-economic disparities in the Mediterranean. Mediterranean Politics 2(1), 93–112, https://doi.org/10.1080/13629399708414606.

Berentsen, H. W. (2006): Changing Regional Inequalities in United Germany. Eurasian Geography and Economics 47(4), 462–477, https://doi.org/10.2747/1538 -7216.47.4.462.

Buckwalter, W. D. (1995): Spatial Inequality, Foreign Investment, and Economic Transition in Bulgaria. The Professional Geographer 47(3), 288–298, https://doi .org/10.1111/j.0033-0124.1995.288_e.x.

Cardoso, R. A. (1993): Regional inequalities in Europe have they really been decreasing? Applied Economics 25(8), 1093–1100, https://doi.org/10.1080 /00036849300000089.

Casellas, A., Galley, C. C. (1999): Regional Definitions in the European Union: A Question of Disparities? Regional Studies 33(6), 551–558, https://doi.org/10.1080 /00343409950078242.

Clercq, de M., Naert, F. (1985): Regional economic disparities in Belgium and regional, national and supra-national policies: an analysis of collective failure. Planning Outlook 28(1), 13–20, https://doi.org/10.1080 /00320718508711734.

Doran, J., Jordan, D. (2013): Decomposing European NUTS2 regional inequality from 1980 to 2009: National and European policy implications. Journal of Economic Studies, 40, 22–38, https://doi.org/10.1108 /01443581311283484.

Dunford, M., Key Laboratory of Regional Sustainable Development Modeling (IGSNRR), Perrons, D. (2012): Regional Inequality in the EU: How to Finance Greater Cohesion. European Planning Studies 20(6), 895–922, https://doi.org/10.1080/09654313.2012.673562.

Ezcurra, R., Pascual, P. (2007): Spatial Disparities in Productivity in Central and Eastern Europe. Eastern European Economics 45(3), 5–32, https://doi.org/10.2753/EEE0012-8775450301.

Ezcurra, R., Rodríguez-Pose, A. (2013): Political Decentralization, Economic Growth and Regional Disparities in the OECD. Regional Studies 47(3), 388–401, https://doi.org/10.1080/00343404.2012.731046.

Ezcurra, R., Rodríguez-Pose, A. (2014). Trade openness and spatial inequality in emerging countries. Spatial Economic Analysis 9, 162–182, https://doi.org/10.1080/17421772.2014.891155.

Galbraith, K. J., Garcilazo, E. J. (2010): Inequalities, employment and income convergence in Europe: evidence from regional data. International Review of Applied Economics 24(3), 359–377, https://doi.org/10.1080/02692171003701594.

Gezici, F., Hewings, J. D. G. (2007): Spatial Analysis of Regional Inequalities in Turkey. European Planning Studies 15(3), 383–403, https://doi.org/10.1080/09654310601017091.

Goschin, Z., Constantin, D. L., Roman, M., Ileanu, B. (2008): The Current State and Dynamics of Regional Disparities in Romania. The Journal of the Romanian Regional Science Association 2(2), 80–105.

Huang, Jr.-T., Kuo, C.-C., Kao, An.-P. (2003): The Inequality of Regional Economic Development in China between 1991 and 2001. Journal of Chinese Economic and Business Studies 1(3), 273–285, https://doi.org/10.1080/476828032000108553.

Kluvánková-Oravská, T. (2004): Structural funds the challenge for sustainable development and regional disparity in the Slovak Republic. Innovation: The European Journal of Social Science Research 17(1), 61–73, https://doi.org/10.1080/1351161042000190745.

Kyriacou, A. P., Roca-Sagalés, O. (2014): Regional disparities and government quality: Redistributive conflict crowds out good government. Spatial Economic Analysis 9, 183–201, https://doi.org/10.1080/17421772.2014.891158.

Lessmann, C. (2014). Spatial inequality and development-Is there an inverted-U relationship? Journal of Development Economics 106, 35–51, https://doi.org/10.1016/j.jdeveco.2013.08.011.

Martin, R. (1997): Regional Unemployment Disparities and their Dynamics. Regional Studies 31(3), 237–252, https://doi.org/10.1080/00343409750134665.

Martínez-Galarraga, J., Rosés, J. R., Tirado, D. A. (2015): The long-term patterns of regional income inequality in Spain, 1860–2000. Regional Studies 49, 502–517, https://doi.org/10.1080/00343404.2013.783692.

Michálek, A., Podolák, P. (2010): Socio-economic disparities and migration in Slovakia. Moravian Geographical Reports 18(2), 36–45.

Mussini, M. (2017). Decomposing changes in inequality and welfare between EU regions: The roles of population change, re-ranking and income growth. Social Indicators Research 130, 455–478, https://doi.org/10.1007/s11205-015-1184-2.

Netrdová, P., Nosek, V. (2009): Přístupy k měření významu geografického rozměru společenských nerovnoměrností. Geografie-Sborník České geografické společnosti 114(1), 52–65.

Özaslan, M., Dincer, B., Ozgur, H. (2006): Regional Disparities and Territorial Indicators in Turkey: Socio-Economic Development Index (SEDI). ERSA Conference Paper ersa06p858, European Regional Science Association.

Ozornoy, I. G. (1991): Some Issues of Regional Inequality in the USSR under Gorbachev. Regional Studies 25(5), 381–393, https://doi.org/10.1080/00343409112331346587.

Pérez-Moreno, S., Angulo-Guerrero, J. M. (2016): Does economic freedom increase income inequality? Evidence from the EU countries. Journal of Economic Policy Reform 1–21, https://doi.org/10.1080/17487870.2015.1128832.

Rocheska, S., Angeleski, M., Milevska, M., Kostoska, O. (2014): Territorial cohesion and regional development in the context of EU integration-the case of Macedonia. Eastern Journal of European Studies 5(1), 67–77.

Sacchi, A., Salotti, S. (2014): How regional inequality affects fiscal decentralisation: Accounting for the autonomy of subcentral governments. Environment and Planning C: Government and Policy 32, 144–162, https://doi.org/10.1068/c1241r.

Sawers, L. (2006): Inequality and the Transition: Regional Development in Lithuania. Baltic Journal of Economics 6(1), 37–51, https://doi.org/10.1080/1406099X.2006.10840428.

Smętkowski, M. (2013): Regional Disparities in Central and Eastern European Countries: Trends, Drivers and Prospects. Europe-Asia Studies 65(8), 1529–1554, https://doi.org/10.1080/09668136.2013.833038.

State Statistical Office of the Republic of Macedonia [SSORM], (2017): Statistical yearbook of the Republic of Macedonia, p. 358, p. 529. Skopje: SSORM.

Torrisi, G., Pike, A., Tomaney, J., Tselios, V. (2015): (Re-)exploring the link between decentralization and regional disparities in Italy. Regional Studies, Regional Science 2(1), 123–140, https://doi.org/10.1080/21681376.2015.1007159.

Zubarevich, N. V., Safronov, S. G. (2011): Regional inequality in large post-Soviet countries. Regional Research of Russia 1, 15–26, https://doi.org/10.1134/S2079970511010138.

Zubarevich, N. V., Safronov, S. G. (2014): Inequality in the Socioeconomic Development of the Regions and Cities of Russia in the 2000s, Sociological Research 53(6), 48–68, https://doi.org/10.2753/SOR1061-0154530603.

The 2015–2016 famine threat in Ethiopia: a study of the relevance of famine archetypes

Dagmar Milerová Prášková

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

ABSTRACT

In recent years, Ethiopia has experienced fast economic growth and has been a major recipient of development and humanitarian aid. However, these developments were unsuccessful in eliminating food insecurity problems, and Ethiopia continues to be a considerably famine-prone country. The aim of this paper is to examine the applicability of Howe's framework of the six archetypal situations symptomatic to famines (watch, price spirals, aid magnet, media frenzy, overshoot, and peaks) to the 2015–2016 food crisis, which left an estimated 15 million Ethiopians in need of acute food assistance. This paper observes that the food crisis proved to have some of Howe's archetypes including watch, price spirals, and, to a lesser extent, media frenzy and peaks. Even though the aid magnet and overshoot were not recorded, the dynamics of the 2015–2016 food crisis confirmed Howe's argument that the current system of humanitarian assistance does not lead to timely and effective responses. In this paper, I also argue that the Ethiopian political context further exacerbates the food insecurity situation of the country.

KEYWORDS

Ethiopia; food crisis; famine; political response

Received: 23 April 2018
Accepted: 4 October 2018
Published online: 16 November 2018

Milerová Prášková, D. (2018): The 2015–2016 famine threat in Ethiopia: a study of the relevance of famine archetypes.
AUC Geographica 53(2), 193–206
https://doi.org/10.14712/23361980.2018.19

1. Introduction

In 2015, East and South Africa were hit by one of the worst droughts in more than 50 years, caused by the El Niño phenomenon. In Ethiopia, two rain seasons brought below-average rainfalls and over 15 million people were in need of acute food assistance in 2016 (FEWSNET 2015a). Some Ethiopian regions were classified as 'emergency' by the Famine Early Warning Systems Network (FEWSNET), which is the phase preceding a full-scale famine according to the FEWSNET classification (FEWSNET 2015b). It implies that the country was facing high risk of mortality and in acute need of humanitarian assistance. The World Health Organization predicted that the situation would worsen until August 2016 and the country would need at least another year to recover from the crisis (WHO 2015). This was already the fifth extensive food crisis since 1984, of which two were declared famines. In December 2016, the number of Ethiopians suffering from undernourishment had reached 28.8% in the 3-year average from 2014 to 2016 (FAO 2017). Paradoxically, Ethiopia experienced strong economic growth with 10.8% between 2003 and 2014 and a slight drop to 8% in 2015 to 2016 (WB 2017). At the same time, Ethiopia is one of the major recipients of development aid (OECD 2016).

This paper examines the applicability of Howe's (2010) six archetypal situations recurring in famines to a 'famine threat'. These archetypes are known as watch, price spirals, aid magnet, media frenzy, overshoot, and peaks. Scrutinizing the presence or absence of factors that underlie a transition from food crisis to famine should help to understand whether they would be relevant in a situation of a severe food emergency not translated into famine. Thus, the contribution of this paper is to appraise the relevancy of these archetypes in non-famine situations which might help humanitarian actors to better assess and react to upcoming food insecurity emergencies. The paper collates various secondary data and information to discuss whether and how these situations took place during the 2015–2016 crisis. The article has the following structure: Section 2 discusses conceptual issues. Section 3 addresses data and methods. Section 4 briefly outlines the history of Ethiopian famines and their patterns. Section 5 provides the main analysis of the 2015–2016 famine threat then examines the crisis origins, food price evolution, movement of population, world media attention, and the humanitarian response. Section 6 discusses the findings and summarizes the main conclusions.

2. Conceptualizing famines

Famine, as an ultimate demonstration of food insecurity, is characterised as high mortality linked to starvation. According to the UN Food and Agricultural Organization (FAO) definition, a famine is declared when at least 20% of households within an affected area face extreme food shortages; acute malnutrition reaches 30% and mortality exceeds two persons per day per 10,000 people (FAO 2008). However, a common definition has not been approved by the international community due to the temporal, scale and sectoral ambiguities and uniqueness that every famine comprises (Howe 2002). For example, there was no expert agreement on the extent of the crises in Sudan in 1998, Ethiopia in 1999 to 2000 and 2002 to 2003 or Malawi in 2002 where famines were not officially declared by the UN even though the countries were hit by severe food emergencies. Whether a famine is officially recognized or not it brings political and emotional connotations. A delayed declaration of famine may lead to a delayed response which may have fatal consequences. A famine definition and conceptualization is thus needed for both practical and political reasons (Howe, Devereux 2004).

Early theories considered famine a consequence of food production failures due to natural disasters. This view was challenged by Amartya Sen who highlighted the critical importance of access to food related to various forms of socioeconomic inequalities. He explained famines by referring to the absence of 'food entitlements' in terms of means and resources to secure food rather than food unavailability as such (Sen 1981a). In subsequent literature, two major famine conceptualizations can be distinguished with one referring to the food availability decline and another referring to the food entitlement decline (Devereux 2009). These two aspects also represent two main pillars of the food security definitions adopted at the World Food Summit (1996). Congruently, a major shift in the focus of famine debates has been the consideration of response failure theories. It accentuates the role of governments and political representation to prevent large-scale food emergencies in the modern era. As Devereux (2007) puts it, new famines are always preventable, so their nature is necessarily political. The persistence of famine in a globalized world signifies political failures of states and the international relief system (Devereux, Howe, Biong Deng 2002).

Various factors can lead to a famine outbreak. These include natural disasters, economic shocks, international and internal conflicts or other political failures. In practice, however, a single cause is often emphasized in the official depiction of famines such as, typically, drought or other natural disasters. A narrow focus on natural causes or immediate triggers and insufficient attention paid to other drivers hinders development actors from defining complex policies and responses (Sandstrom, Juhola 2017). In order to complement early warning systems and reach effective and appropriate responses, Howe and Devereux (2004) suggested a new approach to defining famine based on intensity and magnitude scales. 'Intensity' refers to the severity of the crisis, and 'magnitude' refers to the scale of the crisis. Howe and Devereux (2004) designed a graduated concept of famine to avoid 'a binary conception of famine / no famine' and allow a 'minor famine' to be declared even without high mortality. The authors also stress that stakeholders should concentrate more on prevention policies to reduce large-scale food emergencies.

Equally, famine represents deep social disruption (Rivers et al. 1976). Humanitarian and non-governmental organizations tend to attract international attention in early stages of a food crisis, but a needed reaction usually comes only after shocking images appear in popular media (Hammond, Maxwell 2002). This demonstrates the tremendous power international media can have in international relations. The way a state responds to famine depends mainly on two factors: the ruling regime's political priorities and their relationship with potential donors (Cutler 1991). Donor countries also consider the political situation of the affected country and their geopolitical interests. There is a high probability that aid may not get into the hands of those who really need it if there is ongoing civil conflict or an undemocratic rule. Conversely, the likelihood of famine or starvation on a massive scale is relatively low in a well-functioning state thanks to different protection mechanisms such as early warning systems, social protection programs, regional food reserves or crop insurance.

As Edkins (2002) points out, there are tendencies to comprehend famine as a failure which can be solved solely with technological solutions. Whether deliberate or intentional, Edkins writes that the responsible actors should be brought to justice and political actions should come first to prevent famine. It is no coincidence that the severest food crises occur in war-torn countries and authoritarian or dictatorship regimes. Marcus (2003) also calls for famine criminalization within international law. He defines four types of government behaviour. First, actions of incompetent and corrupt politicians do not respond to a population's acute needs. Second, government activities can be labelled as indifference to the current emergency. Third, policies are dominated by recklessness despite direct impacts on starving people. Forth, hunger is used deliberately as a political weapon to exterminate regime enemies. This last argument underlines intentional behaviour and can be applied to the case of Somalia in 2011 when famine was declared in several regions controlled by the radical group Al-Shabaab which forbade most international aid agencies to provide humanitarian aid (Menkhaus 2012).

In practice, there are several safeguarding mechanisms to alert food insecurity crises such as the Global Information and Early Warning System (GIEWS) managed by FAO and FEWSNET developed by the US Agency for International Development (USAID). GIEWS provides an information analysis of crop prospects and monitoring, markets and trade, price and policy, supply and demand balance, vulnerability and risk. FEWSNET focuses on weather and climate, markets and trade, agricultural production, livelihoods, nutrition, and food assistance. It identifies the following five food insecurity phases as distinguished in the Integrated Food Security Phase Classification (IPC) of Global Partners (FAO 2012): 1) minimal, 2) stressed, 3) crisis, 4) emergency and 5) famine. FEWSNET predicted the food crisis in East Africa in 2011 six months in advance and has thus been labelled an 'excellent and timely tool' and better than GIEWS which has a lower predictive ability (Ververs 2012).

Arguing against the linear view of the famine process, Howe (2010) analyses larger patterns within famines through the lenses of systems thinking analysis. He seeks analogies between the basic elements of the systems approach, such as reinforcing feedback loops, balancing loops, delays, and leverage, to understand specific situations or archetypes of famines. The 'watch' situation is when it is difficult to assess whether a food crisis will develop into an absolute food emergency. Such a situation is usually linked to poor food production and inadequate coping strategies of the affected people. The watch situation is followed by 'prices spirals' which is when food prices increase due to lack of food in markets. The third situation is called 'aid magnet' which refers to the fact that aid centres attract large inflows of hungry people resulting in their capacity being exceeded. Overcrowding then leads to poor sanitation conditions and the needs of the hungry population may not be met due to limited food resources. Subsequent higher mortality provokes the fourth situation 'media frenzy' where media outlets compete for the best stories and images. However, media attention comes too late to prevent the emergency. The fifth situation 'overshoot' relates to ramping up of assistance after the most severe period of the crisis has passed. The sixth situation 'peaks' combines all five situations and refers to successive spikes characterising famines and responses. However, Howe concedes that '*many situations will not conform neatly to the archetypes*'.

Howe argues that the current setting of humanitarian aid does not provide timely and effective responses because aid does not come until after all peaks take place. Howe puts emphasis on system thinking. Even though he does not precisely quantify the specific archetypes (such as the increase of food prices, number of fleeing people to relief sites or number of media input to be considered as media frenzy), he leaves the space for specific conditions, specific situations, and specific considerations. However, this approach has not been applied to analyse a concrete case (Rubin 2018). Later, Howe has expanded this approach by introducing five basic elements characterising formation and evolution of a famine: pressure (on food security), hold (worsening situation), self-reinforcing dynamics (pattern on accelerated changes), famine system, and rebalancing (Howe 2018). Both Howe's works create a coherent framework and a convenient tool for analysis of famines in practice and for further research. On the other hand, it is not clear if these archetypes would be relevant to a situation of famine threat or a severe food emergency.

3. Data and methods

In this article, I collated information from various sources to examine the applicability of Howe's archetypes to the case of the 2015–2016 food crisis in Ethiopia. To assess the presence of the watch situation, reports and alerts of early warning systems and organizations working on spot can be examined. For the purposes of this article, FEWSNET monthly reports were used to create a time sequence of the issued emergencies and a framework of available information about worsening food insecurity in specific regions. For price spirals, I looked at three different indicators: 1) food inflation rates between 2014 and 2016 as provided by the World Food Programme (WFP); 2) staple food prices in the two most insecure regions in terms of food from January 2014 to December 2016 (data provided by FAO GIEWS) and 3) price spikes provided by the WFP Alert for Price Spikes (ALPS). As an indicative assessment of the aid magnet condition, I focused on the number of internally displaced Ethiopians in the period between 2014 to 2016 from the International Organization for Migration (IOM) and the Internal Displacement Monitoring Centre (IDMC). Distinguishing of the cause of home abandonment provides another insight into the peoples' motivations and another possible aspect of this archetype. The media frenzy was analysed through the outputs of the most world known English and French-speaking media and non-governmental organizations working in Ethiopia. The overshoot situation was examined through the humanitarian funding situation with a focus on humanitarian needs and funding gaps. For this purpose, data from the UN Office for the Coordination of Humanitarian Affairs (OCHA) was used. The assessment of the last archetype combined all of the abovementioned information.

4. Previous famines and food crises in Ethiopia

For a long time, Ethiopia has been a symbol of famine due to the affliction of recurrent famines since the 9th century. Their brief characteristic is described in Table 1. About one third of the population most likely died during the 1888–1892 famine (Sen 1981a). Since that, Ethiopia has experienced several severe food crises labelled as famines and there were similar patterns occurring during those famines.

Table 1 suggests that poor rainfalls and subsequent droughts triggered the famines and food crises in Ethiopia. Lack of rain increases hunger incidences and the probability of crisis. Farmers' food stocks usually run out during lean seasons (periods of regular food insecurity and rises in food prices) before new harvests. Lean season usually lasts from April to June in Belg-receiving areas (listed famine-prone

Tab. 1 Character of previous famines and severe food crises.

Year	Affected regions	Death toll	Trigger(s)	Reinforcing factors	Responses
1972–1974	Afar Amhara Somali Tigray	40–80 thousand people	erratic rain drought	delayed warning report by FAO	delayed request for aid by the Ethiopian government, introduction of FEWSNET
1984–1985	Afar Amhara Tigray	2 million 0,6 million forced to resettle	drought, desertification, land degradation	high food prices, military actions of the Marxist rule	delayed international response, only after Live Aid phenomenon
1999–2000	Amhara Oromia SNNPR Somali Tigray	71–122 thousand	drought	restricted access of humanitarian organizations to the Somali region	failed early response of the Ethiopian government
2002–2003 (food crisis)	Amhara SNNPR Tigray	loss of life avoided	drought, insufficient recovery from past droughts, environment degradation	poor market infrastructure, inability to distribute food across local markets, indebtedness of farmers	quick aid by the USA thanks to Ethiopia's engagement in the war on terrorism after 9/11 attacks
2011 (food crisis)	Amhara Afar Oromia Somali	12.4 million of affected people, refugee deaths in hundreds	drought, depleted water resources	high cereal prices, scarcity of pasture and water	late response by the government, UN and donors but sophisticated humanitarian system

Source: Miller, Holt (1975) for the 1972 famine; Hammond (2011) and De Waal (1991) for the 1984 famine; Devereux (2009) and Khalif, Doornbos (2002) for the 1999 famine; Devereux (2007) and Brown (2008) for the 2002 food crisis; Ververs (2012) and FEWSNET (2011) for the 2011 crisis.

regions) and from February to April in pastoralist areas. In the lowlands, pasture and land availability decline thus making livestock products (such as milk) generally unavailable. Moreover, the Ethiopian government and often also international organizations did not respond adequately to recurrent food crises despite their apparent predictability. The Table 1 also indicates that similar Ethiopian regions in the North-east, East and South of the country were hit the most. These are arid and semi-arid ecological zones with the prevalence of low rainfall, high temperatures and ongoing desertification (EPA 1998).

The 2015–2016 crisis recorded similar pattern like previous famines and food crises: trigger (subsequent waves of drought), geographical distribution and extent, reinforcing factors (high food prices) and late responses. Interestingly, no famine has materialised since the implementation of the Productive Safety Net Programme in 2005 implying the mitigation effects of the programme. On the other hand, Ethiopia still faces periodic food emergencies and decades of food aid have not had any substantial impact on hunger and poverty reduction (Sabates-Wheeler, Devereux 2010).

5. The 2015–2016 food emergency

Borrowing the structure of famine dynamics elaborated by Howe (2010), this section analyses how the food crisis cycle evolved during 2015 and 2016 due to the threat of famine outbreak.

5.1 Watch

During this period, early warning systems usually provide information about production and food deficit, but it is not yet clear whether the afflicted population will be able to cope with the upcoming situation. Coping strategies include, for instance, a reduction in food consumption, sale of livestock and assets, seeking of off-farm jobs or even migration. However, this period plays a key role in famine prevention. Yet, there are two dangerous moments. First, possible underestimation of the severity level of the crisis may result in non-response of donors. Second, coping strategies may bring some short-term relief but usually prevent people from reacting to the future crisis while selling land, livestock or tools (Howe 2002).

In order to analyse the beginning and evolution of the food crisis in 2015 to 2016, monthly Famine Early Warning System reports are scrutinised as FEWSNET is considered to be a precise tool for averting emergencies (Ververs 2012). The article takes into account the whole period of 2015 which includes the uncertainty of crisis creation and time for coping strategies. In reality, the watch period lasted for a shorter time. Table 2 shows that the FEWSNET expected food security to decline from stressed (IPC Phase 2) to crisis (IPC Phase 3) from January 2015 in the six affected regions: Afar, Amhara, Oromia, Tigray SNNPR and Somali.

Tab. 2 FEWSNET reports from January to June 2015.

Region	Jan 2015	Feb 2015	Mar 2015	April 2015	May 2015	June 2015
Amhara (N + E)	IPC2 → IPC3 till June			IPC3 from July to September	IPC3 from July to September	IPC3 from July to September
Tigray	IPC2 → IPC3 till June PC2 → IPC3			IPC3 from July to September		
Oromia (E + central)	IPC2 → IPC3 till June	IPC3 in June	IPC3 through June	IPC3 from July to September		
Afar (N + NE)		IPC3 in May and June	IPC3 through June			
SNNPR				IPC3 from July to September	IPC3 from July to September	IPC3 from July to September
Somali						

Source: FEWSNET Food Security Outlook reports (2015)
Notes: Integrated Food Security Phase Classification (IPC): 1) minimal, 2) stressed, 3) crisis, 4) emergency and 5) famine (FAO 2012)

A FEWSNET alert came in June 2015 calling for food emergency assistance. It said that '*20 to 50 percent less rain than average between March and May, (...) planted area was less than 60 percent of average across SNNPR*' (FEWSNET 2015c). In addition, the Ethiopian government declared the Belg rain failed in June (HRD 2016). Acute malnutrition and emergency food assistance needs were also predicted in the following months. In October 2015, several regions were expected to be at risk of or to move to an emergency (IPC Phase 4). Another alert was issued in December 2015 calling for '*immediate sustained, large-scale, multi-sectoral emergency assistance to save lives and livelihoods*' (FEWSNET 2015a). Table 3 shows the prediction of crisis severity change.

The FAO GIEWS system provided information about deteriorating food security conditions in June 2015 observing that major problems would be felt in central-northern SNNPR and '*the estimated number of people in need of relief food assistance is expected to increase significantly*' (FAO GIEWS 2015). The

Tab. 3 FEWSNET reports from June to December 2015.

Region	June	July 2015	Aug 2015	Sept 2015	Oct 2015	Dec 2015
Amhara				All stressed regions likely to remain in IPC3 through Dec 2015 at least		IPC4
Tigray						IPC4
Oromia					IPC4 since Jan–Mar 2016	IPC4
Afar (South)			IPC3 only with humanitarian assistance through Dec 2015 at least		IPC4	ALERT: emergency assistance required immediately
SNNPR	ALERT: emergency food assistance needed till July/ Sept					
Somali (North)		IPC3 only with humanitarian assistance through Dec 2015 at least	IPC3 only with humanitarian assistance through Dec 2015 at least		IPC4	ALERT: emergency assistance required immediately

Source: FEWSNET Food Security Outlook reports (2015); See Notes for Table 2

Figure 1 shows the IPC phases in the food security outlook from October to December 2015 provided by FEWSNET.

It is also important to highlight that the food security situation was deteriorating in the western regions as well, especially due to violence and conflicts in neighbouring countries. By the end of 2015, Ethiopia hosted more than 730 thousand refugees and asylum-seekers from South Sudan, Sudan, Somalia and Eritrea. Most of the refugee camps were located in the Gambella and Benishangul-Gumuz regions. Deficiency in financial means had a strong negative impact on humanitarian conditions in these camps resulting in tensions between local Ethiopians and the incoming people (FAO GIEWS 2015).

5.2 Price spirals

For the price spirals archetypes, it is important to consider the fact that poor households in less developed countries spend as much as 60–80 percent of their incomes on food (WFP 2012), so their vulnerability to rising food price is very high. According to the World Food Programme, food prices showed an upward trend compared to the long-term average in 2010 to 2014 (WFP 2015). Prices of different staples differed in particular regional markets, though. To assess this trend, I focused on food inflation rates from 2014 to 2016, prices of staple food in two regions covered by GIEWS: Mekele (Tigray region) and Dire Dawa (east of Ethiopia, on the border between Oromia and Somali), and data from the Alert for Price Spikes.

Usually, the highest food prices are recorded during regular food insecurity periods in the lean season (February to June). Food inflation was growing throughout the year in 2015 with two-digit values (Figure 2) making food more expensive and thus less available for poor households. The food inflation peak in December 2015 corresponds with a second failed harvest that year. There was a considerable decrease in inflation in 2016.

In relation to food prices, the GIEWS tool focuses on two local markets in two considerably food-insecure regions: Mekele and Dire Dawa (Figure 2). According to the FEWSNET (2015d), the most consumed cereal by the poor in rural areas of Ethiopia is maize, the cheapest is white sorghum, and teff (a local crop) is a very important cereal throughout the country. A significant part of the local diet is also made up of white wheat. Thus, Figure 2 focuses on these crops in order to ascertain price instability which generally further exacerbates food insecurity.

In Mekele, maize recorded the most stable price while the price of sorghum almost doubled in the 18 months following January 2015. The data also disproves the FEWSNET claim that sorghum is the cheapest staple food in Ethiopia. Teff was the most expensive crop, experiencing a steady rise in price since 2015. In Dire Dawa, the price of maize was a little higher than in Mekele but it recorded a stable level. The price of sorghum was about 30% higher than in Mekele in 2015, reaching the same level by April 2016. The price of wheat recorded the highest level in both markets. Information about the price of teff is not available for Dire Dawa.

For the detection of abnormally high food prices, the Alert for Price Spikes focuses on the gap between observed prices and long-term seasonal trends. It characterizes a situation as 'normal', 'stress', 'alert' or 'crisis' (WFP 2017). The following five markets in food-insecure regions (defined by the FEWSNET

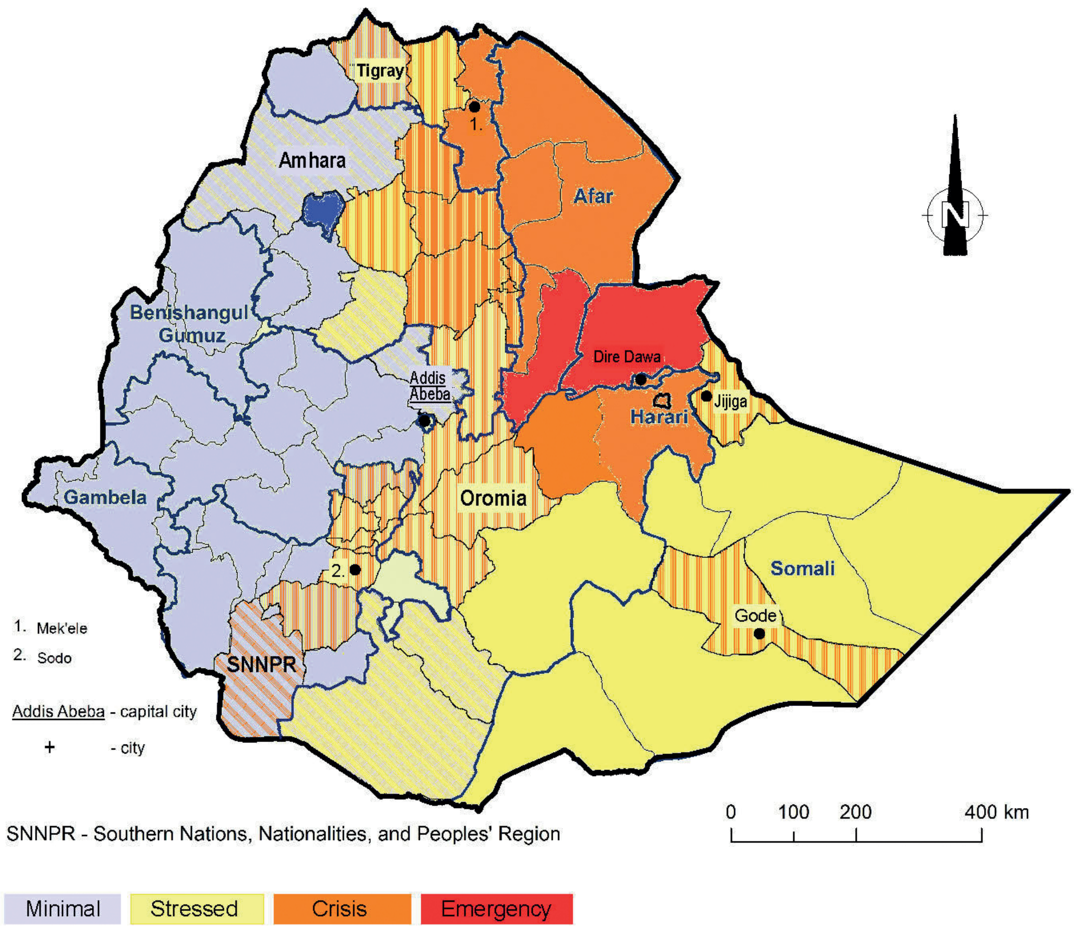

Fig. 1 IPC Phases in Ethiopia.

Source: FEWSNET (2015b)

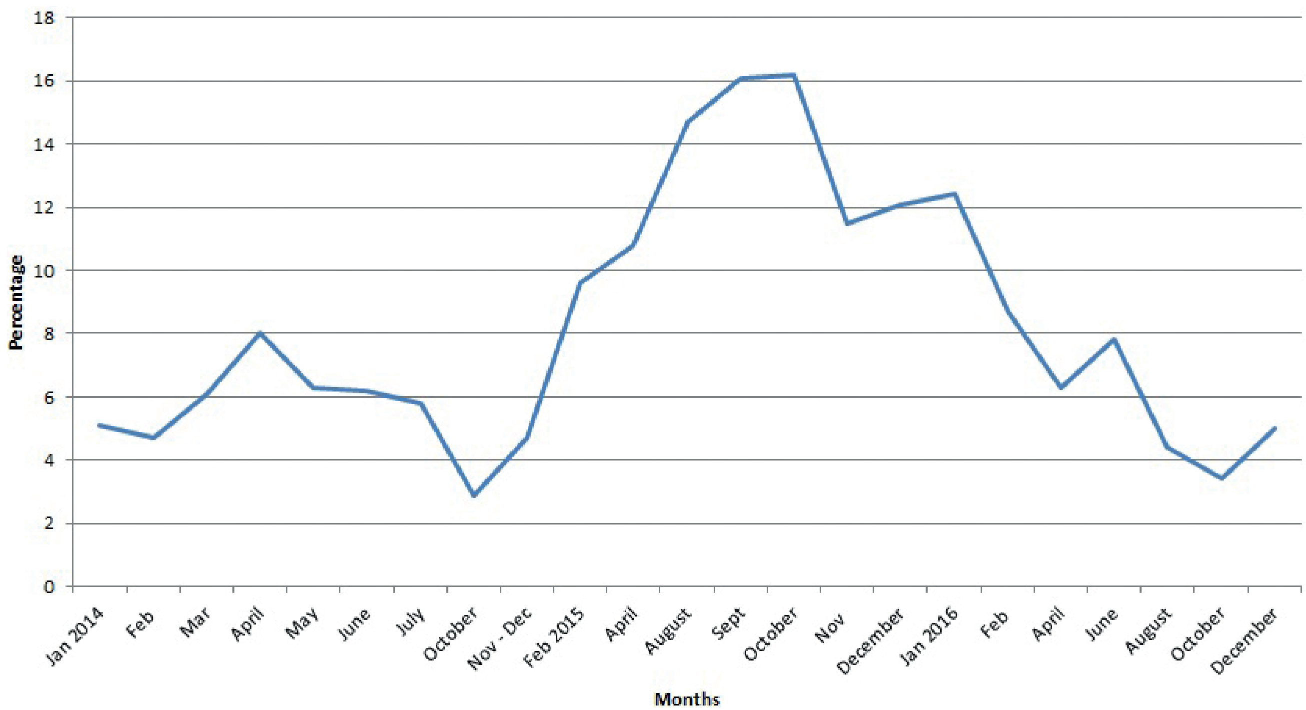

Fig. 2 Food inflation in Ethiopia.

Source: WFP Ethiopia Monthly Market Watch documents

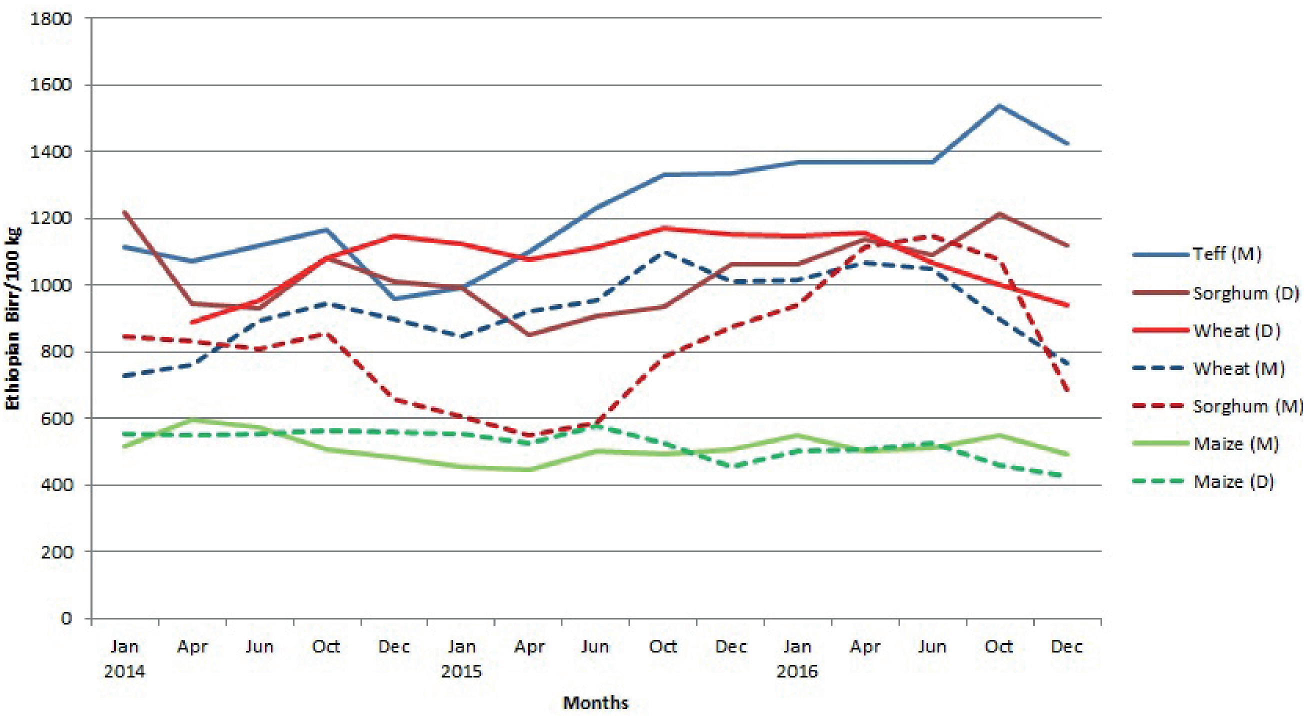

Fig. 3 Wholesale prices of staple foods in the Mekele (M) and Dire Dawa (D) markets (Ethiopian Birr/100 kg).
Source: FAO GIEWS FPMA Tool (2017)

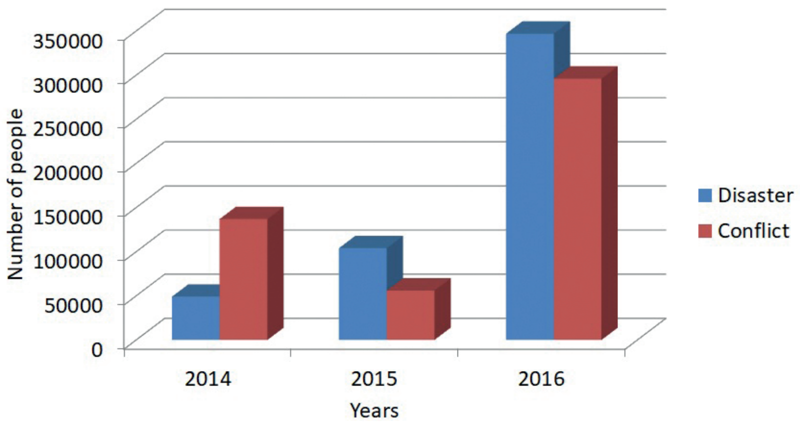

Fig. 4 Internal displacement and causes in Ethiopia.
Source: IDMC (2018)

classification) were considered: Dire Dawa, Mekele, Jijiga (Somali), Sodo (SNNPR) and Gode (Somali). During the examined food crisis, there were two alerts and two crises. The first alert was in Sodo (July 2015, wheat price) and the second was in Jijiga (June 2016, wheat price). The crises were in Gode (January 2015, maize price) and Mekele (October 2015, wheat price), both contributing significantly to the worsening of the already dire situation in poor Ethiopian households.

5.3 Aid magnet

Howe (2010) identifies an 'aid magnet' as a situation where large numbers of hungry people head to relief sites overwhelming the capacity of agencies to meet their needs with the increase in undernutrition and excess mortality. Due to its famine history, humanitarian activities in Ethiopia are built on '*existing robust systems and programmes, including the Productive Safety Net Programme, the health extension programme, the Community-Based Management of Malnutrition programme and the network of mobile health and nutrition teams*' (UNICEF 2017). During 2015–2016, there was an increasing number of health facilities treating severe acute malnutrition because the estimated number of children affected by severe acute malnutrition for 2015 increased from 300 to 350 thousand (UNICEF 2015), with predictions of 420 thousand in 2016 (UNICEF 2016a). UNICEF (2016b) also states that increased country

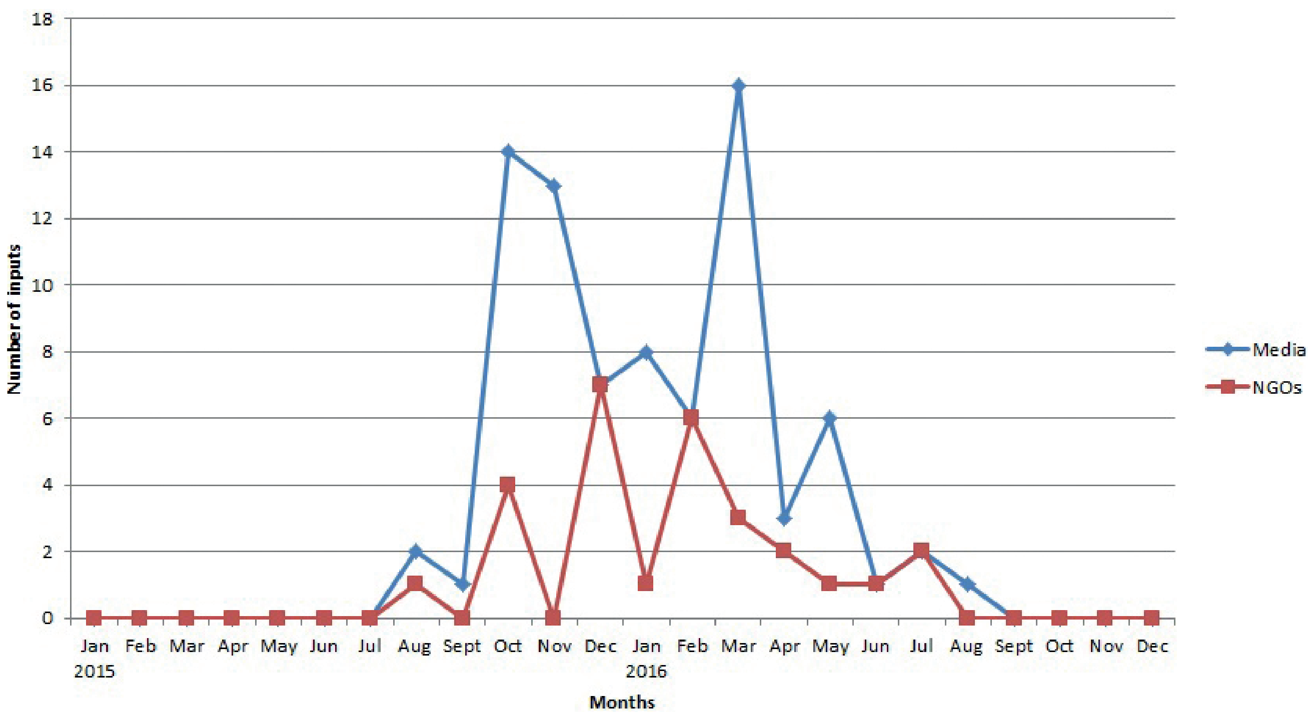

Fig. 5 Media and NGO coverage of drought in Ethiopia.
Source: own research

capacity and delivery of services prevents excess mortality in humanitarian situations.

There were two causes of the new internal displacement in Ethiopia. In 2015, 104 thousand people left their home due to disasters (drought or sudden flooding) and 56 thousand were on the move due to a conflict. Sixty seven percent of the internally displaced were from the Somali region (predominantly pastoralist), 14% were from Oromia, 11% from Afar and the rest from Gambella, Amhara, SNNPR and Hareri (IOM 2016). In 2016, Ethiopia experienced a sharp rise in internal displacement. The number of people leaving their homes due to disasters rose more than three times to 347 thousand and more than five times to 296 thousand due to a conflict (Figure 4).

5.4 Media frenzy

The media frenzy archetype is characterised as competition amongst the media and substantial public interest following exposure to a horrific image from famine hotspots (Howe 2010). For the purposes of this article, the following English and French speaking best-known world media were analysed: The Guardian, BBC, New York Times, Washington Post, Reuters, CNN, Al Jazeera, Le Monde and Radio France International (RFI). The choice was based on my language skills and on testing the presence of the topic within the concrete journal, television or radio. I searched for the words Ethiopia, food crisis, famine, drought, 2015 and 2016 in news and opinion articles and only articles focusing on drought and upcoming food crisis published online were considered. A timeline was formed, and the frequency of coverage was created. Also, I looked for the international non-governmental organizations working in Ethiopia to discover their possible influence on media coverage of the food crisis after their warnings and calls for emergency aid. These were OXFAM, Save the Children, Care International, and Catholic Relief Services. World media news about the drought and its consequences for food security in Ethiopia started to appear in August 2015 (The Guardian, RFI). Clear links to the 1984 famine were made and a famine threat was pictured as real even though famine was not predicted by any early warning system. The number of news and NGO media outputs is indicated in Figure 5.

There was no media coverage on the food crisis in Ethiopia in the first half of 2015. The first media response corresponds with the FEWSNET monthly report in October when an emergency (Phase 4) was predicted and the Ethiopian government asked for international assistance from donors. In November 2015, the media focused on weather conditions and Ethiopia's capacity to respond. In December, FEWSNET issued the alert for emergency assistance and non-governmental organizations published most of their stories, however, media attention was already decreasing. The Ethiopian government asked for foreign aid in January 2016. In March 2016, the media reacted to a United Nations report on the humanitarian situation in Ethiopia and in May 2016, world media publicized the FEWSNET severe food crisis prediction. After August 2016, the Ethiopian food crisis disappeared from international news and did not reappear until March 2017.

5.5 Overshoot

Donor countries may be hesitant to provide aid in the beginning of a crisis which may lead to an overshoot of emergency responses when too much aid arrives at the end of a crisis. It also takes several months for emergency aid to come as the process involves preparation and examination of proposals, fund releasing and aid transportation (Howe 2010). In Ethiopia, the UN Office for the Coordination of Humanitarian Affairs (OCHA 2015) uses the Hotspot woreda (administrative unit) classification to prioritize response, especially in supplementary feeding. The number of areas needing urgent nutrition support quadrupled between February and December 2015. Woredas with Priority 1 increased from 40 in February to 97 in May, and from 142 in August to 196 in December.

There were 10 million people in need of aid in November 2015. This number increased to more than 15 million in 2016 (OCHA 2016), including Ethiopians who receive food and cash transfers through the Productive Safety Net Programme. From a geographical point of view, most of the people in need of food assistance were in the Oromia region (3.7 million), followed by the Amhara (2.2 million), Somali (1.5 million) and Tigray regions (1.2 million) (HRD 2016).

According to the OCHA, there was a significant gap (Table 4) in required humanitarian funding throughout both years, especially in food assistance but also in other sectors such as health, nutrition, water and sanitation, agriculture and education. The greatest emergency was recorded after the failure of Kiremt rains and the subsequent Meher harvest in the autumn of 2015 when humanitarian needs skyrocketed by more than three times and donors released their aid with a delay of several months. In this case, there was clear evidence of the emergency.

Tab. 4 Humanitarian funding situation.

Period	Total needs (millions $)	Funding gap (millions $)	Available (millions $)
Jan 2015	386 (food 305 = 79%)	344	42
Apr 2015	386 (281 = 72%)	228	158
Aug 2015	432 (food 337 = 78%)	329	103
Jan 2016	1400 (food 1200 = 86%)	1042	358
Mar 2016	1400 (food 1200)	724	676
May 2016	1520	692	828
June 2016	1520 (food 1100 = 72%)	530	990
July 2016	1520 (food 1100)	545	975
Aug 2016	1520 (food 1100)	516	1004
Sept 2016	1620 (food 1100 = 68%)	582	1038
Nov 2016	1620 (food 1100)	557	1063
Dec 2016	1620 (food 1100)	537	1083

Source: OCHA Humanitarian Funding Updates and Humanitarian Requirements Documents (see appendix)

The Ethiopian government allocated $ 272 million in 2015 and $ 109 million in 2016 for humanitarian response. The biggest aid donors in order of donation size were:

Tab. 5 2015–2016 donations by donor.

Country	2015	2016
USA	$ 182 million	$ 619 million
UK	$ 76 million	$ 101 million
European Commission	$ 38 million	$ 181 million
Germany	$ 3 million	$ 88 million
Canada	$ 17 million	$ 54 million
Sweden	$ 11 million	$ 28 million
Japan	$ 14 million	$ 25 million
Netherlands	$ 17 million	$ 10 million

Source: OCHA FTS (2016), OCHA FTS (2017)

The United Nations provided humanitarian aid through the Central Emergency Response Fund CERF and specialized agencies such as the World Food Programme, the Food and Agricultural Organization and UNICEF ($ 27 million in 2015/$ 20 million in 2016). Aid was also provided, to a lesser extent, by private donors or non-governmental organizations ($ 17 million in 2015/$ 28 million in 2016) (OCHA 2016b).

5.6 Peaks

This archetype combines all successive spikes characterising famine. Howe (2010) concentrates on peaks of death, peak of food prices, peak of media attention and peak of deliveries. In the case of the 2015–2016 crisis, following peaks can be identified: food price inflation, migration (progressive in 2016) and media attention (two peaks). There was no peak of death related to aid magnets (or internal migration) and specific peak of aid deliveries.

6. Discussion and conclusion

The findings presented in the previous section have several implications. First, the FEWSNET was able to predict the impending emergency and as a result issued two alerts in 2015 (in June and December) calling for immediate emergency assistance. There were two subsequent periods of failed rains in 2015 which intensified the food insecurity of affected people not only after the lean season in June but also due to the limited Meher harvest which started in October. By July 2015, it was clear that the food crisis would turn into an emergency in the absence of external humanitarian aid and that the coping strategies would not be sufficient. The watch situation was therefore rather short, lasting only the first half of 2015. From June onwards, it was evident that the crisis would continue to deepen. Late responses cannot be attributed to the insecurity of the crisis evolution typical for the watch situation. Moreover, aid donors are long familiar with the situation in Ethiopia. Thus, subsequent waves of drought, warning reports and chronic food insecurity were clear signs for the Ethiopian government and the international donor community to respond on time.

Second, rises in food prices were recorded during the crisis. In 2015, food inflation reached two-digit values with the peak being in October (usual start of harvesting and time of low prices) and decreasing in 2016. The year 2015 was exceptional in this regard in comparison with 2014 and 2016. Local markets in the two most food-insecure regions recorded a significant rise in prices of sorghum (100% rise) and teff in the Mekele market. Wheat prices rose steadily from the end of 2014 to mid-2016. The Alert for Price Spikes issued two alerts in both lean seasons of 2015 and 2016 when prices are usually high. Thus, Howe's price spike archetype was recorded without a specific peak of all examined indicators.

Third, regarding the aid magnet situation, there were no excessive mortality rates in aid centers thanks to long-term operations of the UN humanitarian organisations. On the other hand, there was a sharp rise in internal displacement in Ethiopia, especially in 2016. Although Howe (2010) focuses on people's motivation to leave their homes (aid provision), this article highlights the causes of internal migration (disasters and conflicts). It cannot be unambiguously stated that internal migration was solely triggered by aid provision as displacement in Ethiopia was a present phenomenon. Displacement is caused also by competition over production resources and clashes between communities. In addition, extreme vertical hierarchies in the society reinforce geographic, economic and political marginalisation (Hammond 2011) and land confiscation or redistribution by regional authorities or federal government discredits traditional customary law (Abbink 2006). Land tenure stands for a highly sensitive political issue. During the examined food crisis, Ethiopia experienced political turmoil caused by ethnic and land conflicts. Protests in the Oromia region spread due to increasingly sensitive power divisions along ethnic lines (BBC 2014). Demonstrations against displacement of Oromo farmers across Oromia were suppressed with a disputed number of victims throughout 2015 when security forces were using 'unlawful force' and 'arbitrarily arrests' (HRW 2016). Other protests against unequal distribution of power and economic benefits spread to Amhara and SNNPR in 2016 (HRW 2017). In order to subdue the 'anti-peace forces', the government declared the state of emergency in October 2016 (BBC 2017). It is important to emphasize that these regions were hit the most by the 2015–2016 food crisis because access to land and production resources represents an indivisible part of food production and food security. Its loss and internal conflict have severe consequences on the state of food insecurity and they further deepened the discussed food crisis.

Fourth, media frenzy was proved. World media reacted to the prediction of FEWSNET in October when the food crisis was already ongoing. There were two media output peaks (October/November 2015 and March 2016) but otherwise media attention went down gradually. Even though a severe emergency was foreseen, there was no official prognosis of famine outbreak by any warning system. We can say that the famine threat was rather a media bubble to catch attention but, in any case, this claim should not underestimate the situation perceived by some Ethiopians themselves whose stories, when presented in the media, were reminiscent of the year 1984. Also, no other links to the crisis were made and drought caused by the El Niño confirmed the generally narrow point of view in relation to food crises. The most active media outlets were the Guardian, Reuters and the BBC followed by RFI and Le Monde. RFI also informed about the Ethiopian government's efforts to make press and non-governmental organizations remained silent to not spoil the picture of a booming economy or associate the country with misery (RFI 2015a). RFI also claimed that some drought-affected areas were almost impossible to reach by foreign journalists (RFI 2015b). The political influence on media and the non-governmental sector over providing accurate information represents a major hindrance and unacceptable intervention in crisis management. In this case, the media were able to report on the crisis despite Ethiopian efforts to conceal the reality. This fact supports the argument of Devereux (2007) about the political nature of famines which applies to severe food crises as well.

Fifth, humanitarian needs in Ethiopia skyrocketed in autumn 2015 but most of the aid came with several month-delays in the first quarter of 2016. There were also significant gaps in funding. The greatest gap was recorded in December 2015 when the total requirements reached $ 1.4 billion and the gap was $ 1.042 billion. This fact suggests that Howe's situation of overshoot emergency response was recorded during this food crisis but on the other hand, it does not confirm the condition that there was '*too much aid too late*'. It is not clear whether ramping up aid without fulfilling all humanitarian needs suffices to realisation of this archetype. The response rose in connection with growing requirements in 2016 but funding gaps persisted. Personally, I am inclined to think that the overshoot archetype is less applicable in this case. The late response might be attributed to donors' stretched budgets due other urgent crises (Syria, Yemen, Iraq or South Sudan) or their fatigue. It is also important to emphasize, however, that the funding was adequate to prevent a full-scale famine. What seems to be problematic is the aid structure. As Sandstrom and Juhola (2017) point out, most of the humanitarian budget is usually comprised by food aid as it was in the case of the Ethiopian food crisis in 2011 (58%) and during the emergency in Kenya and Somalia in 2004 to 2005 (over 80%). Numbers from the 2015–2016 food crisis also confirm this trend with food aid ranging from

68% to 86%. A stronger focus should be put on other sectors linked to food insecurity such as health, nutrition, education, agriculture or water, sanitation and housing.

Sixth, the dynamics of the food crisis of 2015 to 2016 does correspond to Howe's peaks archetype, to a lesser extent. The most urgent need of humanitarian assistance can be identified in the period of autumn and winter 2015. A rise in food prices as well as internal displacement was rather progressive processes. Media attention was experienced in two peaks and there was no peak in deliveries. Yet, the progressive characteristic of the crisis suggests a long-term food crisis in Ethiopia. It is also necessary to highlight that Ethiopia suffered from another wave of drought exacerbated by disease outbreaks and displacement in 2017. Eight and a half million Ethiopians faced crisis-level food insecurity (OCHA 2017). Parts of Amhara, Tigray, Oromia and SNNPR remained in IPC Phase 3 in September and an emergency (IPC Phase 4) situation was recorded in parts of the Somali region with delayed humanitarian assistance (FEWSNET 2017). The humanitarian needs of the country were $ 1.471 billion ($ 893 for food) with a $ 267 million gap by the end of 2017 (OCHA 2018).

There was no famine outbreak but on the other hand, famine was never officially predicted by any warning system. Application of the Howe's (2010) archetypal framework suggests its relevance in famine threat situations, although the magnitude of particular archetypes and their peaks may be more significant in full-scale famines. Even though not all archetypes took place, the dynamics of the 2015–2016 crisis proved Howe's argument of belated responses to crises. Furthermore, the tendency to see the failure of food systems only due to negative weather conditions with no acknowledgment of the complexity of food insecurity leads to limited impacts of humanitarian activities. There is an acute need for changes in humanitarian thinking towards more systemic and complex solutions. As the political context deepens the food crisis in Ethiopia, chronic food insecurity and the probability of future food emergencies in the country remain very high.

Acknowledgements

The author gratefully acknowledges comments from the anonymous reviewers. This paper was undertaken through the financial support of the specific university research (SVV) of Charles University.

Appendix

OCHA Humanitarian Funding Updates and Humanitarian Requirements documents

Ethiopia: Humanitarian Snapshot (as of January 2015)
Ethiopia: Humanitarian Snapshot (as of 15 April 2015)
Ethiopia: Humanitarian Snapshot (as of 31 August 2015)
Ethiopia: Humanitarian Snapshot (as of 4 December 2015)
Humanitarian Requirements Document (as of 5 January 2016)
Humanitarian Funding Update (as of Mar 30, 2016)
Humanitarian Funding Update (as of 17 May 2016)
Humanitarian Funding Update (as of 8 June2016)
Humanitarian Funding Update (as of 13 July 2016)
Humanitarian Funding Update (as of 5 August 2016)
Humanitarian Funding Update (as of 21 Sept 2016)
Humanitarian Funding Update (as of 30 December 2016)

References

Abbink, J. (2006): Ethnicity and Conflict Generation in Ethiopia: Some Problems and Prospects of Ethno-Regional Federalism. Journal of Contemporary African Studies 24(3), 389–413. https://doi.org/10.1080/02589000600976729

Brown, M. E. (2008): Case Study: Ethiopia and the 2002–2003 Food Security Crisis. Famine Early Warning Systems and Remote Sensing Data. Springer, Berlin Heidelberg.

Cutler, P. (1991): The Political Economy of Famine in Ethiopia and Sudan. Ambio 20(5), 176–178.

De Waal, A. (1991): Evil Days: 30 year of war and famine in Ethiopia. New York, Washington and London, Africa Aid Watch Report.

Devereux, S. (2009): Why does famine persist in Africa? Food Security 1(1), 25–35. https://doi.org/10.1007/s12571-008-0005-8

Devereux, S. (eds.) (2007): The New Famines: Why Famines Persist in an Era of Globalization. London and New York, Routledge.

Devereux, S. (2006): Distinguishing between chronic and transitory food insecurity in emergency needs assessment. Rome, World Food Programme.

Devereux, S., Howe, P., BIONG DENG, L. (2002): Introduction: The "New Famines". IDS Bulletin 33(4), 1–11. https://doi.org/10.1111/j.1759-5436.2002.tb00038.x

Edkins, J. (2002): Mass Starvations and the Limitations of Famine Theorising. IDS Bulletin 33(4), 12–18. https://doi.org/10.1111/j.1759-5436.2002.tb00039.x

FAO (2008): Integrated Food Security Phase Classification. Technical Manual, Version 1.1. Rome. [Online]: http://www.fao.org/docrep/010/i0275e/i0275e.pdf [cit. 15. 12. 2017]

Hammond, L. (2011): Governmentality in motion: 25 years of Ethiopia's experience of famine and migration policy. Mobilities 6(3), 415–432. https://doi.org/10.1080/17450101.2011.590038
Hammond, L., Maxwell, D. (2002): The Ethiopian Crisis of 1999–2000: Lessons Learned, Questions Unanswered. Disasters 26(3), 262–279. https://doi.org/10.1111/1467-7717.00204
Howe, P. (2018): Famine systems: A new model for understanding the development of famines. World Development, 105, 144–155. https://doi.org/10.1016/j.worlddev.2017.12.028
Howe, P. (2010): Archetypes of famine and responses. Disasters 24(1), 20–54. https://doi.org/10.1111/j.1467-7717.2009.01113.x
Howe, P. (2002): Reconsidering "Famine". IDS Bulletin 33(2), 19–27. https://doi.org/10.1111/j.1759-5436.2002.tb00040.x
Howe, P., Devereux, S. (2004): Famine Intensity and Magnitude Scales: A Proposal for an Instrumental Definition of Famine. Disasters 28(4), 353–372. https://doi.org/10.1111/j.0361-3666.2004.00263.x
Khalif, M. H., Doornbos, M. (2002): The Somali region in Ethiopia: a neglected human rights tragedy. Review of African Political Economy 29(91), 73–94. https://doi.org/10.1080/03056240208704585
Marcus, D. (2003): Famine Crimes in International Law. The American Journal of International Law 97(2), 245–281. https://doi.org/10.2307/3100102
Menkhaus, K. (2012): No access: Critical bottlenecks in the 2012 Somali famine. Global Food Security 1(1), 29–35. https://doi.org/10.1016/j.gfs.2012.07.004
Miller, D., Holt, J. F. (1975): The Ethiopian famine. Proceedings of the Nutrition Society 32(3), 167–172. https://doi.org/10.1079/PNS19750034
Rivers, J. P. W., Holt, J. F., Seaman, J. A., Bowden, M. R. (1976): Lessons for Epidemiology from the Ethiopian Famines. Annales de la Société Belge de Médicine Tropicale 4–5(56), 345–357.
Rubin, O. (2018): The Precarious State of Famine Research. The Journal of Development Studies. https://doi.org/10.1080/00220388.2018.1493196
Sabates-Wheeler, R., Devereux, S. (2010): Cash transfers and high food prices: Explaining outcomes on Ethiopia's Productive Safety Net Programme. Food Policy 35(4), 274–285. https://doi.org/10.1016/j.foodpol.2010.01.001
Sandstrom, S., Juhola, S. (2017): Continue to blame it on the rain? Conceptualization of drought and failure of food systems in the Greater Horn of Africa. Environmental Hazards 16(1), 71–91. https://doi.org/10.1080/17477891.2016.1229656
Sen, A. (1981a): Poverty and Famines: An Essay on Entitlement and Deprivation. Oxford, Claredon Press.
Sen, A. (1981b): Ingredients of Famine Analysis: Availability and Entitlements. The Quarterly Journal of Economics 96(3), 443–464. https://doi.org/10.2307/1882681
Ververs, M.-T. (2012): The East African Food Crisis: Did Regional Early Warning Systems Function? The Journal of Nutrition 142(1), 131–133. https://doi.org/10.3945/jn.111.150342

Other sources

BBC (2017): Ethiopia rights body: "More than 600 protest death". [Online] http://www.bbc.com/news/world-africa-39619979 [cit. 15. 12. 2017]
BBC (2014): Anger over 'violent crackdown' at protest in Ethiopia. [Online] http://www.bbc.com/news/av/world-africa-28522986/anger-over-violent-crackdown-at-protest-in-ethiopia [cit. 15. 12. 2017]
EPA (1998): National Action Programme to Combat Desertification. Addis Ababa: Environmental Protection Authority Ethiopia. [Online] http://www.unccd.int/ActionProgrammes/ethiopia-eng2000.pdf [cit. 26. 5. 2016]
FAO (2017): FAOSTAT Ethiopia. [Online] http://www.fao.org/faostat/en/#country/238 [cit. 30. 10. 2017]
FAO (2012): IPC Integrated Food Security Phase Classification, Technical Manual Version 2.0: Evidence and Standards for Better Food Security Decisions. Rome.
FAO (2008): IPC Integrated Food Security Phase Classification, Technical Manual Version 1.1: Evidence and Standards for Better Food Security Decisions. Rome.
FAO GIEWS (2017): FAO GIEWS FPMA Tool. [Online] http://www.fao.org/giews/food-prices/tool/public/#/dataset/domestic [cit. 10. 12. 2017]
FAO GIEWS (2015): GIEWS Country Briefs Ethiopia. [Online] http://www.fao.org/giews/countrybrief/country.jsp?code=ETH [cit. 10. 12. 2015]
FEWSNET (2015a): Food security emergency in central/eastern Ethiopia follows worst drought in more than 50 years. [Online] http://www.fews.net/east-africa/ethiopia/alert/december-4-2015 [cit.10. 12. 2015]
FEWSNET (2015b): Food Security Outlook: Ethiopia. [Online] http://www.fews.net/east-africa/ethiopia/food-security-outlook/october-2015 [cit. 27. 11. 2015]
FEWSNET (2015c): Alert: Well below average Belg harvest likely in central SNNPR. [Online] http://www.fews.net/east-africa/ethiopia/alert/june-26–2015 [cit. 28. 11. 2015]
FEWSNET (2015d): Ethiopia Price Bulletin January 2015 [Online] http://www.fews.net/sites/default/files/documents/reports/Ethiopia_2015_01_PB.pdf [cit. 11. 10. 2017]
FEWSNET (2013): East Africa, Ethiopia: Seasonal calendar, Typical Year. [Online]
http://www.fews.net/east-africa/ethiopia/seasonal-calendar/december-2013 [cit. 9. 10. 2017]
FEWSNET (2011): Ethiopia Food Security Update. [Online] http://fews.net/sites/default/files/documents/reports/ET_dekadal_2011_07_26_final.pdf [cit. 16. 8. 2018]
HRD (2016): Joint Government and Humanitarian Partners' Document: Ethiopia 2016 Humanitarian Requirements Document. [Online] https://reliefweb.int/sites/reliefweb.int/files/resources/ethiopia_hrd_2016.pdf [cit. 11. 12. 2017]
HRW (2017): World Report 2017. [Online] https://www.hrw.org/world-report/2017/country-chapters/ethiopia [cit. 10. 11. 2017]
HRW (2016): World Report 2016. [Online] https://www.hrw.org/world-report/2016/country-chapters/ethiopia [cit. 10. 11. 2017]
IDMC (2016): Global Report on Internal Displacement. [Online] http://www.internal-displacement.org

/assets/publications/2016/2016-global-report -internal-displacement-IDMC.pdf [cit. 18. 11. 2017]
IOM (2016): Internal Displacement Monitoring Report: October – December 2015. [Online] https://reliefweb.int/sites/reliefweb.int/files/resources/idmr_-_oct_to_dec_2015.pdf [cit. 18. 11. 2017]
OCHA (2018): Ethiopia: Humanitarian funding update (as of 11 January 2018). [Online] https://reliefweb.int/sites/reliefweb.int/files/resources/hrd_funding_for_2017_-_december_31_as_of_jan_11_2018v1.pdf [cit. 10. 2. 2018]
OCHA (2017): FSNWG Snapshot: Food Security & Nutrition Situation (as of 19 October 2017). [Online] https://reliefweb.int/sites/reliefweb.int/files/resources/FSNWG_EA_snapshot_18102017.pdf [cit. 9. 12. 2017]
OCHA (2016): Ethiopia: Humanitarian funding update (as of 30 December 2016). [Online] https://reliefweb.int/sites/reliefweb.int/files/resources/hrd_funding_update_latest_30_dec_2016.pdf [cit. 9. 12. 2017]
OCHA (2015): Ethiopia: Humanitarian Snapshot (as of 4 December 2015). [Online] https://reliefweb.int/report/ethiopia/ethiopia-humanitarian-snapshot-4-december-2015 [cit. 9. 12. 2017]
OCHA FTS (2017): Ethiopia emergencies for 2016. [Online] https://fts.unocha.org/countries/71/donors/2016?order=total_funding&sort=desc [cit. 27. 12. 2017]
OCHA FTS (2016): Ethiopia emergencies for 2015. [Online] https://fts.unocha.org/countries/71/donors/2015?order=total_funding&sort=desc [cit. 27. 12. 2017]
OECD (2016): Aid and at a glance charts. [Online] https://public.tableau.com/views/AidAtAGlance/DACmembers?:embed=y&:display_count=no?&:showVizHome=no#1 [cit. 30. 10. 2017]
RFI (2015a): Ethiopie: les besoins en aide alimentaire augmentent. [Online] http://www.rfi.fr/afrique/20151003-ethiopie-besoins-aide-alimentaire-augmentent-faim-malnutrition-onu [cit. 3. 12. 2017]
RFI (2015b): Cette famine annoncée que l'Ethiopie ne veut pas voir. [Online]
http://www.rfi.fr/afrique/20151009-secheresse-faim-menace-ethiopie-climat-famine-afar [cit. 9. 12. 2017]
UN Committee for Development Policy (2017): List of Least Developed Countries. [Online] https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/ldc_list.pdf [cit. 29. 11. 2017]
UNICEF (2017): UNICEF Annual Report 2016. [Online] https://www.unicef.org/about/annualreport/files/Ethiopia_2016_COAR.pdf [cit. 18. 8. 2018]
UNICEF (2016): Ethiopia: Situation Report #9. [Online] https://reliefweb.int/sites/reliefweb.int/files/resources/UNICEF%20Ethiopia%20Humanitarian%20SitRep%20%239%20-%20October%202016.pdf [cit. 13. 9. 2018]
UNICEF (2016b): UNICEF Annual Report 2015. [Online] https://www.unicef.org/about/annualreport/files/Ethiopia_2015_COAR.pdf [cit. 13. 8. 2018]
UNICEF (2015): Ethiopia: Situation Report #5. [Online] https://reliefweb.int/sites/reliefweb.int/files/resources/UNICEF%20Ethiopia%20Humanitarian%20Situation%20Report-%20May%202018.pdf [cit. 13. 9. 2018]
WB (2017): Ethiopia Overview. [Online] http://www.worldbank.org/en/country/ethiopia/overview [cit. 29. 11. 2017]
WB (2016): World Bank national accounts data [Online] https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?locations=ET [cit. 30. 10. 2017]
WFP (2017): Alert for Price Spikes (ALPS). [Online] http://foodprices.vam.wfp.org/ALPS-chart.aspx?iso=ETH [cit. 27. 12. 2017]
WFP (2015): Ethiopia Monthly Market Watch December 2015. [Online] http://documents.wfp.org/stellent/groups/public/documents/ena/wfp281062.pdf?iframe [cit. 28. 12. 2017]
WFP (2012): How High Food Prices Affect the Worlds' Poor. [Online] https://www.wfp.org/stories/how-high-food-prices-affect-worlds-poor [cit. 14. 8. 2018]
WHO (2015): Ethiopia needs major health support to stop devastation from El Niño. [Online] http://www.who.int/hac/crises/eth/releases/4december2015/en/ [cit. 10. 12. 2015]
World Food Summit (1996): Rome Declaration on World Food Security. [Online] http://www.fao.org/docrep/003/w3613e/w3613e00.HTM [cit. 22. 11. 2016]

Approaches for land cover monitoring over Europe based on backscatter and coherence properties of Envisat and ERS SAR data

Kyriakos Ntouanoglou[1,*], Antonios Mouratidis[2]

[1] Hellenic Air Force, Greece
[2] Aristotle University of Thessaloniki, Department of Physical and Environmental Geography, Thessaloníki, Greece
* Corresponding author: kyrntouan@yahoo.gr

ABSTRACT

The purpose of this study is the exploitation of Synthetic Aperture Radar (SAR) satellite data for land cover monitoring. The properties of the recorded radar data heavily depend on the type of land cover, season and weather conditions. Therefore, it is possible to utilize the variability of these factors, in order to develop various techniques and methodologies that can be used for classifying land surfaces. In this context, this paper proposes approaches, which include the application of mathematical expressions and application of thresholds on multi-temporal data, for recognizing and classifying various types of land cover, on the basis of ERS and Envisat C-band SAR backscatter and coherence properties. These can be useful for any kind of contemporary SAR data, such as those of the current two Sentinel-1 satellites. Although this study focuses on four main land cover types (urban, mountainous, agricultural-low vegetation and forested areas) over specific areas in Europe, the same principles can be extended worldwide, leading to useful insights for designing future SAR satellite missions or for establishing guidelines for in-depth studies of specific land cover types.

KEYWORDS

remote sensing; SAR; land cover; classification; backscatter; coherence

Received: 11 January 2018
Accepted: 12 October 2018
Published online: 17 December 2018

Ntouanoglou, K., Mouratidis, A. (2018): Approaches for land cover monitoring over Europe based on backscatter and coherence properties of Envisat and ERS SAR data.
AUC Geographica 53(2), 207–219
https://doi.org/10.14712/23361980.2018.20

1. Introduction

1.1 Background and Scope

A plethora of Space-based Remote Sensing (RS) data for almost the entire Earth are nowadays available, providing access to different information for a variety of users worldwide. Land Cover (LC) classification is amongst the most popular applications of such data. Nevertheless, the "Achilles heel" of all RS-derived information is its accuracy, reliability and the challenge for their constant increase (Kaoru and Toshiharu 1996).

Land cover classification can be defined as "the ordering or arrangement of objects into groups or sets on the basis of their relationships" (Sokal 1974). Typically, the faster it is performed, the lower classification accuracy is to be expected (Ackermann 2011).

The most important factor to be considered when choosing a classification approach or designing a classification system, is the user needs. Variables such as scale, resolution and study area also have to be taken into account. In general, classification techniques can be grouped as supervised or unsupervised, hard or soft (fuzzy), per-field, sub-pixel or per-pixel and parametric or non-parametric (Lu and Weng 2007).

Additionally, it is highly advisable that the data, format and classification procedures are determined in advance. For example, if a single-date image is to be used, unlike in a time-series approach, there is no need for atmospheric corrections (Song et al. 2001). On the other hand, if the study area is rugged or mountainous, topographic correction is an important aspect to consider (Hale and Rock 2003).

An alternative to the most typically used optical (visible and infrared) bands for LC classification is the microwave (radar) part of the electromagnetic spectrum, which provides a different kind of information on the observed objects (Lusch 1999). The two main types of microwave RS are a) active and b) passive. In the former case, the radar antenna transmits its own signal and eventually records part of its return, after "reflection" on the ground surface. On the other hand, passive radar systems do not send their own signal, but only receive and record the microwave radiation emitted from the ground objects (Long 2008).

This paper deals with the active type of radar and in particular with Synthetic Aperture Radar (SAR), typically used since almost 30 years for the purposes of Earth observation from Space. Just as in optical remote sensing, there are several ways of processing and classifying SAR images. Nevertheless, a special characteristic of SAR is that – unlike optical imaging, where the user has several spectral bands available for interpretation – typically only one, two or in the best case a maximum of four bands (for fully polarimetric SAR data, which are generally rare) are available (Haack et al. 2000).

With a focus on LC classification, the objective of this study is to analyze and understand the temporal de-correlation effect of the recorded SAR signal – commonly known as backscattering, as well as the stability of the SAR signal phase – quantified as coherence values. Phase coherence is one of the main products of SAR interferometry (InSAR) and there are several factors that may affect it (Hanssen 2001):

1. The unknown integer number of phases (2kπ).
2. The phase component due to topography.
3. The phase component due to distortion in the observation direction of the radar.
4. The phase component due to surface reference and errors of satellite orbit.
5. The phase component due to atmospheric delay to which the signal is subject.
6. The phase component due to any changes in the scattering characteristics of the Earth's surface between two observations.
7. The phase component due to all kinds of noise, such as thermal, the error of images writing or interference errors.

With reference to the sixth aforementioned factors, temporal changes over a ground target may occur, owing to geometric (e.g. due to wind) and/or dielectric variations (e.g. due to precipitation). This has an impact – to a different extent – on many LC types observed by a SAR sensor, which include e.g. the decrease of coherence magnitude or the increase of phase noise (Lavalle 2013). The stability of a SAR target, as denoted by the absence of backscatter and/or coherence change, may also be particularly informative and helpful for classification purposes (e.g. in the case of the built environment).

In this context, this paper proposes approaches for recognizing and classifying various types of LC, on the basis of SAR backscatter and coherence properties.

The study areas, in each of which the focus is on a particular type of terrain (urban, mountainous, agricultural and low vegetation areas, forests), are spread across Europe (Fig. 1).

In all cases, Envisat/ASAR Single Look Complex (SLC) data have been used, with the exception of forested areas, where the use of ERS data has been employed. The processing of the SAR data has been carried out using the Next ESA SAR Toolbox (NEST) and the ENvironment for Visualizing Images (ENVI™) software.

In order to process SAR images and classify them, there are several steps to be followed. One drawback of radar imaging is the lack of many different spectral bands for each acquisition. Nevertheless, alternative approaches, based on the inherent properties of the SAR signal, can be employed for tackling this obstacle. Thus, before proceeding to the description of the methodological approach of this study, a short description of relevant SAR characteristics is presented.

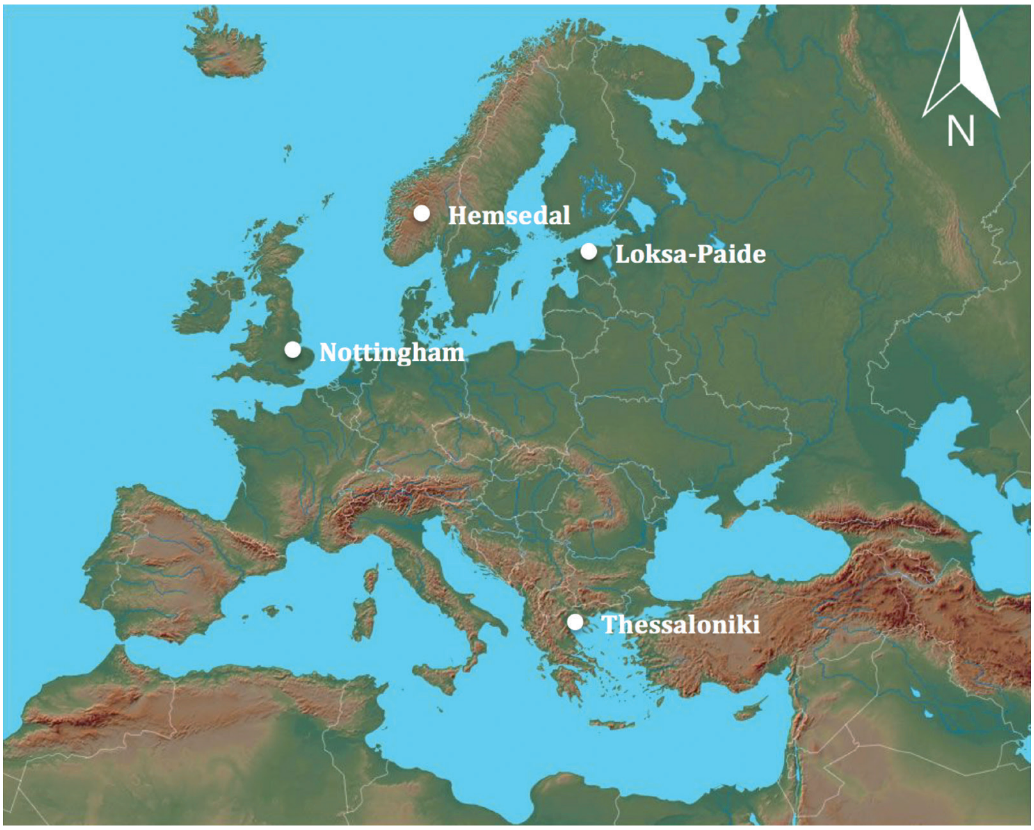

Fig. 1 Study areas across Europe (Geology.com 2016).

1.2 Overview of main SAR characteristics

What occurs when the antenna that transmits is also the one that receives the resulting single-signal power (for cases of point targets) is explained by Eq. 1 (Toomay and Hannen 2004), which denotes the relationship between the different parameters:

$$P_r = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4}, \quad (1)$$

where:
P_r = power received (watts),
σ = target radar cross section square meters (m^2),
P_t = power transmitted (watts),
G = antenna gain (dimensionless),
R = radar to object range (m),
λ = wave length (m).

In Eq. 1, the first part of the fraction $(\frac{P_t G^2}{(4\pi)^3})$ depends on the system and along with the second part $(\frac{\lambda^2}{R^4})$, which is related to the signal propagation, it can be determined by calibration procedures. Thus, the remaining coefficient (σ) contains the highest level of uncertainty and is also the most interesting one for the purposes of this study.
In the case of an extended target, such as the area (A) of the resolution cell/pixel of the radar, "σ" is represented by the backscattering coefficient (σ^0), which is the average value of the radar reflective area per area unit $(\frac{d\sigma}{dA})$, expressed in m^2/m^2. As a result, Eq. 1 can be modified as follows:

$$P_r = P_t (\sigma^0 A) \left(\frac{G^2 \lambda^2}{(4\pi)^3 R^4} \right), \quad (2)$$

As per the parameter of coherence, consider the simple geometry shown in Fig. 2, including two radar sensors (SAR_1 and SAR_2) with a zero baseline, look angle (θ), along-track distance (x), across-track distance (y), radial distance (r), vertical elevation above the surface (z), center of the radar resolution cell (O) and location (P) of the resolution cell at an arbitrary coordinate (x, y, z). If c is the correlated part of the signal and n the uncorrelated noise due to baseline, thermal, rotation, temporal and other unknown factors, then the measure for the calculation of the complex correlation coefficient γ (i.e. related to coherence) for radar data of the first and second acquisition (s_1 and s_2) is:

$$\gamma = \frac{|\langle s_1 s_2^* \rangle|}{\sqrt{\langle s_1 s_1^* \rangle \langle s_2 s_2^* \rangle}}, \quad (3)$$

where $s_1 = c + n_1$ and $s_2 = c + n_2$ (s^* represents the complex conjugate).

The main components contributing to the total decorrelation are: 1) thermal ($\gamma_{thermal}$ or γ_{SNR}), 2) spatial ($\gamma_{spatial}$ or γ_{proc}) and 3) temporal decorrelation (γ_{temp}).

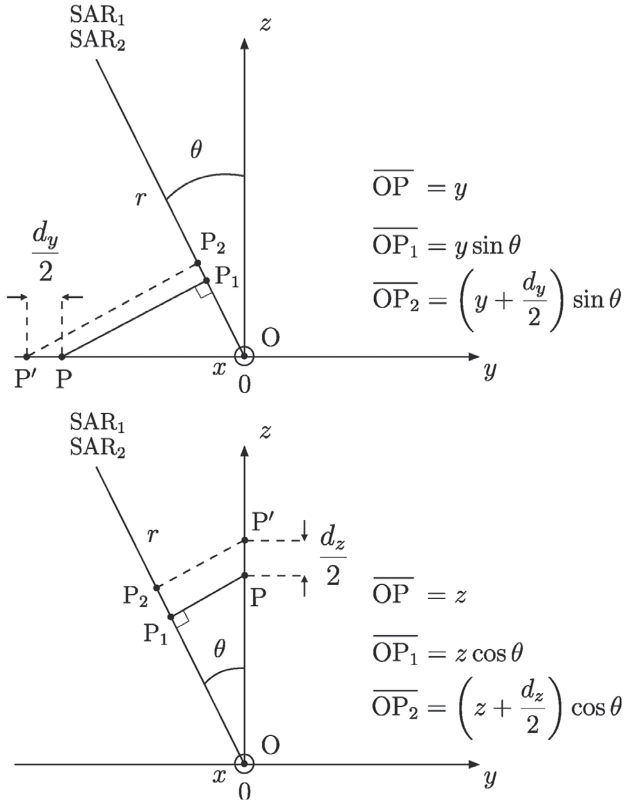

Fig. 2 Resolution cell and imaging geometry of repeat-pass radar interferometer with zero spatial baseline. The *P* point moves into *P′* between the acquisitions, the displacement of P and P′along y and z results in a phase offset that depends on the look angle θ (Lavalle et al. 2012).

The relationship between $\gamma_{thermal}$ and Signal to Noise Ratio (SNR) is (Wei and Sandwell 2010; Zebker and Villasenor 1992):

$$\gamma_{\mathrm{thermal}} = \frac{1}{1+\mathrm{SNR}^{-1}} = \frac{|c|^2}{|c|^2+|n|^2}, \tag{4}$$

Zebker and Villasenor (1992), give the Gaussian distributed temporal de-correlation equation:

$$\gamma_{\mathrm{temp}} = \exp\left[-\frac{1}{2}\left(\frac{4\pi}{\lambda_{\mathrm{eff}}}\right)^2\left(\sigma_y^2\sin^2\theta + \sigma_z^2\cos^2\theta\right)\right], \tag{5}$$

where $\lambda_{eff} = \frac{4\pi}{k}\cos\theta$, which introduces another type of correlation, due to the rotation of the target in relation to the radar look direction:

$$\gamma_{\mathrm{rotation}} = -\frac{2\sin\theta|d\varphi|R_x}{\lambda}, \tag{6}$$

where $d\varphi = \varphi_1 - \varphi_2$ is the distance after the small rotation.

Many parameters appear in the various situations, when researchers attach different notability to these components, such as Santoro et al. (2007), by decomposing the $\gamma_{spatial}$ to $\gamma_{surface}$ and γ_{volume}, because two images taken from different angles are being used. Furthermore, Siqueira et al. (2014), in order to be able to study flora through interferometry, calculates $\gamma_{obs\ (observation)}$ as,

$$\gamma_{\mathrm{vol(volume)}} = \frac{\gamma_{\mathrm{obs}}}{\gamma_{\mathrm{SNR}}\gamma_{\mathrm{spatial}}\gamma_{\mathrm{temp}}} = f(h_v), \tag{7}$$

where $h_v = f^{-1}(\gamma_{vol})$ and $\gamma_{vol} \leq \sin c\left(\frac{\kappa_z n_v}{2}\right)$, while k_z is the vertical wavenumber and h_v flora (e.g. tree) height. An InSAR system requires a pair of images taken from two different points of space, called S_1 and S_2 respectively. These two images must be co-registered and range filtered for uncommon parts of the spectra of S_1 and S_2, in order to increase the coherence (ρ):

$$\rho = \frac{E\{S_1S_2^*\}}{\sqrt{E\{|S_1|^2\}E\{|S_2|^2\}}}, \tag{8}$$

where *E* is the expectation value, |.| represents the absolute value, * refers to the complex conjugation, and the processes S_1, S_2, and $S_1S_2^*$ are assumed stationary and jointly stationary. Assuming that there are no multiple realizations of the images and $S_1S_2^*$ are ergodic, and with *m* and *n* referring to the image dimensions, while *M* and *N* refers to the averaged pixels in each dimension, then:

$$\hat{\rho}_{\mathrm{MLT}} = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} S_1(m,n)\,S_2^*(m,n)}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}|S_1(m,n)|^2\sum_{m=1}^{M}\sum_{n=1}^{N}|S_2(m,n)|^2}} = \frac{|\langle s_1s_2^*\rangle|}{\sqrt{\langle s_1s_1^*\rangle\langle s_2s_2^*\rangle}}, \tag{9}$$

where $|\hat{\rho}_{MLT}|$ corresponds to the maximum likelihood estimator of $|\rho|$. The calculation of the coherence |ρ| is possible for any type of multi-dimensional SAR data, as Interferometric Synthetic Aperture Radar (InSAR) or Polarimetric Synthetic Aperture Radar (PolSAR) (Martinez and Pottier 2005).

2. Methodology

2.1 Urban Areas

Typically, the built environment and the stable man-made structures within urban areas are characterized by high coherence (no or little change), as a result of the preservation of SAR signal phase. On the other hand, low coherence due to phase decorrelation may correspond to agricultural areas, dense high-growing vegetation (forest), layover areas and areas of low backscatter (smooth surfaces or steep back-slopes) (Barbieri and Lichtenegger 2005).

The selected data were acquired over the city of Thessaloniki, Greece, with about 1 million residents

Fig. 3 Study area 1: urban environment in Thessaloniki, northern Greece.

in its Larger Urban Zone (LUZ). More than 790 thousands of them are living in the main urban area, which covers about 112,000 km^2, while the construction network and build-up areas are very dense (Fig. 3). Elevations in the area range from MSL (Mean Sea Level) up to 1201 meters (Hellenic Statistical Authority 2011).

The methodology used for the urban areas refers to co-registered images of ascending and descending passes. More specifically, the SAR signal stability for an urban area is analyzed, by averaging the coherence of the two different passes.

Four ascending and six descending ASAR data during 2005 were used. Firstly, the methodology sequence (Fig. 4) consisted of a separate coregistration of the ascending and descending datasets, after applying precise orbits. Consequently, two different coregistered stacks were created – one for ascending and one for descending pass – and then the average coherence was estimated for each of the two passes. The coherence window size was 10 pixels (azimuth) × 2 pixels (range).

In order to maximize the coherence differences between urban and non-urban areas, a multiplication process took place in two steps. Initially, after studying the relevant histograms and identifying the minimum coherence values for each pass, all the pixels were multiplied with the same integer number, in order to obtain values higher than one. That is, if e.g. the minimum coherence was 0.015 then it was multiplied by 100, in order to obtain a value higher than 1 (0.015 × 100 = 1.5). Subsequently, each resulting image was multiplied by itself as many times as considered adequate to achieve the desired enhanced output. That is, the multiplication power depended on the coherence value differences between urban and non-urban areas, i.e. the smaller the original difference the higher the power of multiplication needed to maximize it. For better visualization purposes, the image contrast was also manipulated accordingly. In the next step the two products were orthorectified.

Ultimately, a secondary product was computed, i.e. the average coherence from both ascending and descending data at a resolution of 20 m × 20 m, which

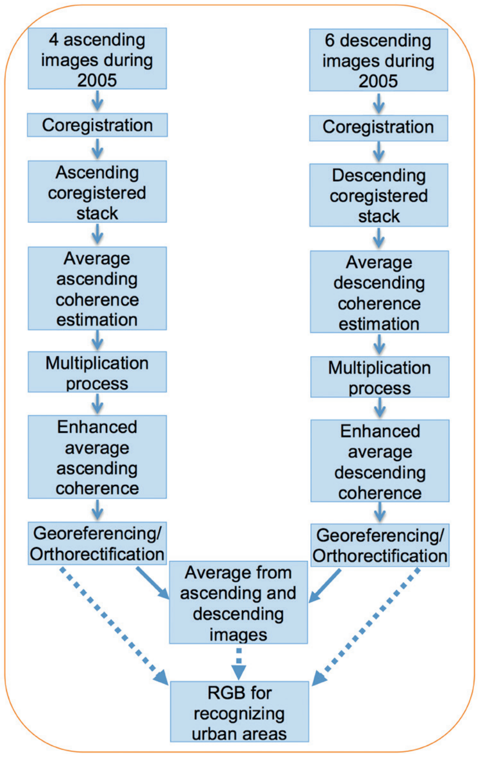

Fig. 4 Methodology followed for the identification of urban areas.

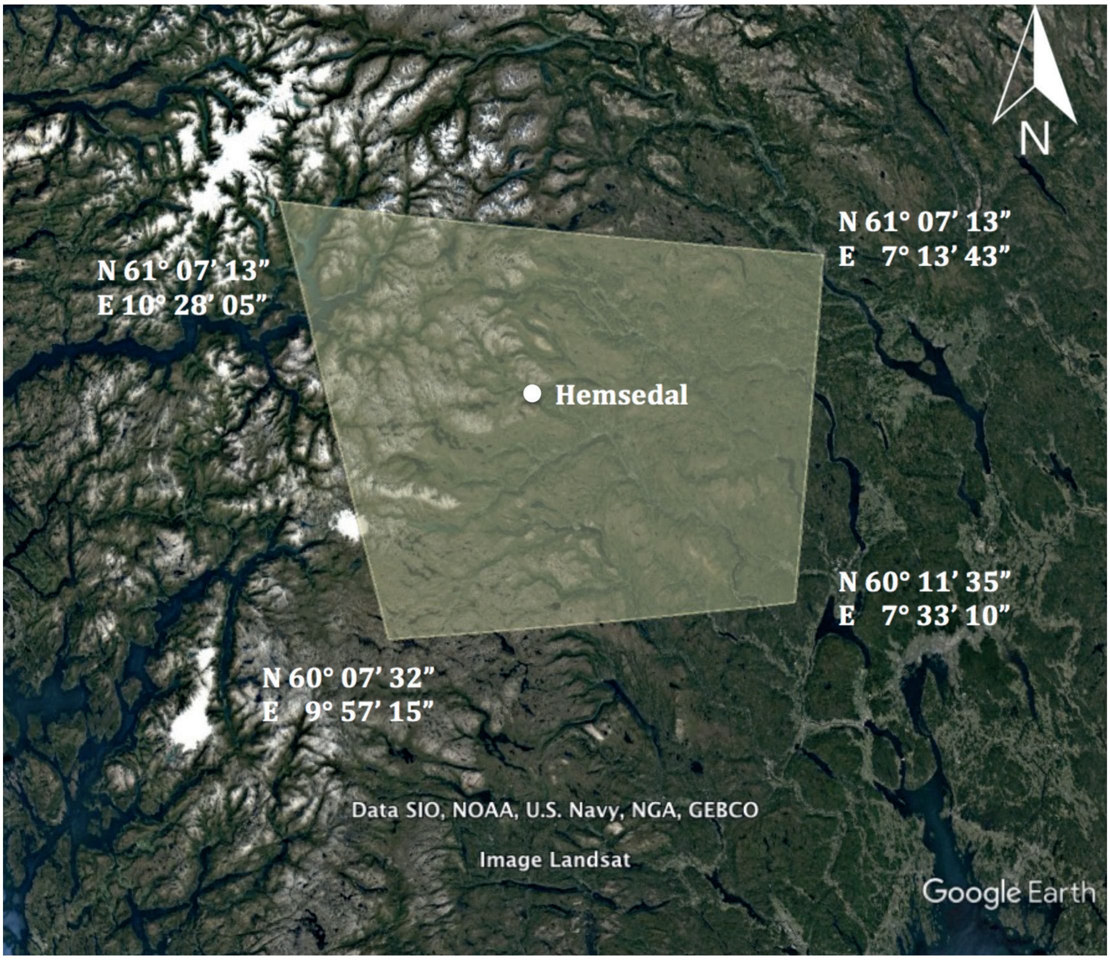

Fig. 5 Study area 2: mountainous terrain near Hemsedal, Norway.

was also used in the creation of an ensemble RGB product for recognizing urban areas.

2.2 Mountainous Terrain

For the analysis of mountainous terrain that is covered by snow or ice, images from Norway and especially images taken near the area of Hemsedal (Fig. 5) have been chosen for backscatter analysis.

In order to increase the reliability of the results, data for three years have been incorporated, i.e. from 2003 to 2005. Taking climatic conditions into account, the acquisition dates of these images were separated in three periods; a) from January to February b) from July to August and c) during May. Both ascending and descending passes were used.

The first period corresponds to winter conditions, during which the mountains in the area are covered by snow and ice. The second period concerns summer conditions, characterized either by high soil moisture content (stemming from the melted snow and ice) or by the presence of bare rocks. In this respect, relatively high backscatter values are expected in the aforementioned two periods (Lu and Meyer 2002). Regarding May, according to the World Weather & Climate Information (2014), it is the driest month in the area, hence significantly lower backscatter values are to be expected.

In order to distinguish the backscatter differences and observe their monthly variances, both descending and ascending passes were used. Separated into semesters and pass type, the images were calibrated, resulting in the calculation of actual backscatter values in db (decibel). Subsequently images were coregistered and orthorectified and statistics were exported for each image category (asceding-descending) (Fig. 6).

The most important element in the analysis is the Snow Water Equivalent (SWE), because of the information on the differences in the backscatter that it

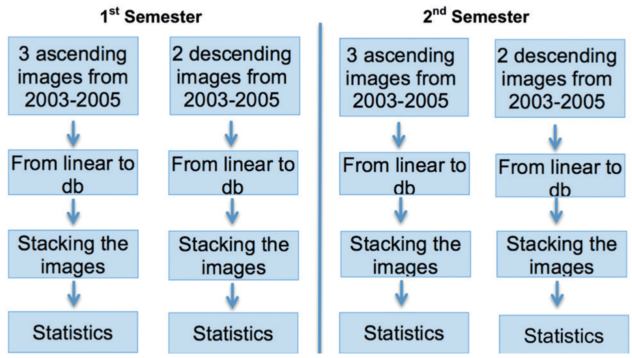

Fig. 6 Methodology followed over mountainous areas.

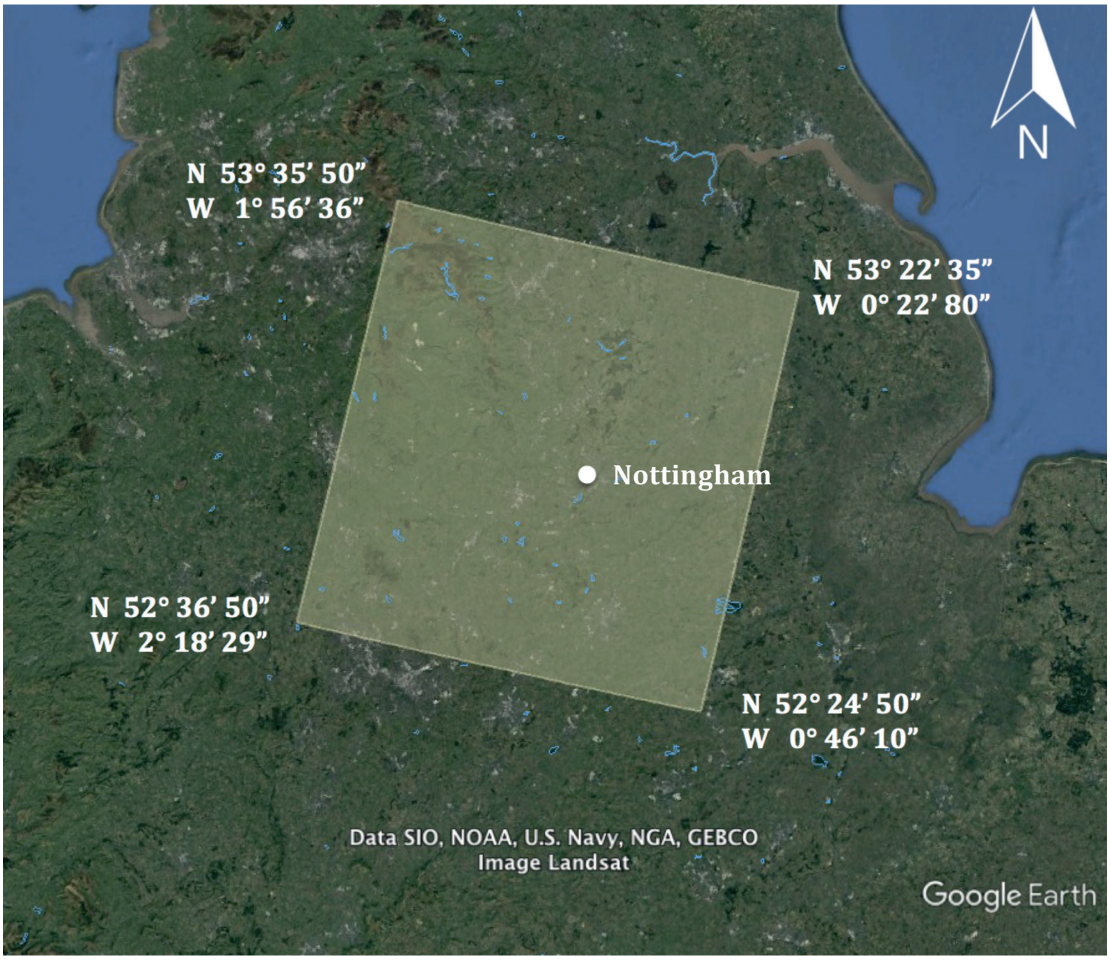

Fig. 7 Study area 3: agricultural and low vegetation in the vicinity of Nottingham, UK.

provides (Storvold et al. 2006). A set of experiments carried out by using ERS-1 data (5.3 GHz, 23° incidence angle) in the Norwegian mountains, has shown that the Snow Covered Area (SCA) could be monitored, as – in special conditions – the snow-free terrain gives high backscatter. Nevertheless, these conditions are very rare and for operational purposes it is better not to use the aforementioned incidence angle of ERS (Solberg 1993). In most cases the snow-covered terrain gives slightly higher backscatter compared to bare ground, but it also depends significantly on the incidence angle and on the snow wetness; "semi-wet" snow results in slightly reduced backscatter (Malnes and Guneriussen 2002).

2.3 Agricultural And Low Vegetation Areas

In order to take sample images of agricultural and low vegetation areas, one such area near Nottingham, UK, has been chosen. One reason for this choice lies in the fact that in the UK agriculture occupies approximately 70% or 9.2 million hectares of its territory (BBC 2007).

Based on the image availability from Envisat ASAR and Land Cover (LC) maps of 2007 (Morton et al. 2011), the area of Central East UK was selected, including most of the Midlands and a little of Eastern areas. The highest point in these areas is Kinder Scout, with 636 meters elevation and an average altitude of almost 150 meters (Fig. 7).

Maps of agricultural production of the years 2000 and 2010 were used (Defra 2006, 2013). Although there are no maps available for the period 2007–2009 (which corresponds to the SAR image dates), it can be assumed that there are no significant changes in agricultural production between 2000 and 2010.

The major part of agricultural production consists of wheat, barley and maize, followed by oilseed rape and potatoes, while sugar beet, potatoes, peas and field beans are also being farmed (Defra 2013, UK Agriculture 2014). The cultivation and harvesting season for the main four products is shown in Table 1.

Tab. 1 The harvesting and cultivation periods for the products studied.

Products	Harvesting period	Cultivation period
Wheat	August	October
Barley	June	October
Maize	April	September
Oilseed rape	Late July	Late August

Source: UK Agriculture, 2014

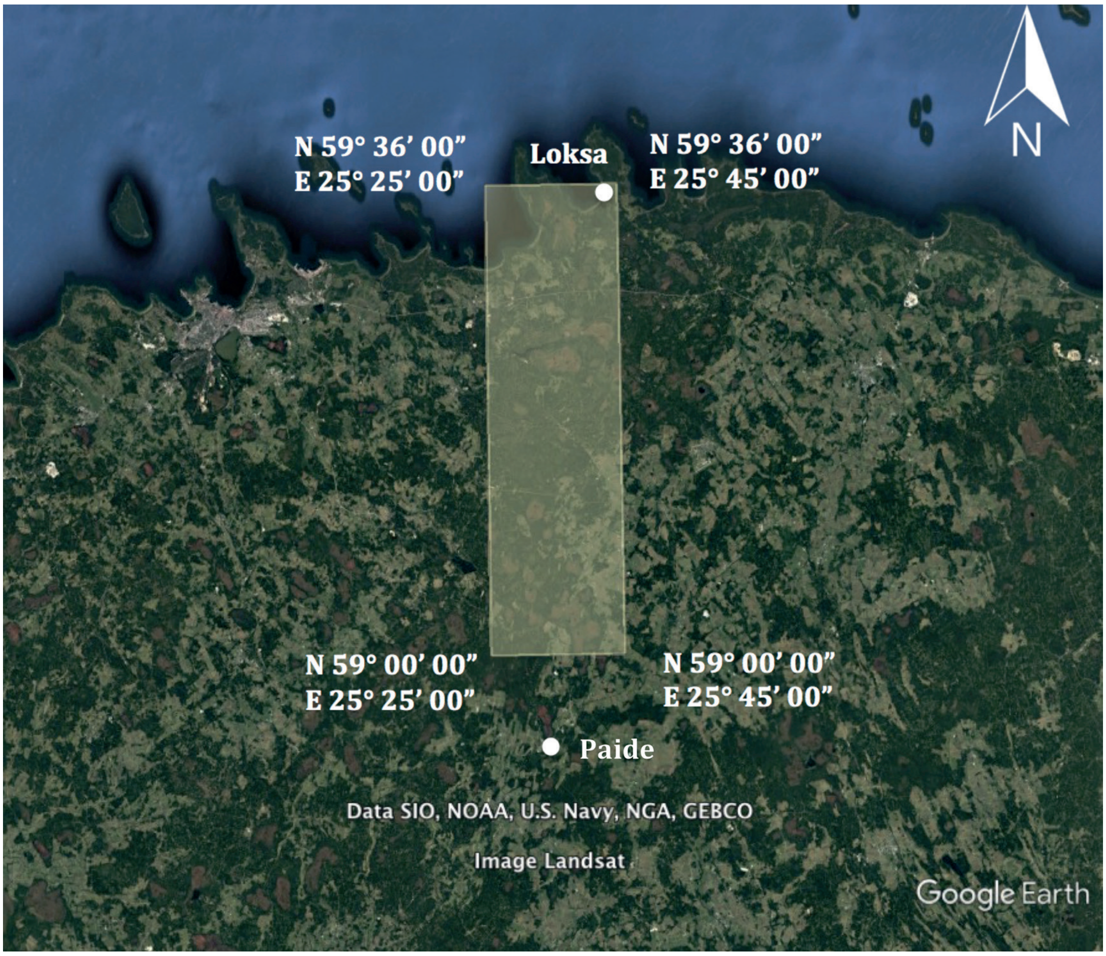

Fig. 8 Study area 4: forested area in Estonia.

Different varieties of wheat and barley exist, such as vernal or winter varieties, with their own production cycles. Cereals constitute the main product of the United Kingdom's crops (Defra 2006), with the most popular of them being the spring wheat which makes up 95% of the total production, opposed to the 5% of the winter wheat's production (Nabim 2014).

For this reason, the selected SAR image acquisition periods were between May and June, representing the backscatter from the products before the harvest. Conversely, the images taken between September and October correspond to the backscatter of the products after the harvest. The analysis focused on the behaviour of backscatter values over several years.

More specifically, from a total of 27 available SLC Envisat/ASAR images, 17 were retrieved during the period between November 2007 and November 2009. From the remaining ten images, five of them refer to the months of May and June for the period from 2004 to 2006, while the remaining five refer to the months of September and October of the same years.

All the images were calibrated and terrain corrected, while backscatter statistics were retrieved for each of the 17 images. The remaining two sets of 5 images were stacked, in order to calculate the average backscatter for each set.

2.4 Forested Areas

The study area is located between the towns of Kehra and Tapa in Estonia (Fig. 8). Although small in extent, Estonia has a large part of its area covered with forest – 2.3 million hectares – which is 51.5% of the mainland territory. The dominant tree species of the Estonian forests are Scots pine (covering 32% of the forested area), followed by birch species (31%), Norway spruce (19%), grey alder (8.5%) and aspen (5%) (Nordic Forest Research 2004).

One of the most important factors that have to be taken into account when studying a forested area with SAR is the wavelength of the available band. Depending on the wavelength used, the penetration of the SAR signal may vary, similarly to other vegetated or agricultural areas (Fig. 9). Thus the information retrieved may be misleading, if not properly interpreted.

For this land cover type nine sets of two images (C band and VV polarization) from the tandem mission of ERS-1 and ERS-2 were chosen and processed for interferometric analysis (coherence estimation and evaluation). From the total of nine available ERS tandem pairs originally processed, only four were eventually selected for further analysis.

From this set of four pairs, two were acquired around the summer period, while the other two were

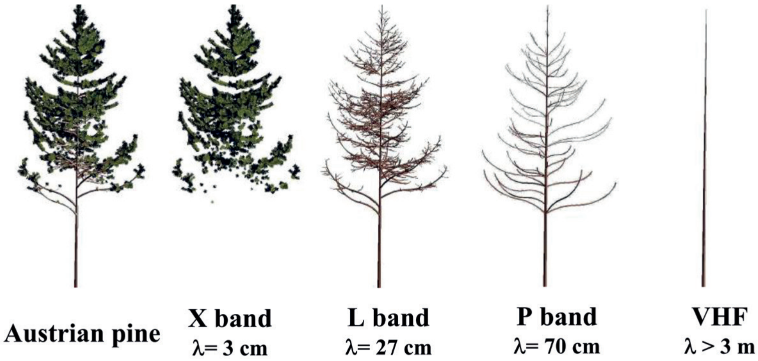

Fig. 9 Dependence of SAR backscatter of biomass (leaves, branches, trunks) from the radar wavelength (Le Toan 2007; Le Toan et al. 2001).

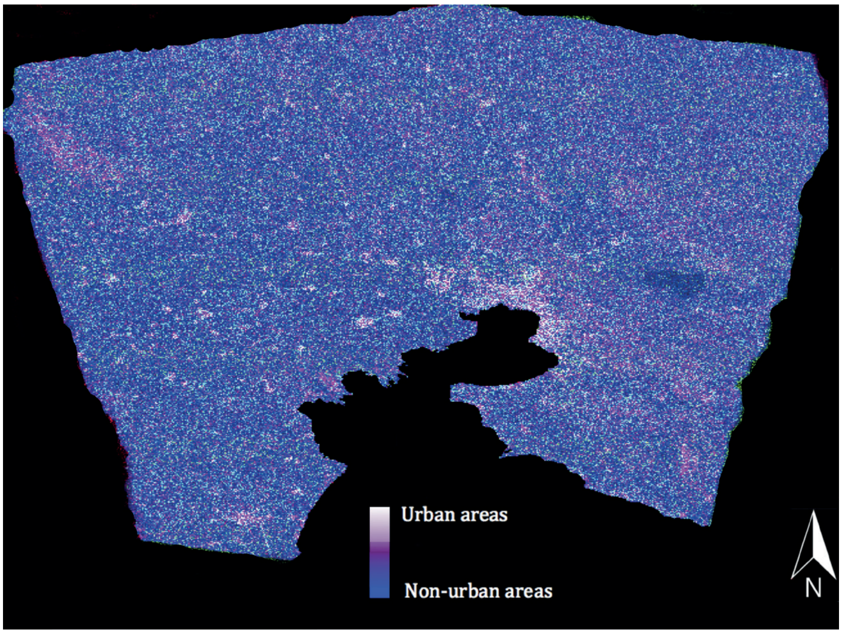

Fig. 10 RGB image of Thessaloniki area, produced by the average coherence of 4 ascending and 6 descending SAR images (Red: average of descending pass; Green: average of ascending pass; Blue: average of both passes).

acquired during winter. The fact that the images were spread during two semesters provided the opportunity to study the area during two extreme conditions. These conditions also represent the extremes in the trees' annual life cycle (especially for deciduous trees) – being devoid of foliage (winter) and full of foliage (summer).

3. Results and Discussion

3.1 Urban Areas

Concerning the urban versus non-urban areas, an attempt to classify them based on coherence, showed considerably higher values for the former and lower values for the latter, as represented in the final orthorectified RGB product (Fig. 10). The higher backscatter values of the urban areas are something to be expected, as there are many corner reflectors for the SAR signal within the built environment.

Furthermore, by having a pair of images or a series of images taken over a considerable period of time, one can further estimate the construction activity of an area.

3.2 Mountainous Areas

In the analysis of this study area it was highlighted how the backscatter variation is connected to each semester and the related weather conditions. In the first semester the mean backscatter value is the highest of the year, possibly due to the high volume of snow- or ice-covered terrain. On the other hand, in the second semester, most of the mountain area is snow-free, which results in lower backscatter values in the area, compared to the winter period.

Finally, in May images, the backscatter has the lowest values of the year. By studying the specific month, a "snowline" can be identified between the highest parts – where there is continuous snow cover – and the lowest parts of the basin – where there is snow-free ground as well (Andersen 1982). Bearing this in mind, experiments in three subsets of the average May image were carried out, in order to study the backscatter values at three respective altitude zones. The results provide some indication on how the backscatter decreases progressively, when moving from the highest part of the mountains to lower elevations, but further investigation and more data is needed, in order to come to meaningful conclusions.

3.3 Agricultural and Low Vegetation Areas

Regarding the agricultural and low vegetation areas, taking into account annual rainfall data, the interpretation of backscatter variation is based on soil moisture and roughness variations. When these two variables have low values (dry and smooth surface), the backscatter is also low. When either of the two parameters is characterized by high values (wet and/or rough surface), the resulting backscatter is high.

Thus the reason why the backscatter in winter is high can be attributed to the increased soil moisture. During winter months, the agricultural production is not fully-grown and most of the SAR signal returns directly from the ground, which is also wet, resulting in high backscatter values. It ought to be noted that this behavior is similar to that of trees; before the signal returns to the receiver, it can follow different pathways, which may thus result in different backscatter values.

In more detail, as it can be observed in Fig. 11, from March to September, the average image backscatter has the lowest values. From November to February, the values start to rise.

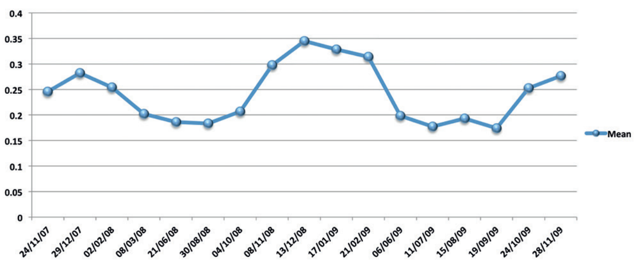

Fig. 11 Periodicity of the mean backscatter values for the agricultural and low vegetation areas.

Additionally, from the two sets of 5 images the typical backscatter values before and after the harvest for the agricultural areas were calculated; the most distinct differences can be observed in terms of mean values for the May/June (0.1428+/−0.0528) and September/October (0.2447+/−0.0766). These values and periodicity can be attributed to the backscatter variations of crops rather than urban areas, as shown in Fig. 12.

Taking into consideration the different stages of growth for each crop type (Fig. 13), as well as the average monthly precipitation in the study area (Fig. 14), further investigation could be performed. This ought

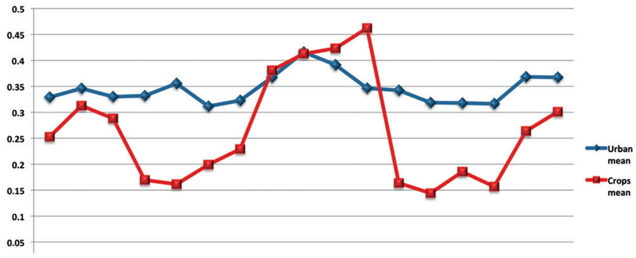

Fig. 12 Comparison of the mean backscatter values for crops and urban areas.

to focus on the backscatter variations, owing to the changes in the two main factors affecting the SAR signal, i.e. surface roughness and dielectric constant (directly related to the water content). Nevertheless, this would require extensive in-situ data and an analysis per crop type, as their cultivation periods differ significantly.

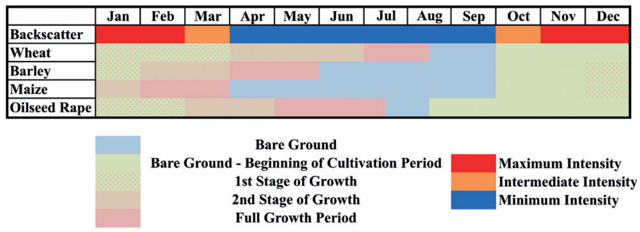

Fig. 13 Monthly backscatter with respect to the different stages of growth, for each crop type.

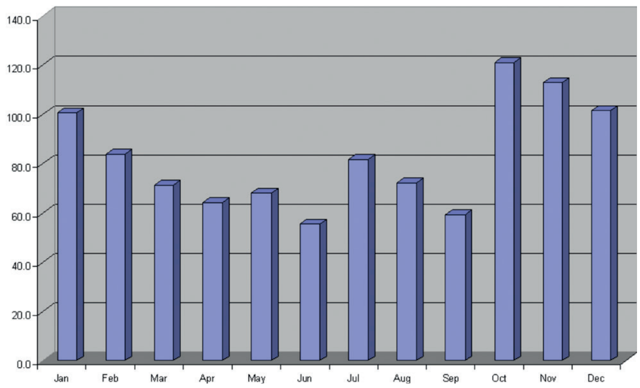

Fig. 14 Average monthly precipitation for study area 3 (agricultural and low vegetated area).

3.4 Forested Areas

Finally, for the forested areas, the coherence results for the two winter datasets, show a concentration of the pixel values between 0.4 and 0.7, with a maximum accumulation at 0.65 and a mean value of 0.51. On the other hand, for the summer image sets, the pixel values are concentrated between 0.2 and 0.35 with a maximum accumulation at 0.25 and a mean value of 0.35. Especially for August and February, greater data dispersion was identified, with many pixels having values up to 0.7 and down to 0.4 respectively. Furthermore, the pixels values from the two sets of coherence were averaged, in order to get the mean values for each pixel for each semester.

Subsequently, using the CORINE land cover database (scale of 1 : 100,000) and the processed coherence images, all the relevant classes and their coherence could be identified in a GIS (ArcGIS™) for each semester (Table 2).

As regards to the coherence for the forested areas (coniferous, mixed and broad-leaved), the results show that during summer the values are characteristically lower compared to the other classes. Furthermore, even though coherence increases during winter, it has still lower values for the aforementioned forested areas than in the other classes, but with a wider range. Between the different species, low coherence results throughout the year are encountered (from lowest to highest) over coniferous forests, followed by mixed, broad-leaved forest and finally transitional woodland shrub. For the different species, lower coherence was also associated with higher dispersion.

Tab. 2 Summer and winter coherence for the different classes.

Class	Summer Coherence	Winter Coherence
Inland marshes	0.48–0.64	0.57–0.73
Transitional woodland shrub	0.32–0.67	0.53–0.83
Pastures	0.22–0.51	0.44–0.77
Land occupied by agriculture	0.20–0.66	0.37–0.77
Mixed forest	0.12–0.45	0.22–0.70
Coniferous forest	0.09–0.66	0.14–0.78
Water bodies	0.16–0.49	0.53–0.74
Industrial or commercial units	0.25–0.46	0.32–0.58
Mineral extraction sites	0.45–0.65	0.55–0.79
Discontinuous urban fabric	0.21–0.55	0.35–0.71
Non-irrigated arable land	0.37–0.68	0.55–0.81
Complex cultivation patterns	0.36–0.70	0.45–0.77
Broad-leaved forest	0.20–0.57	0.34–0.74
Natural grassland	0.30–0.50	0.40–0.78
Peat bogs	0.34–0.74	0.44–0.82
Moors and heathland	0.22–0.53	0.49–0.73
Road and rail networks	0.44–0.54	0.58–0.74

4. Conclusions

In this paper, different approaches for land cover monitoring through classification with the use of SAR data were tested over Europe. In particular, ERS and Envisat/ASAR datasets were employed in order to identify urban, mountainous, agricultural/low vegetation areas and forests, by exploiting SAR backscatter and coherence. The period of SAR observations included data from all four seasons, thus covering a variety of meteorological conditions, as well as agricultural/vegetation stages.

For the urban and non-urban areas an attempt to classify them based on coherence, showed great value differences between the two aforementioned categories. For the selected location, images throughout the year were used and no deterioration of the results due to moisture or rain was observed. On these grounds, it can be assumed that this methodology is robust enough to be used for studies related to urban sprawl monitoring.

On the other hand, in the case of the mountainous areas at high latitudes under study, it has proven difficult to separate the mountainous snow-free terrain from the snow-covered, without weather information. The results for winter/summer show little difference between the mean, maximum and minimum backscatter. The reason is that in the first semester (winter) the mountains are full of snow, while in the second semester (summer) the snow and ice are melted and stagnated in the area. As a result, a relatively high backscatter is observed during both semesters. Nevertheless, during May, which is the driest month of the year in the study area, backscatter values are significantly lower. Consequently, during this period, it may be relatively easy to use backscatter values for snow line mapping, if considerably more satellite images are employed.

Regarding the agricultural and low vegetation areas, the interpretation of results ought to be performed in conjunction with the annual rainfall in the UK and by considering soil moisture and roughness. Especially with respect to the relative weight (importance) of these two last parameters, which are critical for SAR backscatter, this case study from the area in the UK has provided with some interesting insights.

More specifically, in this case study, winter backscatter values of agricultural areas were high, while the respective summer values were relatively low.

This information should be combined with the fact that during winter, surface roughness is low (theoretically low backscatter), as the agricultural areas have been harvested and seeded for the next growth period, while soil moisture is high (theoretically high backscatter).

On the contrary, during summer, the surface is rather rough, as the agricultural production is fully grown (theoretically high backscatter), while soil moisture is relatively low (theoretically low backscatter).

Therefore, it seems that for agricultural and low vegetation areas, soil moisture is a more important factor than roughness and it more or less determines the strength of SAR backscatter.

In any case, working with this type of terrain requires more in-situ information, also taking into account daily meteorological data and the agricultural cycle of different crops.

Finally, for the forested areas, a methodology based on Corine Land Cover (CLC) maps has been adopted; in order to classify all the areas based on backscatter and coherence values and to discover the differences between the various forest types. Common ground for all of these kinds of forest is the low coherence and the high variability, all over the year, in relation to all the other types of land cover. Therefore, it is relatively easy to discriminate forest from non-forested areas. On the other hand, the coherence values of the different types of forest are very similar (between 0.09–0.67 for the summer and 0.14–0.83 for the winter). Thus, in order to discriminate between different types of forest, in-situ information and/or different kind of SAR data (more polarizations) is needed.

In conclusion, the heritage and experience gained from ERS and Envisat/ASAR SAR imaging for land cover classification approaches shall be invaluable for the exploitation of current and future SAR data. In the European context in particular, the Sentinel-1 satellites (since 2014) within the realization of the Copernicus Programme, guarantee not only the continuity of European C-band SAR satellite missions, but also bring along significant improvements. These are mainly in terms of spatial and temporal resolution, as well as systematic acquisitions in at least two polarizing modes, resulting in an overall improved capability of monitoring land surface and discriminating between different land cover types.

Acknowledgements

This study could not have been completed without the various SAR data from ENVISAT and ERS-1/2, for which the European Space Agency (ESA) provided access through EOLiSA, in the framework of a category-1 project (ID: 28350). The first author would also like to thank Dr. Stephen Hobbs and Dr. Konstantinos Salonitis, for giving him the opportunity to realize this study as part of his thesis at the Department of Aeronautics and Space Engineering of Cranfield University, but also for taking part in the project GEoSTARe.

References

Ackermann, R. E. (2011): Sequential land cover classification (unpublished MSc thesis), University of Pretoria, South Africa.

Andersen, T. (1982): Operational snow mapping by satellites. Hydrological Aspects of Alpine and High Mountain Areas. Proceedings of the Exeter Symposium, July 1982, IAHS Publication 138, 149–154.

Barbieri, M., Lichtenegger, J. (2005): Introduction to SAR for Geology. In K. Fletcher (ed), Spaceborne radar applications in Geology, ESA TM-17: 1–54, Noordwijk, European Space Agency.

BBC (2007): Snapshot of farming in the UK. http://news.bbc.co.uk/2/hi/uk_news/magazine/6919829.stm

DEFRA (2006): Arable crops grown in the UK. http://archive.defra.gov.uk/foodfarm/growing/crops/

DEFRA (2013): Farming statistics provisional crop areas, yields and livestock populations at June 2013. United Kingdom, National Statistics Publications. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/364157/structure-jun2013prov-UK-16oct14.pdf

Geology.com (2016): Physical map of Europe. http://geology.com/world/europe-physical-map.shtml

Haack, N. B., Herold, D. N., Bechdol, A. M. (2000): Radar and optical data integration for land-use/land-cover mapping. Photogrammetric Engineering and Remote Sensing 66(6), 709–716. http://asprs.org/a/publications/pers/2000journal/june/2000_jun_709-716.pdf

Hale, S. R., Rock, B. N., (2003): Impacts of topographic normalization on land-cover classification accuracy. Photogrammetric Engineering and Remote Sensing 69(7), 785–791, https://doi.org/10.14358/PERS.69.7.785.

Hanssen, R. F. (2001): Radar Interferometry: Data Interpretation and Error Analysis, Springer, Berlin, https://doi.org/10.1007/0-306-47633-9.

Hellenic Statistical Authority (2011): Statistical Yearbook of Greece, 2009 and 2010. Piraeus, Greece.

Kaoru, T., Toshiharu, K. (1996): GIS-aided land cover classification assessment based on remote sensing images with different spatial resolutions. Application of Geographic Information Systems in Hydrology and Water Resources Management. Proceedings of the Vienna Conference HydroGIS '96, IAHS Publication 235, 659–665.

Lavalle, M. (2013): Tackling temporal decorrelation in repeat-pass polarimetric interferometry. 2nd Advanced Course on Radar Polarimetry, 21–25 January 2013, ESA-ESRIN. Frascati (Rome), Italy.

Lavalle, M., Simard, M., Hensley, S. (2012): A Temporal Decorrelation Model for Polarimetric Radar Interferometers. IEEE Transaction on Geoscience and Remote Sensing, 50(72012), 2880–2888, https://doi.org/10.1109/TGRS.2011.2174367.

Le Toan, T. (2007): SAR data for forestry and agriculture. Advances Training Course on Land Remote Sensing, 5 September 2007, Lecture D3L3. http://earth.esa.int/landtraining07/D3L3-LeToan.pdf

Le Toan, T., Picard, G., Martinez, J. M., Melon, P., Davidson, M. (2001): On the relationships between radar measurements and forest structure and biomass. Proceedings of the Third International Symposium on Retrieval of Bio- and Geophysical Parameters from SAR Data for Land Applications, 11–14 September 2001, Sheffield, UK. ESA Publications Division (ESA SP-475), Noordwijk, Netherlands, 3–12.

Long, G. D. (2008): Microwave Sensors – Active and Passive. Provo, UT.

Lu, D., Weng, Q. (2007): A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing 28(5), 823–870, https://doi.org/10.1080/01431160600746456.

Lu, Z., MEYER, D. J. (2001): Study of high SAR backscattering caused by an increase of soil moisture over a sparsely vegetated area: implications for characteristics of backscattering. International Journal of Remote Sensing 23(6), 1063–1074, https://doi.org/10.1080/01431160110040035.

Lusch, P. D. (1999): Introduction to Microwave Remote Sensing. Center for Remote Sensing and Geographic Information Science. Michigan State Univesity, BSRSI, November 1999, Michigan, 1-84. http://web.pdx.edu/~emch/ip2/RADAR.pdf
Malnes, E., Guneriussen, T. (2002): Mapping of snow covered area with Radarsat in Norway. Procceedings IGARSS'02, 28 June 2002, Vol. 1, 683–685, Toronto, Canada, https://doi.org/10.1109/IGARSS.2002.1025145.
Martinez, C. L., Pottier, E. (2005): Topography Independent InSAR Coherence Estimation in a Multiresolution Scheme. Proceedings of IEEE International Geoscience and Remote Sensing Symposium, IGARSS '05,25-29 July 2005, Vol. 4, 2689–2692, Seoul, Korea, https://doi.org/10.1109/IGARSS.2005.1525620.
Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R., Simpson, C. I. (2011): Final Report for LCM 2007 – the new UK Land Cover Map. CS Technical Report No 11/07, Countryside Survey, July 2011. http://www.countrysidesurvey.org.uk
NABIM (2014): Wheat Production. http://www.nabim.org.uk/wheat/wheat-production
Nordic Forest Research (2004): Forest Research in the North: Estonia. Scandinavian Journal of Forest Research (News & Views). http://www.nordicforestresearch.org/wp-content/uploads/2011/03/NV_2004_1.pdf
Rhode Hill Gardens (2018). Rainfall at Rhode Hill Gardens. http://homepages.nildram.co.uk/~rhg/rain.html#ad
Santoro, M., Askne, J. H., Wegmuller, U., Werner, C. L. (2007): Observations, Modeling, and Applications of Ers-Envisat Coherence Over Land and Surfaces. Geoscience and Remote Sensing, IEEE Transactions 45(8), 2600–2611, https://doi.org/10.1109/TGRS.2007.897420.
Siqueira, P., Hensley, S., Chapman, B., Bergen, K.: Temporal decorrelation studies relevant for a vegetation InSAR mission, NASA, (unpublished poster). http://cce.nasa.gov/mtg06_ab_presentations/163_138_ab_pres.pdf
Sokal, R. (1974): Classification: purposes, principles, progress, prospects. Journal Science 185(4157), 1115–1123, https://doi.org/10.1126/science.185.4157.1115.
Solberg, R. (1993): Mapping Snow Cover by ERS-I SAR. Proceedings of IGARSS '93 Symposium, Vol. 3, 1273–1275, Tokyo, Japan. https://doi.org/10.1109/IGARSS.1993.322100.
Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., Macomber, S. A. (2001): Classification and change detection using Landsat TM data: when and how to correct atmospheric effect. Remote Sensing of Environment 75(2), 230–244, https://doi.org/10.1016/S0034-4257(00)00169-3.
Storvold, R., Malnes, E., Larsen, Y., Høgda, A. K., Hamran, E. S., Muller, K., Langley, A. K. (2006): SAR Remote Sensing of snow parameters in Norwegian areas – current status and future perspective. Journal of Electromagnetic Waves and Applications 20(13), 1751–1759. https://doi.org/10.1163/156939306779292192.
Toomay, J. C., Hannen, J. P. (2004): Radar principles for the Non-specialist. 3rd Edition, Scitech Publishing, New York, https://doi.org/10.1049/SBRA032E.
UK Agriculture (2014): Crops in the UK. http://www.ukagriculture.com/crops/crops.cfm
Wei, M., Sandwell, T. D. (2010): Decorrelation of L-Band and C-Band Interferometry Over Vegetated Areas in California. IEEE Transaction on Geoscience and Remote Sensing 48(7), 2942–2952, https://doi.org/10.1109/TGRS.2010.2043442.
World Weather & Climate Information (2014): Average weather in Bergen, Norway. http://www.weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine,bergen,Norway
Zebker, A. H., Villasenor J. (1992): Decorrelation in Interferometric Radar Echoes. IEEE Transaction on Geoscience and Remote Sensing 30(5), 950–959, https://doi.org/10.1109/36.175330.

Rainfall analysis of debris flows in the Obří důl Valley in the Krkonoše Mts., Czechia

Zuzana Drábová*

Charles University in Prague, Faculty of Science, Department of Physical Geography and Geoecology, Czechia
* Corresponding author: zuzana.drabova@natur.cuni.cz

ABSTRACT
Rainfall is the most important triggering factor responsible for the occurrence of debris flows in the Obří důl Valley in the Krkonoše Mts. The critical rainfall conditions for slope failures are not the same for different debris flows, and may be strongly influenced by regional geological and geomorphological conditions. Nevertheless, analysis of hourly intensities, daily rainfall, cumulative data and the antecedent precipitation index (API) revealed that several of the above-mentioned factors are necessary to trigger the debris flow. On the other hand, a significant amount of daily rainfall (e.g. 225 mm) could trigger a debris flow without the support of any other rainfall characteristics in the monitored area and period under review. We used several rain gauges from the study area but the local differences in rainfall were so high that data from more remote stations was difficult to include in the Obří důl Valley. This is why only a limited amount of precise data is available for some years.

KEYWORDS
debris flow; rainfall analysis; Krkonoše Mts.

Received: 22 June 2018
Accepted: 7 November 2018
Published online: 17 December 2018

Drábová, Z. (2018): Rainfall analysis of debris flows in the Obří důl Valley in the Krkonoše Mts., Czechia.
AUC Geographica 53(2), 220–237
https://doi.org/10.14712/23361980.2018.21

1. Introduction

The main aim of the paper was to analyse rainfall data from several rain gauges in order to determine the rainfall threshold for triggering debris flows in the Obří důl Valley in the Krkonoše Mts. They have morphological as well as precipitation preconditions for debris flow occurrence. Due to this fact debris flows were described in the Krkonoše National Park in the frame of inventory of geomorphological features since early 1970s (Pilous 1973). This type of gravitational movement has been described in detail and analysed with respect to the genesis by Pilous (1973, 1975 and 1977). The Obří důl Valley has the largest occurrence of debris flows in the Krkonoše Mts. It is a well-developed glacial valley between Sněžka Mt. (the highest mountain in this range) and Pec pod Sněžkou Mt. The position of the surrounding mountains is clear from Fig. 1. Glacial modelling of headwalls and nivation hollows in the Obří důl Valley was analysed by Šebesta and Treml (1976), who studied debris flows in landscape development.

Debris flow paths in the Krkonoše Mts. are also connected with the occurrence of snow avalanches, which have been observed since the winter of 1961/62 according to the International avalanche classification (De Quervain et al. 1981). This classification was adopted for the Krkonoše Mts. by Spusta and Kociánová (1998), Spusta et al. (2003), Spusta et al. (2006), Kociánová and Spusta (2000), Kociánová et al. (2004), Vrba and Spusta (1975, 1991). The area of the Krkonoše Mts. has also been analysed from the point of view of its susceptibility to snow avalanches (Blahůt 2008; Suk and Klimánek 2011; Juras et al. 2013).

Rainfall is the most frequent triggering factor for shallow slope deformations (Záruba and Mencl 1982; De Vita and Reichenbach 1998; Schuster and Wieczorek 2002; Glade and Crozier 2005). These phenomena usually occur in places where surface and subsurface runoff is concentrated and where a sufficient amount of loose material is located. The most significant debris flows in the Obří důl Valley occurred during extreme rainfall events of 1882 and 1897, when

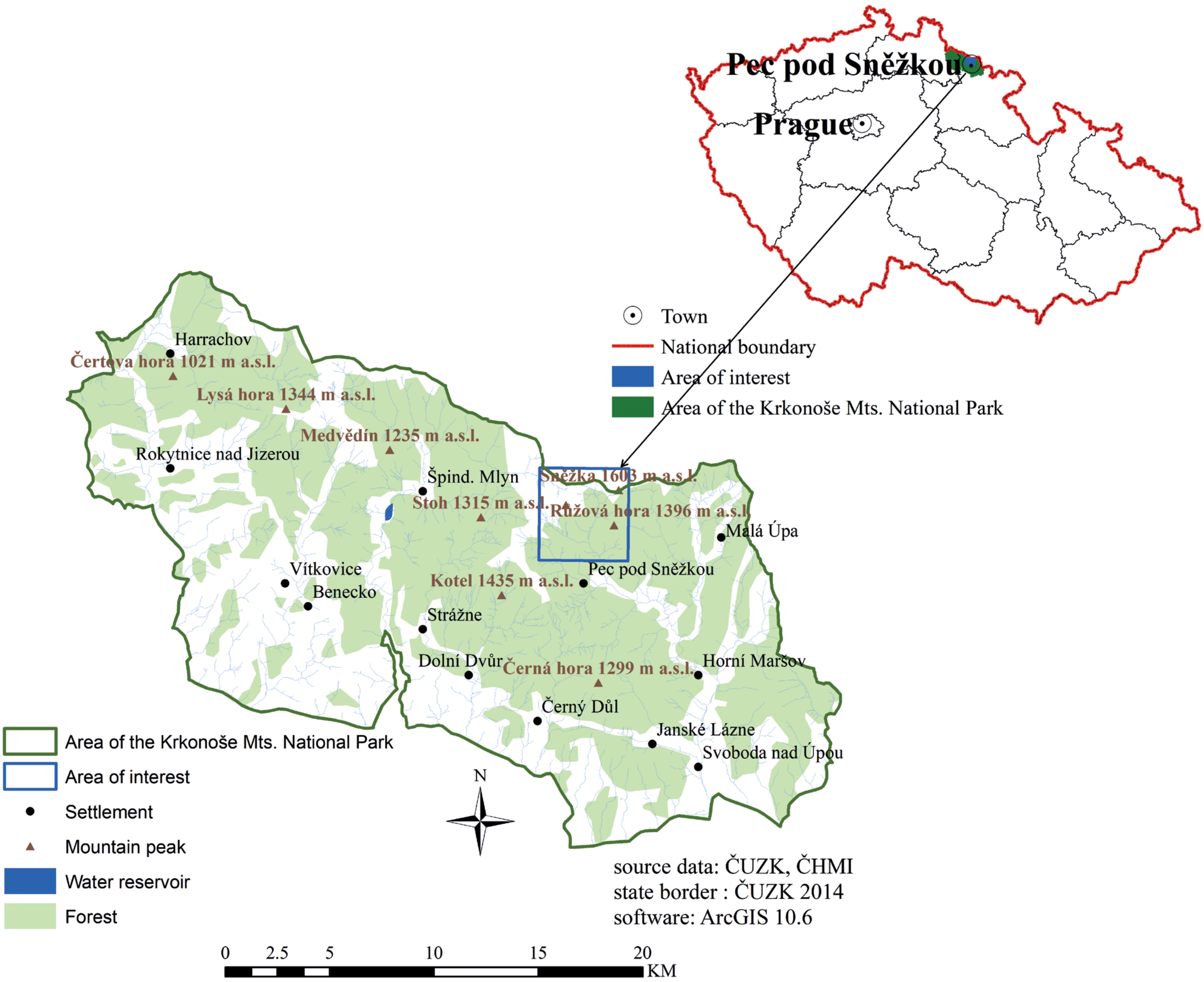

Fig. 1 The study area.

two houses were destroyed and seven people died (Pilous 2016).

In order to analyse landslides generated by rainfall it is important to establish the rainfall threshold, which is often connected with the relation between the duration and intensity of the rainfall (e.g. Turner and Schuster 1996; Novotný 2000; Schuster and Wieczorek 2002). Even rather low rainfall can trigger shallow landslides in the recently deforested areas or arid regions (Schuster and Wieczorek 2002). According to Turner and Schuster (1996) and Guzzeti et al. (2004). It is possible to establish rainfall thresholds for shallow landslides (up to 8 meters) on the basis of the optimal rate between the duration and intensity of the rainfall. Rybář and Novotný (2005) stressed the importance of seasonal and multiannual cycles in rainfall and temperature curves for landslide activity. There are two basic methods to define rainfall thresholds – physical and empirical based models. The physical model is based on the process of the landslide origin and conceptual approach comes out from historical or statistic data (Corominas 2000; Crosta and Frattini 2001; Aleotti 2004; Wieczorek and Glade 2005; Guzzetti et al. 2007). The thresholds can be considered as being global, regional or local (Guzzetti et al. 2004, 2007).

1.1 Antecedent precipitation index (API)

Antecedent rainfall is believed to play an important role in the initiation of debris flows because it reduces soil suction and increases the pore-water pressure (Thach et al. 2002). The API shows the rainfall situation retrospectively and is used to define the antecedent moisture condition (Mishra and Singh 2003). The API is also used to assess the saturation of the watershed. Soil moisture has a considerable influence on the physical properties of soil, e.g., pore-water pressure and shear strength (Zhao et al. 2011), which can affect the initiation of debris flows as discussed by Brand (1989), Marchi et al. (2002), and Wieczorek and Glade (2005). In addition, Crozier and Eyles (1980) consider antecedent climatic conditions to be crucial for the triggering of debris flows. The influence of antecedent rainfall is determined by seasonal variations in rainfall and temperature, which affect evapotranspiration. Intense convective storms occur during the summer when evapotranspiration can quickly remove much of the soil moisture. Consequently, the significance of antecedent rainfall may vary depending upon the regional climate (Wieczorek and Glade 2005). The average level of moisture in a catchment varies daily. It is replenished by rainfall and subsequently depleted by evaporation and evapotranspiration (Mishra and Singh 2003).

1.2 Determination of rainfall thresholds

A rainfall threshold is defined as the minimum rainfall conditions for triggering landslides in a particular region (Guzzetti et al. 2007). The determination of rainfall thresholds for landslide initiation is considered as a basic task in landslide hazard assessment, and various methods have been proposed to establish rainfall thresholds (Crosta 1998; Corominas and Moya 1999; Glade 2000; D'Odorico and Fagherazzi 2003; Zezere et al. 2005; Godt et al. 2006; Guzzetti et al. 2007; Marques et al. 2008; Dahal and Hasegawa 2008; Dahal et al. 2009; Frattini et al. 2009; Saito et al. 2010; Giannecchini et al. 2012).

One of the most difficult tasks in using antecedent rainfall for debris flow prediction is determining the number of days to be used (Guzzetti et al. 2007). A detailed literature review revealed a complex relationship between the number of days of the antecedent rainfall and the triggering of a landslide. Terlien (1998) considered 2, 5, 15 and 25 days for the Manizales area (Colombia). Kim et al. (1991) used 3 days, Heyerdahl et al. (2003) used 4 days, Glade (2000) used 10 days, Aleotti (2004) considered 7, 10, and 15 days, Zezere et al. (2005) used 1, 5, 10, 15, 30, 45, 60, 75 and 90 days., and Polemio and Sdao (1999) considered 180-day cumulative daily rainfall data. In summary, antecedent rainfall of between 3 and 120 days (Pasuto and Silvano 1998) can be used to explain the occurrence of landslides (Dahal et al. 2009). The large variability in the number of antecedent rainfall days may be influenced by factors such as: diverse lithological, morphological, vegetation, and soil conditions; different climatic regimes and meteorological circumstances leading to slope instability; and heterogeneity and incompleteness in the rainfall and landslide data used to determine the thresholds (Guzzetti et al. 2007).

2. Geological and geomorphological settings

The lithology of the area is not uniform. The northern slopes of the Obří důl Valley are composed of Krkonoše crystalline complex from the early Proterozoic – several hundred-meter-thick successions of beds of grey mica schist are interbedded with quartz, erlan and gneiss (Chaloupský and Teisseyre 1968). The weathered mantle on the slopes also varies in thickness. In the area of the rock outcrops, the mantle is limited only on local depressions, while in other areas it fluctuates between 20 cm and 2 meters. Talus deposits under the rock walls or talus cones on the foothills represent an even greater thickness (Pilous 1973). Boulders and stones prevail over gravel and sand; nevertheless, smaller particles (up to clayey) are also included. The largest boulders are around 2 meters in diameter but can vary from 20 to 100 cm (Pilous 1973). Debris flows have occurred in sediments with various grain-sizes, e.g. in stony-debris (the Nad Kovárnou site) as well as in areas where sandy fractions prevail (the Rudník site). Rankers have evolved on the top of the silicate rocks (Horník et al. 1986) at different sites, e.g. on crests, slopes or deluvial deposits, and most are covered with forest vegetation (Hraško et al. 1991).

The processes of erosion and denudation has gradually transformed the original Palaeozoic Variscan mountain range into a planation surface (i.e. slightly protruding highlands with softly undulated landforms) that reflect the structural conditions of the bedrock (Pilous 2016). According to the latest research based on the thermochronological analysis of samples from the Krkonoše Mts. (Danišík et al. 2010) erosion was responsible for the removal of between 3.6 and 6 km of rock between 100 and 75 Ma.

Research of neotectonic processes and the resulting landforms has long been focused on the Polish side of the Krkonoše Mts. (Migoń 1992). Its results support older opinions (Ouvrier 1933) that the Krkonoše Mts. represent a segmented horst. In the case of the Czech part of the Krkonoše Mts., tectonic processes have only been mentioned in relation to the landforms as a general background or in a very few case studies (Migoń and Pilous 2007). New research shows that their influence is probably much greater than first thought (Lysenko 2007). Hard rocks of the contact zone have also had an impact on the appearance of the landforms. Their extremely steep slopes create a typical hogback shape in the area of Čertův hřeben Ridge (Pilous 2016).

Glaciation of the Krkonoše Mts. is the feature most commonly studied by geomorphologists. Partsch (1894) and Migoń (1999) drew attention to the important role of the periglacial landscape, and particularly the extent and compactness of deflation zones on summit planation surfaces in the Sudetes mountain range. The latest and most complex findings on the extent of glaciation within the Krkonoše Mts. have been presented by Engel (1997, 2003, 2007). Glacial features dominate the relief of the central part of the Krkonoše Mts., where cirques and troughs are deeply incised into summit plateaus (Engel et al. 2014).

Contemporary research mainly involves cross-sectional and longitudinal profiling, radiometric dating and macroscopic analysis of glacial sediments, which, among other things, are performed to contribute to the correlation of moraines on both sides of the mountain range (Engel et al. 2010, 2014). The latest findings of Engel et al. (2014) show that only small cirque glaciers occurred during the last glaciation period.

The periglacial landforms in the Krkonoše Mts. were described by Králík and Sekyra (1969). These landforms and their genesis particularly in the last two decades were also studied by Sekyra and Sekyra (1995), Křížek (2007), Treml et al. (2010), Křížek and Uxa (2013). The most conspicuous and largest landforms are cryoplanation terraces occurring at various extents, lengths and perfections on almost every peak in the Krkonoše Mts. (Pilous 2016). The landforms on Studniční hora Mt. and Sněžka Mt. jutting above the alpine timberline are perfectly developed (Dvořák et al. 2004). Nivation forms can be found in the Obří důl Valley (Pilous 2016).

The analysis of factors determining the occurrence of debris flow can be divided into: geological, geomorphological, climatic and vegetational. The occurrence of debris flow is probably more connected to the ability of layers to contain water rather than the differences in grain-size distribution (Pilous 1973). The presence of mica in rocks supports the sliding process especially in the zone of saturation. The water from the flat area of the Krkonoše Mts. (levelled surface) infiltrates into the weathering mantle and through the system of joints and tectonic fractures into the massif and supports the saturation of loose material on the slopes of deeply incised valleys like the Obří důl Valley (Czerwiński 1967).

Geomorphological factors include: inclination and aspect of the slope, relative height, possible presence of snow avalanches and windward (or leeward) effect. The average slope inclination of the debris flow scarp and transitional area varies between 24° and 46°; however, at Čertova zahrádka and Čertova Gorge it can be between 33° and 90°. The vertical differentiation between the source and accumulation area is between 925 and 1460 m, while the length is from 30 to 680 m (Pilous 1977). The slope aspects with relation to prevailing winds were analysed by Sokol and Vavřík (1971) who revealed that the west oriented slopes (NW–W–SW) are the most affected in this area. A total of 83% of debris flows are fixed to these slopes in the Obří důl Valley (Pilous 1973). In terms of time distribution, most of the debris flows occur in June when the weathered mantle and soil are still well saturated from melted snow and storms can happen (Sokol and Vavřík 1971). Most of the debris flow scarp areas are below the timber line (Pilous 1977) where the vegetation is shallowly rooted (spruce), and trees can work during strong winds as crowbars.

3. Historical overview of debris flows

Debris flows in the Krkonoše Mts. are the largest and oldest, according to literature descriptions and documentation. The oldest engraving is from 1804 in Rudník, where one debris flow path can be documented. Other engravings and lithography are from the mid-19th century (Tittel 1830; Knippel 1850 and Tauber around 1850) and the locations are again Rudník and additionally Sněžka Mt.

In the first stage of the inventory, 35 debris flow paths in the Obří důl Valley were identified (Pilous 1973). The next inventory from 1977, which was based on the analysis of photos, field inspections, monitoring and more accurate data from the first inventory, revealed 51 events (Pilous 1973, 1977). In one particular case, more accurate dating appeared, i.e. for 25 August 1938 (Šourek 1977).

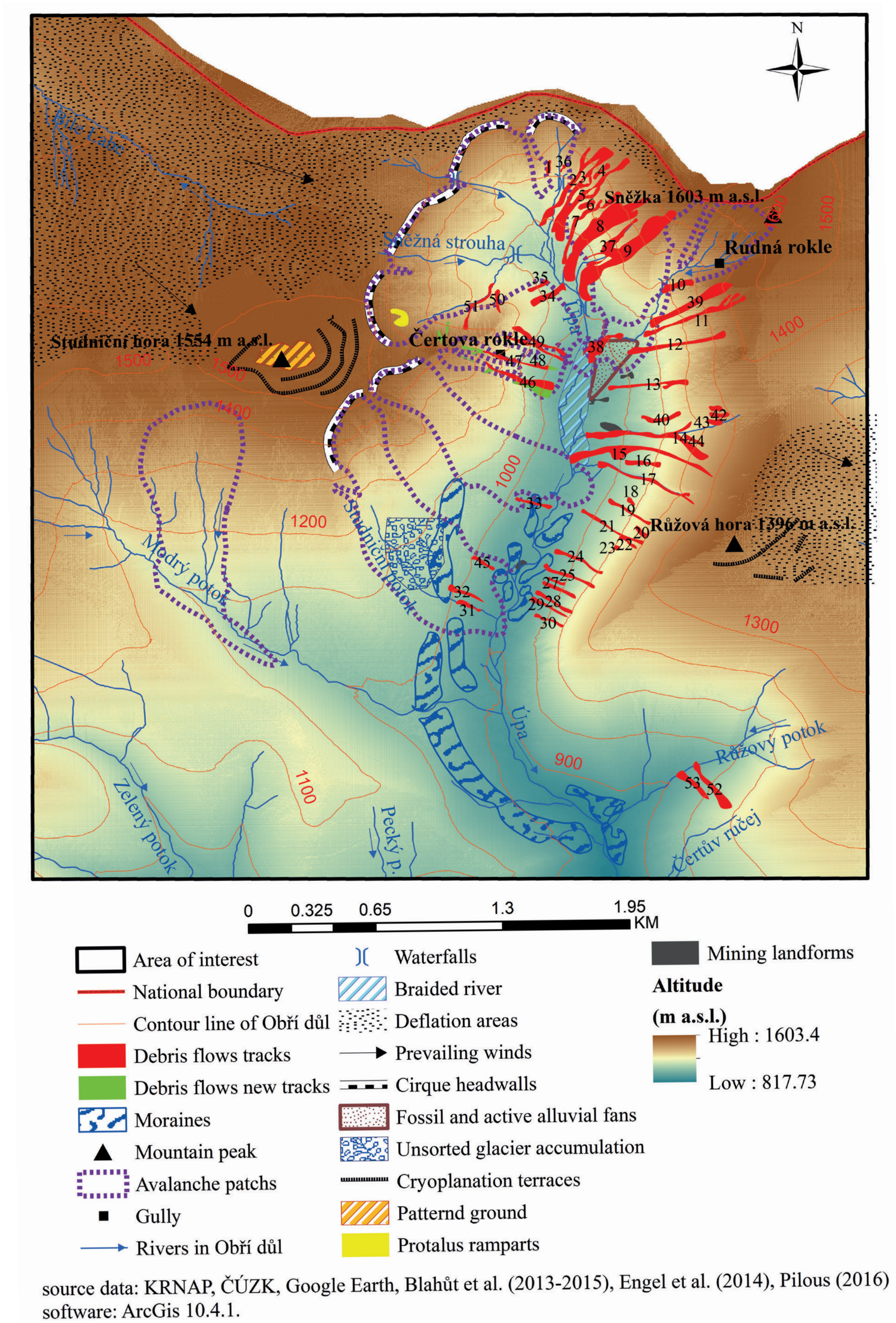

Fig. 2 Geomorphological map.

Tab. 1 List of debris flows (elaborated according to Pilous 1973, 1977). Colour: yellow and red – first stage of the inventory (Pilous 1973); only red – two houses destroyed and seven people killed (29/7/1897); green – next stage of the inventory (Pilous 1977); blue – the newest origin of debris flows in the Růžová důl Valley (Myslil 1997).

No.	Location	Date of origin	No.	Location	Date of origin
1	Úpská jáma Cirque	14/7/1964 (the youngest)	27	SE slope Růžová hora Mt/	(2–3/7/1926)
2	Úpská jáma Cirque	?	28	SE slope Růžová hora Mt.	(2–3/7/1926)
3	Úpská jáma Cirque	?	29	SE slope Růžová hora Mt.	(2–3/7/1926)
4	Úpská jáma Cirque	> 1882	30	Modrý důl Valley	?
5	Úpská jáma Cirque	> 1882	31	Výsluní Modrý důl Valley	?
6	Úpská jáma Cirque	> 1882	32	Výsluní Modrý důl Valley	?
7	Úpská jáma Cirque	> 1882	33	Velká studničná jáma Cirque	?
8	Úpská jáma Cirque	> 1882; 17/7/1882	34	Čertův hřeben Ridge	17/7/1882
9	Zadní Rudník	> 1882 17/7/1882	35	Čertův hřeben Ridge	17/7/1882
10	Rudná rokle	> 1882 17/7/1882	36	Úpička drainage through	?
11	Rudná rokle	> 1882 17/7/1882 (the oldest)	37	Zadní Rudník	?
12	Rudná rokle	(> 1882) 17/7/1882 29/7/1897	38	Rudník	?
13	Pod Kovárnou	29/7/1897	39	SW slope Sněžka Mt. /Rudná rokle	?
14	Růžová hora Mt.	29/7/1897	40	SW slope Sněžka Mt.	?
15	Růžová hora Mt.	29/7/1897	41	Růžová hora Mt.	?
16	Růžová hora Mt.	(29/7/1897)	42	Růžová hora Mt.	?
17	Růžová hora Mt.	(29/7/1897)	43	Růžová hora Mt.	?
18	Růžová hora Mt.	(29/7/1897)	44	Růžová hora Mt.	?
19	Růžová hora Mt.	(29/7/1897)	45	Výsluní Modrý důl Valley	12/7/1937
20	Růžová hora Mt.	(29/7/1897)	46	Čertova rokle	18/6/1974
21	Růžová hora Mt.	(29/7/1897)	47	Čertova rokle	22/6/1975
22	Růžová hora Mt.	(29/7/1897)	48	Čertova zahrádka	18/6/1974
23	SE slope Růžová hora	?	49	Čertova zahrádka	18/6/1974
24	SE slope Růžová hora	?	50	Čertův hřeben Ridge	18/6/1974
25	SE slope Růžová hora	?	51	Čertův hřeben Ridge	18/6/1974
26	SE slope Růžová hora	?	52+53	V Korytech Růžový důl Valley	7/1997

From the debris flow paths identified in the study area the oldest in Tab. 1 are numbers 4, 5, 6, 7, 8, 9, 11 and partially 12. The largest number of events occurred during intensive rainy periods (17/7/1882, 29-30/7/1897, 2-3/7/1926) and affected the whole of the Krkonoše Mts. Two of the above-mentioned events from Růžová hora Mt. (No. 14 and 15) had a direct catastrophic influence on a small settlement at the bottom of the Obří důl Valley, where two houses were completely destroyed and the debris flow left behind seven fatalities. Unfortunately, historical sources do not always provide the precise locations, like after the 1926 rainfall period. Ouvrier (1933) mentioned only that debris flows were identified in all of the larger valleys in the Krkonoše Mts.

Ouvrier (1933) described the debris flows that occurred during the night of 2 July 1926, in the Krkonoše Mts., but without any precise location. If these debris flows originated in the Obří důl Valley, then they were probably in the forested area of the Růžová hora Mt.

4. Data and methods

4.1 Rainfall data

The rainfall data were taken from fourteen rain gauges located near the Obří důl Valley (Fig. 3). The precise amount of rainfall is usually not known in the head scarp of the slope deformation because there are no rain gauges directly in that area. Therefore, we had to use the nearest stations and because of the rainfall variability due to possible orographic effects we also tried to consider the position of the rain gauge compared to the detachment area (detailed data are included in Tab. 2). Eleven rain gauges operated by the Czech Hydrometeorological Institute (CHMI) are located within a distance of 0.5 to 23 km (two of them are located between 0.5 and 5 km) from the triggering area of the debris flows (Tab. 2). Data from the National Center for Environmental Information – National Oceanic and Atmospheric Administration, National Climatic Data Center, U.S. Department of Commerce (NCDC NOAA) are available for stations at Sněžka Mt. and Pec pod Sněžkou for 6 h, 12 h, and daily amounts (see Tab. 2).

Rain gauges are located in a similar climatic region – very cold and cold with abundant precipitation (Quitt 1971) with the different geological setting – gneisses and migmatites that underwent retrograde metamorphism and biotite metagranites to metagranodiorites and orthogneisses and porphyritic biotite granite (Cháb et al. 2009). Nevertheless, they are at different elevations and windward and leeward slopes have created the variability. The rain gauge on Sněžka Mt. (S in Tab. 2) is the most relevant because of its elevation (1602 m a.s.l.) and close vicinity to the Obří důl Valley. Historical rainfall data are also available from this station (see Tab. 2) (Schneider 1897; Demuth 1901; Jitrasek 1915). Rainfall data were analysed for the available period up to 2006, because data after 2006 suggested rather low rainfall values. The rain gauge at Luční bouda only has available data since 2009 and was also not analysed due to low rainfall values.

4.2 Data analyses

Daily rainfall, hourly rainfall, antecedent rainfall and cumulative rainfall data were used to estimate the rainfall threshold.

4.2.1 Daily rainfall

Daily rainfall data represent the total amount of rainfall measured in the selected CHMI rain gauges (LCB,

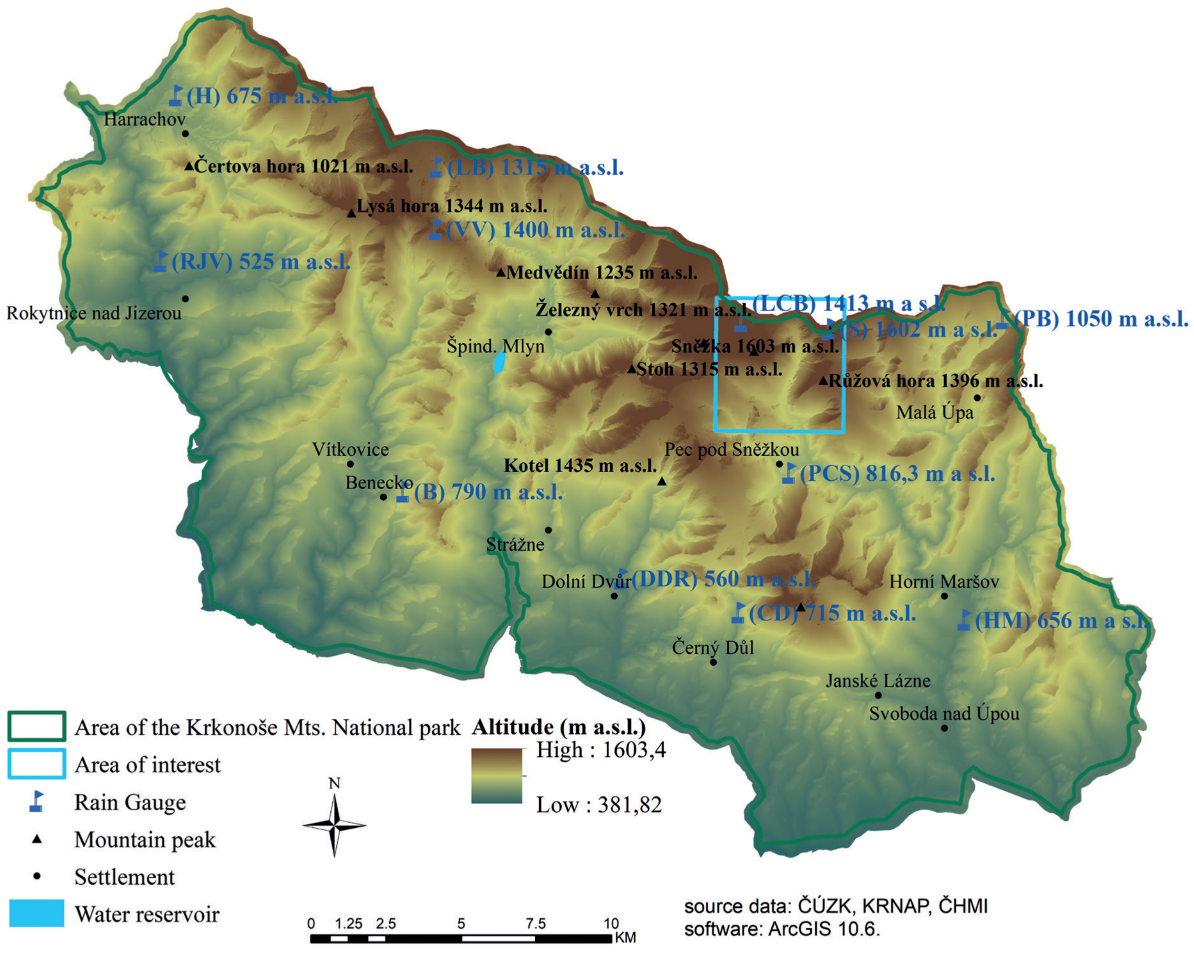

Fig 3. Map of rain gauges: (S) – Sněžka, (LCB) – Luční bouda, (PCS) – Pec pod Sněžkou, (PB) – Pomezní boudy, (CD) – Černý důl, (DDR) – Dolný Dvůr-Rudolfov, (HM) – Horní Maršov, (LB) – Labská bouda, (B) – Benecko (according to CHMI).

Tab. 2 Location of the selected rain gauges in the area of interest (according to CHMI and NCDC NOAA); rain gauge: see Fig. 3, type of rain gauge: AWS – automatic weather station; AWS_3 – automatic weather station without operator; ACS_1 – first-order automatic climatological station; ACS_4 – fourth-order automatic climatological station; APS – automatic precipitation station; MPS – manual precipitation station.

Rain gauge	Code	Elevation (m a.s.l.)	Distance from the study area (km)	Measurement period (CHMI)/ Historical data	Measurement period (NCDC NOAA)	Type of rain gauge
(S)	H7SNEZ01	1602	0.5–1.65–2.7	Historical data	01/11/1952–31/12/1963	AWS_3
					01/10/1973– 01/11/2017	
(LCB)	HLUCB01	1413	1.6–3	20/01/2009–31/12/2014	–	ACS_1
(PCS)	HPECS01	8163	2.4–5	15/06/1962–31/12/1971	01/11/1936–31/12/1941	AWS
				01/02/1988–31/12/2014	01/10/1988–01/11/2017	
(PB)	H1POMB01	1050	6.3–7.5	10/05/1994–31/12/2014	–	APS
(CD)	H1CDUL01	715	7.3–10	24/06/2005–31/12/2014	–	APS
(DDR)	H1DDVU01	560	8.3–10.5	16/05/1894–31/03/1939	–	APS
				01/05/1941–31/07/1945		
				01/04/1963–31/12/2014		
(HM)	H1HMAR01	565	9.4–11.4	01/01/1961–31/12/2014	–	APS
(LB)	H1LBOU1	1315	12.8–14.3	01/01/1961–31/04/1997	–	ACS_1
				21/06/1997–31/09/1999		
				01/10/2002–31/12/2014		
(B)	P2BENE01	790	13.5–14.5	01/01/1932–31/12/1937	–	MPS
				01/01/1938–31/12/2014		
(RJV)	P2ROKY01	525	21–22	22/07/1958–31/12/2014	–	MPS
(H)	P2HARR01	675	22–23	01/01/1951–31/12/2014	–	ACS_4
(VV)	HVITKO1	1400	12.3–13.3	01/07/1945–31/05/1974	–	–
				01/07/1974–31/12/1978		
(OD)	–	–	–	Historical data	–	–
(SE)	–	–	–	Historical data	–	–

PCS, PB, CD, DDR, HM, LB, B, RJV, H and VV in Tab. 2) and in the selected National Centers for Environmental Information (NCDC NOAA) (rain gauge: S and PCS) during a single day (from 7:00 a.m. of the first day to 7:00 a.m. the next day). These records were provided by CHMI for the period 1894–2014 for the selected rain gauges. Daily rainfall was analysed with the selected maximum values from the rain gauges. We selected data over 100 mm and then compared them with values from other rain gauges, to avoid low rainfall values which were common. For the station on Sněžka Mt. (rain gauge S in Tab 2.) the daily rainfall data were taken from historical chronicles (see in Tab. 2) (Anonymous 1889–1941, 1897a, 1897b; Schneider 1897; Demuth 1901; Jitrasek 1915) and from the National Centers for Environmental Information (rain gauge S_2) (NCDC NOAA).

4.2.2 Hourly rainfall

One-hour, two-hour, two-hour and thirty minute, and four-hour rainfall data were used from historical chronicles (Anonymous 1889–1941, 1897a, 1897b; Schneider 1897; Demuth 1901; Jitrasek 1915) also from the Sněžka Mt. rain gauge. Hourly rainfall data from the other rain gauges were not available. Six-hour and twelve-hour rainfall data were used from the National Centers for Environmental Information (NCDC NOAA) for the Sněžka Mt. and Pec pod Sněžkou rain gauges.

4.2.3 Antecedent precipitation index

The API was firstly expressed by Kohler and Linsley (1951). The equation is generally defined as follows:

$$API_n = \sum_{i=1}^{n} c^i \times P_i \; [mm]$$

Where n is the total number of days prior to the causal rainfall, usually 5, 10 or 30;
i is the number of days counting backwards from the date on which the API is determined;
c is an evapotranspiration constant (for the Czech Republic it is c= 0.93 (Steinhart 2010);
P_i is the amount of precipitation in days prior to the causal rainfall (mm).
The API was calculated for a number of days (n = 5, 10, and 30) from the daily amounts of rainfall in the period when debris flows occurred and selected dates when daily rainfall reached 100 mm.

4.2.4 Cumulative rainfall data

We considered 90-day cumulative daily rainfall data only from the selected CHMI rain gauges (CD, PCS, PB, DDR, HM, LB, B, RJV, H, VV), due to their availability. We compared the cumulative rainfall data from when debris flows occurred with the daily rainfall data over 100 mm from the other rain gauges (during the period between 1897 and 2006). The rainfall values from 2006 to 2014 were relatively low in all of the available rain gauges.

4.2.5 Determination of rainfall thresholds

The rainfall conditions that trigger debris flows can be different. The rainfall thresholds were determined based on daily, hourly, 90-day cumulative rainfall data, and the antecedent precipitation index for 5, 10 and 30 days. It was important to determine the number of days for the antecedent rainfall and to analyse the correlation between the daily rainfall in relation to the debris flows events and the corresponding antecedent rainfall (Zezere et al. 2005) for three periods: 5, 10 and 30 days from two selected rain gauges, Pec pod Sněžkou and Labská bouda.

5. Results

5.1 Daily rainfall

The highest daily rainfall in the years when debris flows occurred and from all the available rainfall data was 266 mm in the Obří důl Valley on 29 July 1897 (see Fig. 4). On the same day, the Sněžka Mt. (S) rain gauge measured 239 mm, whereas the meteorological station at Sedmidolí registered only 93 mm. This means that there were rather large differences over relatively short distances of 0.5–2.7 km (see Tab. 2). Other debris flows (Fig. 4) were triggered during rather low daily rainfall levels (0.2 mm (DDR), 18.5 mm (RJV), 21.6 mm (DDR), 26.5 mm (B), 29.3 mm (B) and 65.7 mm (B)), it which that daily rainfall records could not be considered alone. This is also supported by the fact that we found several days with rather high daily rainfall without any debris flow (Fig. 4). The highest daily rainfall when no debris flow was registered is from 7 August 2006 (193.1 mm at rain gauge B), the second highest rainfall value was measured on 31 August 2002 at rain gauge PB (191 mm). The rainfall recorded at several of the gauges reached the limit of 100 mm. Five of the selected rain gauges (PCS, PB, DDR, LB and VV) recorded 20 days when the daily rainfall levels were up to 100 mm. Rain gauge PB is only 6.3–7.5 km from the study area, which is relatively close.

5.2 Hourly rainfall

First of all, we checked the highest hourly rainfall levels during debris flow events. The highest was registered during 17 July 1882 and reached 45 mm (rain gauge S). The same station also measured very high daily rainfall (see the chapter above). The other debris flow events could not be explained by hourly rainfall extremes. On 18 June 1974, station S measured almost 73 mm in 3 hours during a heavy storm. What is again surprising is that no debris flow occurred on

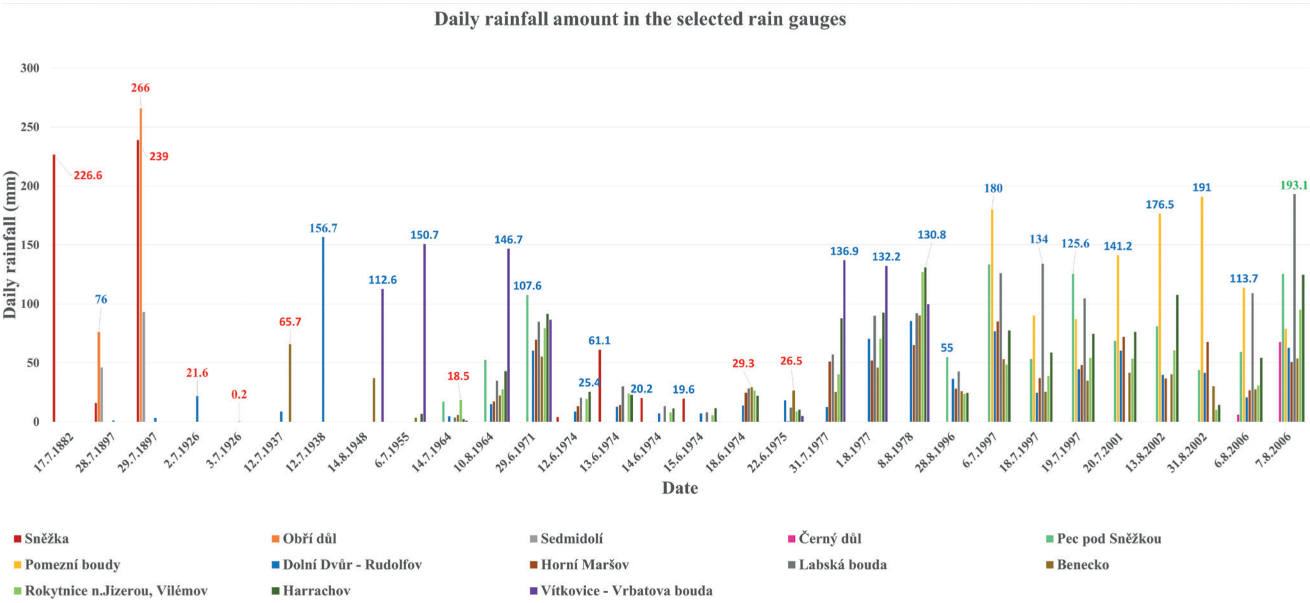

Fig. 4 Daily rainfall levels in the selected rain gauges (Sněžka, Obří důl, Sedmidolí, Černý důl, Pec pod Sněžkou, Pomezní boudy, Dolní dvůr Rudolfov, Horní Maršov, Labská bouda, Benecko, Rokytnice n. Jizerou – Vilémov, Harrachov, Vítkovice-Vrbatova bouda) in the period from 1882 to 2006. Colours: red – occurrence of debris flows (source: CHMI and Anonymous 1889–1941, 1897a, 1897b, Schneider 1897, Demuth 1901, Jitrasek 1915); blue – debris flows did not occur; green – the total maximum daily rainfall when debris flows did not occur (source: CHMI).

Tab. 3 Hourly rainfall levels from the Sněžka Mt. and Pec pod Snežkou rain gauges in the period from 1882 to 2002. Colours: blue – occurrence of debris flows; white – no debris flows; the number in green is the absolute highest maximum twelve-hour rainfall data (mm); dash – data not available (source: NCDC NOAA).

Date	S						PCS	
	1 h	2 h	2 h 40 min	4 h	6 h	12 h	6 h	12 h
17/7/1882	45.0	–	–	178.0	–	–	–	–
28/7/1897	–	–	–	–	–	–	–	–
29/7/1897	11.0	–	–	–	–	–	–	–
2/7/1926	–	–	–	–	–	–	–	–
3/7/1926	–	–	–	–	–	–	–	–
12/7/1937	–	–	–	–	–	–	–	–
12/7/1938	–	–	–	–	–	–	–	–
14/8/1948	–	–	–	–	–	–	–	–
6/7/1955	–	–	–	–	–	–	–	–
14/7/1964	*storm 48.2	25	–	–	–	–	–	–
10/8/1964	–	–	–	–	–	–	–	–
29/6/1971	–	–	–	–	–	–	–	–
12/6/1974	–	–	–	–	–	5.9	–	–
13/6/1974	–	–	–	–	–	74.9	–	–
14/6/1974	31.0	–	–	–	–	–	–	–
15/6/1974	–	–	–	–	–	17.0	–	–
18/6/1974	–	–	72.9	–	–	70.1	–	–
22/6/1975	–	–	–	–	–	10.1	–	–
31/7/1977	–	–	–	–	1.01	–	–	–
1/8/1977	–	–	–	–	–	80.01	–	–
8/8/1978	–	–	–	–	–	10.9	–	–
28/8/1996	–	–	–	–	–	209.6	–	–
6/7/1997	–	–	–	–	–	14.9	–	0
18/7/1997	–	–	–	–	–	37.1	7.9	–
19/7/1997	–	–	–	–	–	79	7.1	–
20/7/2001	–	–	–	–	–	23.1	–	65
13/8/2002	–	–	–	–	–	13.2	–	–
31/8/2002	–	–	–	–	–	0.8	–	–

28 August 1996 when 209.6 mm were registered at station S during a 12-hour period. Nevertheless, only moderate hourly rainfall was identified (see Tab. 3) at all of the other stations in the region.

5.3 Antecedent precipitation index (API)

The antecedent precipitation index was calculated for all debris flow events and days were selected when daily rainfall levels reached at least 100 mm. No significant values were determined for years when debris flows occurred and on the other hand many high values were identified for the following years without any debris flow. The highest API_5 was recorded in 2006 and did not result in a debris flow, whereas years when debris flows occurred, i.e. 1897, 1926, 1937, 1964, 1974 and 1975, show much lower API_5. The highest API_{10} was also recorded in 2006, but the highest API_{30} was calculated in 1997 (Tab. 4). From this analysis it is clear that the antecedent precipitation index itself (in this area) could not be responsible for any debris flows during the monitored period. We have to look at combinations with daily or hourly rainfall.

5.4 Cumulative rainfall data

Fig. 5 includes selected data of cumulative rainfall – both the highest values when debris flows occurred and did not occur. From the days when debris flows occurred, the 90-day cumulative rainfall reached a maximum of 413.7 mm on 3 July 1926. This value is approximately 370 mm lower than the cumulative rainfall from days when no debris flows occurred. The average amount from all of the calculated cumulative rainfall data is 348.8 mm. Only two values when debris flows occurred (in 1926) were above this cumulative rainfall average, i.e. values in 1897 were comparable and other values were lower. Unfortunately, no cumulative rainfall data were available for the debris flow

Tab. 4 Antecedent precipitation index for 5, 10 and 30 days (calculated from the source data: CHMI) (Blue colour represent debris flow events).

	PCS			PB			DDR			HM			LB			B			RJV			H			VV		
Date/ API	5	10	30	5	10	30	5	10	30	5	10	30	5	10	30	5	10	30	5	10	30	5	10	30	5	10	30
17/7/1882																											
28/7/1897							26.4	33.9	53.1																		
29/7/1897							33.9	57.2	71.3																		
2/7/1926							2.7	11.1	61																		
3/7/1926							2.9	11.3	63.2																		
12/7/1937							11.9	22	35.3							25	25	40.7									
12/7/1938							51.7	63.2	73.2																		
14/8/1948																25.3	31.6	55.2							127.5	14	157.6
6/7/1955																17.6	20.2	30.1				22.5	28.4	39	49.1	59.7	72.3
14/7/1964	2.9	11.7	31.1				6.4	8.6	27.4	12.1	14.5	27.4	14.5	19.4	41	13.7	16.8	33.3	8.3	13.6	24.9	5.6	11.1	28.2	11.9	17.8	42.5
10/8/1964	41.6	60	71				42.1	55.8	74	40.9	48.4	57.4	30.3	43.6	47.8	34.6	50.3	55.1	26.9	44.3	54.9	26.7	47.7	55.6	44.9	56.2	62.7
29/6/1971	16.4	26	47.3				5.1	25.4	43.5	7.2	15.9	31.4	10.9	27.9	50.7	12.4	30.4	55.8	15.5	33.3	56.6	20.7	41.1	67.1	20.8	42	74.6
12/6/1974							6.8	9.7	20.3	9.9	12.7	21.9	6	6	16.9	5.8	9.7	19.5	19.5	27.6	40.8	19.1	22.9	47.4			
13/6/1974							12.4	17.1	26.9	20.5	24.1	32.7	23.1	24.5	34.7	15.8	20.7	29.9	33.9	43.5	55.9	40.2	44.9	67.7			
14/6/1974							19.8	25.2	36.9	28.4	33.4	42.7	46.7	50.7	58.3	26.1	30.1	40.1	47.9	55.8	71.1	50.8	59.6	81			
15/6/1974							24.6	30	39.4	25.7	31.6	38.9	55.2	59.4	65.2	29	33.4	42.1	51.1	59.1	73	56.9	66	83.9			
18/6/1974							26.2	37	42	12.1	26.3	33.1	54.7	70.7	75.8	29.4	33.4	41.1	28.3	52	59.2	35.1	63	76.7			
22/6/1975							22.8	33.2	37.1	16.3	38.5	42.7	17.5	23.6	28.9	30.7	32.3	33.4	15.2	33.1	34.2	19.4	42.2	52.8	12.7	35.7	48.3
31/7/1977							0	15	29.5	0.4	13.4	27.4	0	12.3	29	0.4	11.8	24.3	0.5	16.7	30.6	0.7	17.2	30.5	0.6	14.7	40.2
1/8/1977							11.6	24.6	39	47.4	59.9	72.7	53	64	80	23.9	34	46.1	37.7	52.8	65.7	82.1	97.4	109.9	127.9	140.6	164.2
8/8/1978							2.8	3.9	12.1	8.1	8.6	15	8	8	22.5	4.5	4.6	15.6	9.1	9.1	9.1	16.6	16.7	27	11	11.1	26.8
28/8/1996	7.4	13	54.1				23.4	23.4	42.1	22.4	22.4	44.2	26.4	26.6	73.9	28.5	31.5	52.7	27	29.6	47	18.7	23	42.5			
6/7/1997	34.3	50.9	76.7	45.3	64	78.9	19.5	35	55.5	31	44.6	62.7	56.2	78	86	23.2	35.2	59.8	33.9	39.4	55.9	27.2	33.4	51.4			
18/7/1997	44.4	47.1	163.9	28.9	75	190.2	51	52.8	136.2	21.2	21.2	110	22	24	146.7	36.3	38.3	103.4	8.7	8.7	67.1	6.1	6.2	75.3			
19/7/1997	90.8	93.3	201.9	42.3	42.4	192.3	70.4	72.1	149.6	54.1	54.1	136.4	146	146	261.1	57.4	59.2	117.9	44.1	44.1	97.6	60.1	60.3	124.5			
20/7/2001	14.2	24.6	33.1	21.2	24	41.8	10.8	21.8	44.9	20.5	24.4	38.2				27.8	32.9	54.8	25.9	31.8	55.1	27.4	31.8	57.1			
13/8/2002	14.5	15.5	26	23.6	23.6	42.6	49.2	50.7	63.8	16.1	16.9	41.2				23.2	28.9	38.7	32.9	41.9	50.1	35.7	42.7	54.2			
31/8/2002	10.1	15.1	43	9.9	14.5	75.1	4.9	8.3	33.6	1.7	6.6	23.8				12.9	15.1	33.7	12.7	14.9	43.6	5.3	8.6	51.3			
6/8/2006	44.4	53.2	56.2	119.5	124.5	130.7	19.7	26.5	33.6	31.4	32.8	38.3	89.6	105.7	112.4	18.9	24.2	27.6	24.6	42.6	46.3	40.1	59.1	62.3			
7/8/2006	95.5	104.6	107.3	216.6	221.5	227.3	37.6	44	50.6	53.7	55.3	60.4	183.8	199.6	205.9	43	47.9	51.1	51.6	68.2	71.7	87.6	105.3	108.2			

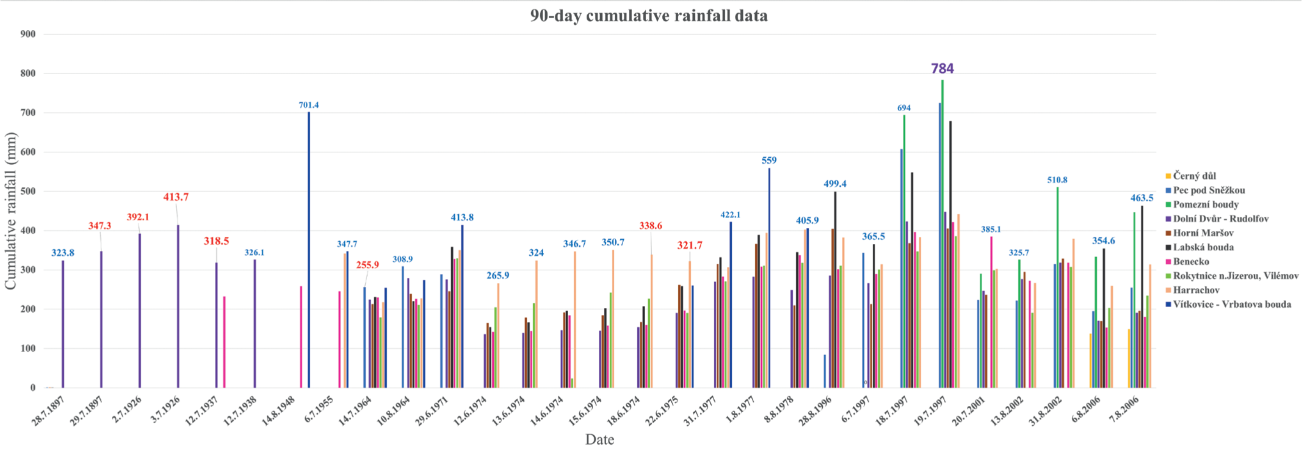

Fig. 5 Cumulative rainfall data for 90 days. Colours: red – occurrence of debris flows; blue – no occurrence of debris flows; violet is the absolutely highest value of cumulative rainfall (calculated from the source data: CHMI).

events (Nos. 5, 7, 8, 9, 10, 34 and 35) – (Tab. 1). In July 1897, July 1937, July 1964, June 1974 and June 1975 when debris flows occurred, values of cumulative rainfall were below the average. In August 1948 and in July 1997 when debris flows did not occur, values of cumulative rainfall were two times higher than the average of all cumulative rainfall data.

5.5 Relation between daily and antecedent rainfall

In order to determine the number of days for the antecedent rainfall, we considered the correlation analysis between the daily rainfalls in relation to the debris flows events and the corresponding antecedent rainfall (Zezere et al. 2005) for two rain gauges, Pec pod Sněžkou and Labská bouda for three periods: 5, 10 and 30 days. The results are shown in Fig. 6. The red points depict the debris flow events whereas the black points show the selected years with high rainfall from a period of 42 years (1964 to 2006). Fig. 6 shows that debris flows occurred after low daily rainfall in combination with small to moderate amounts of antecedent rainfall. On the contrary, high rainfall events in combination with high levels of antecedent rainfall did not create debris flows. We have to stress that debris flow events from the end of the 19th century are not included in Fig. 6, because it was not possible to calculate the antecedent rainfall.

5.6 Combination of different rainfall characteristics

If we consider all of the above analysed factors (hourly and daily rainfall, API, cumulative rainfall), the debris flows from 1882 and 1897 may be explained by the huge daily rainfall amounts, partly supported by hourly rainfall. The influence of API and cumulative rainfall from 1882 could not be evaluated because of missing rainfall data. API data from 1897 are available only for a rain gauge located far from the debris flow source area. The occurrence of the other debris flows in 1926, 1937, 1964 and 1975 is not significantly supported by the available data. Possibly only the 1974 (see Tab. 6) debris flow (18 June 1974) occurred after a combination of moderate hourly rainfall and moderate API, but these values are nothing compared to 1882 and 1897. We have to stress that there were several years over the last few decades when stations registered high values of API and daily rainfall (e.g. middle of July 1997 or beginning of August 2006) but no debris flows were registered. This means that there must be significant differences in the rainfall distribution over rather short distances in the mountainous area of the Krkonoše Mts.

6. Discussion

Debris flows triggered by rainfall have not been studied in detail in the Krkonoše Mts. The majority of the published works are focused on particular debris flow events describing the associated rainfall amounts.

The estimated thresholds for daily rainfall intensity are 225 mm, without the support of API and cumulative rainfall. A comparison of these thresholds with other debris flow studies from the Czech Republic is problematic due to the fact that many of the authors worked on deep seated landslides (Gil and Dlugosz 2006), described isolated events and estimated threshold values for landslide initiation for the Outer Western Carpathians (Bíl et al. 2016). Only a few works focused on individual landslides where numerous events took place (Krejčí et al. 2002; Pánek et al. 2011). However, they did not attempt to estimate the threshold. Work from Smědavská hora Mt. in the Jizerské hory Mts. describes the initiation of debris flows on Smědavská hora Mt. (Smolíková et al. 2016). From this, the authors concluded that for the initiation of debris flows, a combination of API, daily/hourly rainfall and values of short intensities of 10/15 min is far more important than individual extremes. Significant rainfall events have been recorded in the last 30 years without any debris flow events (Smolíková et al. 2016).

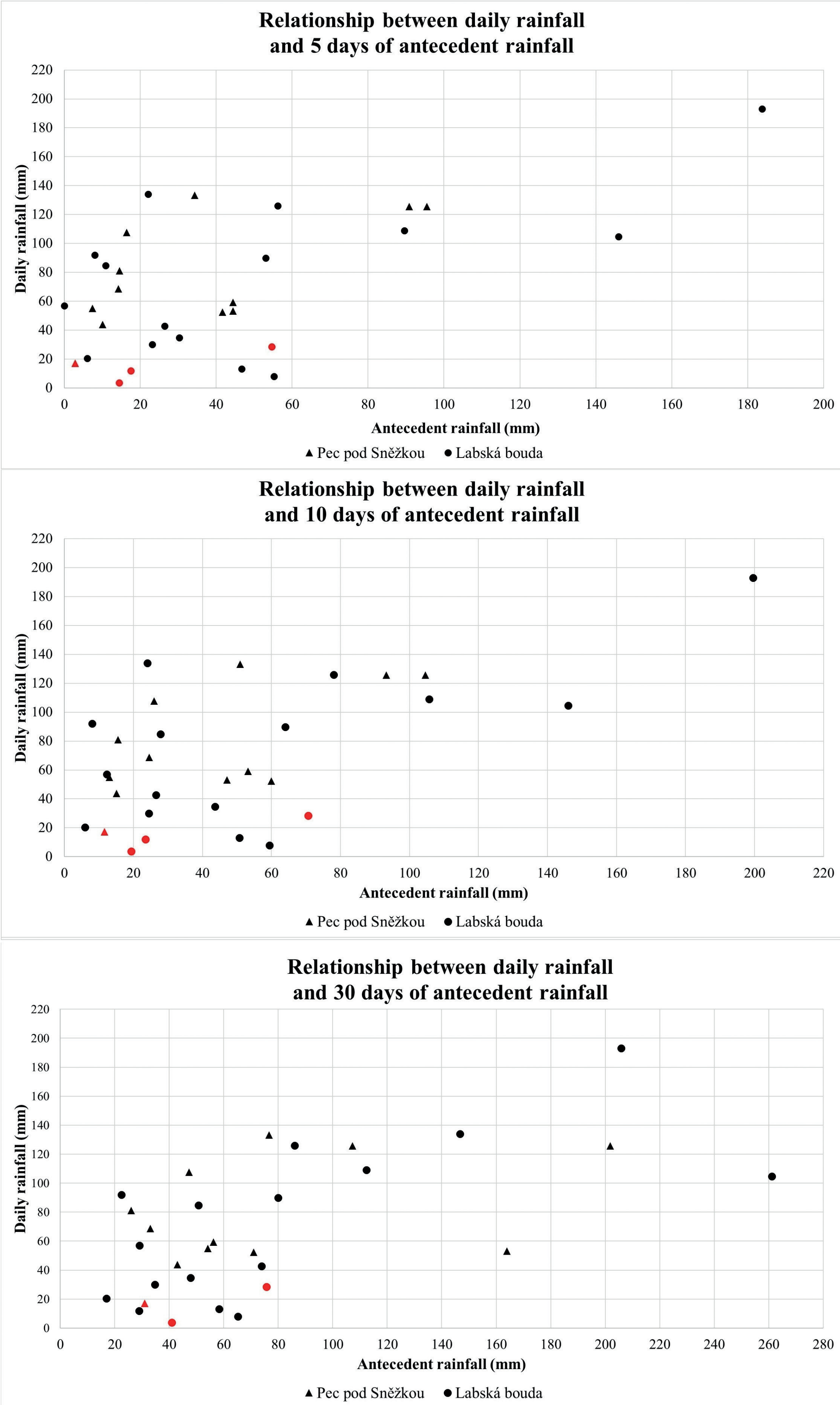

Fig. 6 Relationship between daily rainfall and antecedent rainfall for the period 1964–2006 from Pec pod Sněžkou and Labská Bouda rain gauges. Red points depict debris flow events and black points show years with high rainfall, which did not trigger any debris flows (calculated from the source data: CHMI).

Tab. 5 Daily rainfall thresholds for the initiation of landslides in comparison with our derived value (R > 225 mm).

Continent	Country	Area	Type of landslide	Value of threshold type	Reference
Europe	Italy	Sarno, Campania	All types of landslides	R > 55 mm (lower threshold)	Biafiore et al. (2002)
Europe	Italy	Sarno, Campania	All types of landslides	R > 75 mm (upper threshold)	Biafiore et al. (2002)
Europe	Spain	Llobregat valley, Pyrenees Mts.	Shallow landslide, Debris flows	R > 160–200 mm	Corominas and Moya (1996)
Asia	Japan	Hokkaido Island	All types of landslides	R > 200 mm	Endo (1970)
North America	USA	Alamanda Country	All types of landslides	R > 180 mm	Nilsen et al. (1976)
North America	USA	Los Angeles Area	All types of landslides	R > 235 mm	Campbell (1975)

Tab. 6 Summary of detailed rainfall data in selected dates from all rain gauges (blue lines describe the debris flows events, white – no debris flows; RG is rain gauge; daily amount – the total highest daily value from all rain gauges; hourly rainfall data – 1 h, 2 h, 6 h, 12 h; API 5/10/30 – the total highest API value; 90 days of cumulative rainfall (CR) – the total highest value of CR) (Source data: see previous Chapter 5 and Fig. 4, 5, 6, 7).

	Daily rainfall	RG	1h	RG	2h	RG	6h	RG	12h	RG	API 5	RG	API 10	RG	API 30	RG	90 days CR	RG
17/7/1882	226.6	(S)	45	(S)														
28/7/1897	76	(OD)									26.4	(DDR)	33.9	(DDR)	53.1	(DDR)	323.8	(DDR)
29/7/1897	266	(OD)	11	(S)							33.9	(DDR)	57.2	(DDR)	71.3	(DDR)	347.3	(DDR)
2/7/1926	21.6	(DDR)									2.7	(DDR)	11.1	(DDR)	61	(DDR)	392.1	(DDR)
3/7/1926	0.2	(DDR)									2.9	(DDR)	11.3	(DDR)	63.2	(DDR)	413.7	(DDR)
12/7/1937	65.7	(B)									25	(B)	25	(B)	40.7	(B)	318.5	(DDR)
12/7/1938	156.7	(DDR)									51.7	(DDR)	63.2	(DDR)	73.2	(DDR)	326.1	(DDR)
14/8/1948	112.6	(VV)									127.5	(VV)	31.6	(B)	157.6	(VV)	701.4	(VV)
6/7/1955	15.7	(VV)									49.1	(VV)	59.7	(VV)	72.3	(VV)	347.7	(VV)
14/7/1964	18.5	(RJV)			25	(S)					14.5	(LB)	19.4	(LB)	42.5	(VV)	255.9	(PCS)
10/8/1964	146.7	(VV)									44.9	(VV)	60	(PCS)	74	(DDR)	308.9	(PCS)
29/6/1971	107.6	(PCS)									20.8	(VV)	42	(VV)	74.6	(VV)	413.8	(VV)
12/6/1974	25.4	(H)							5.9	(S)	19.5	(RJV)	27.6	(RJV)	47.4	(H)	265.9	(H)
13/6/1974	61.1	(S)							75	(S)	40.2	(H)	45	(H)	67.7	(H)	324	(H)
14/6/1974	20.2	(S)							31	(S)	50.8	(H)	60	(H)	81	(H)	346.7	(H)
15/6/1974	19.6	(S)							17	(S)	56.9	(H)	66	(H)	83.9	(H)	350.7	(H)
18/6/1974	29.3	(B)							70	(S)	54.7	(LB)	70.7	(LB)	75.8	(LB)	338.6	(H)
22/6/1975	26.5	(B)							10	(S)	30.7	(B)	42	(H)	52.8	(H)	321.7	(H)
31/7/1977	136.9	(VV)					1.01	(S)			0.7	(H)	17	(H)	30.6	(RJV)	422.1	(VV)
1/8/1977	132.2	(VV)							80	(S)	127.9	(VV)	140.6	(VV)	164.2	(VV)	559	(VV)
8/8/1978	130.8	(H)							11	(S)	16.6	(H)	17	(H)	27	(H)	405.9	(VV)
28/8/1996	55	(PCS)							15	(S)	28.5	(B)	31.5	(B)	73.9	(LB)	499.4	(LB)
6/7/1997	180	(PB)							210	(S)	56.2	(LB)	78	(LB)	86	(LB)	343.5	(PCS)
18/7/1997	134	(LB)					7.9	(PCS)	37	(S)	51	(DDR)	75	(PB)	190.2	(PB)	607.5	(PCS)
19/7/1997	125.6	(PCS)					7.1	(PCS)	79	(S)	146	(LB)	146	(LB)	261.1	(LB)	784	(PB)
20/7/2001	141.2	(PB)							23	(S)	27.8	(B)	32.9	(B)	57.1	(H)	385.1	(B)
13/8/2002	176.5	(PB)							65	(PCS)	49.2	(DDR)	50.7	(DDR)	63.8	(DDR)	325.7	(PB)
31/8/2002	191	(PB)							0.8	(S)	12.9	(B)	15.1	(PCS)	75.1	(PB)	510.8	(PB)
6/8/2006	113.7	(PB)									119.5	(PB)	124.5	(PB)	130.7	(PB)	354.6	(LB)
7/8/2006	193.1	(LB)									216.6	(PB)	221.5	(PB)	227.3	(PB)	463.5	(LB)

Rainfall is not the only causative factor for debris flow initiation (Aleotti and Chowdhury 1999), geological, geomorphological, soil and vegetation conditions also have to be taken into consideration. In particular, local climatic and the geomorphological conditions of the geological structure of the area differ from one threshold to another (Guzzetti et al. 2007). Nevertheless, we tried to compare our derived daily rainfall threshold for the Obří důl Valley with others already published (Tab. 5).

The limits of our presented research are also influenced by the following circumstances: the data from 1897 were measured at different rain gauges, thus they may be less significant for determining thresholds. A significant problem was the incompleteness of the rainfall data from the selected rain gauges. The rain gauges are located at different elevations and distances from the triggering area of the debris flows. The orographic effect also has an influence on rainfall distribution.

7. Conclusions

The rainfall analysis for debris flows from the available data revealed that in the area under study daily rainfall above 225 mm could trigger debris flows (e.g. years 1882 and 1897), even without support of API. A combination with hourly rainfall cannot be demonstrated, because we do not have sufficiently accurate data from these years. We have data from the last 40 years from several stations in the surroundings of the Obří důl Valley, which revealed that daily rainfall between 150 and 200 mm/day does not create debris flows, even with the support of API (see Tab. 6). Also, the heavy storm during 28 August 1996 when 209.5 mm/12 hours were measured did not create a debris flow. It seems that the local rainfall threshold is above 225 mm/day, without the possibility to consider its combination with hourly rainfall or API.

The API itself probably has only a limited influence on the triggering mechanism and has to be considered in combination with daily or hourly rainfall. Cumulative rainfall and API are not good indicators for predicting debris flows in our study area as most of the historical debris flows did not occur during the highest cumulative rainfall events or significant API values. A plausible explanation for this is that the infiltration of intensive rainfall is limited by the permeability of the soils. Daily rainfall could result in rapid oversaturation of the weathered mantle and soils and eventually cause the slope to be more susceptible to failure.

There are significant differences in the measured data even over rather short distances. Unfortunately, it did not help us to substitute the data from one station with data from another station. We also have to resolve the issue that some of the debris flows remain undated.

The temporal probability of the occurrence of debris flows was directly estimated based on the statistical relationship between the historical debris flow events and rainfall data. The rainfall thresholds were estimated using all debris flow episodes without considering the specifics of the sizes of the debris flows or the number of debris flows in the episodes.

Acknowledgements

This study was supported by the Grant Agency of Charles University in Prague (GAUK No. 367015). We are grateful to the Czech Hydrometeorological Institute for providing the rainfall data. We would also like to thank the administration of the Czech national park for providing permission to work in the region. We gratefully acknowledge both of the anonymous reviewers for their valuable comments, which significantly improved the manuscript.

References

Aleotti, P. (2004): A warning system for rainfall-induced shallow failures. Engineering Geology 73, 247–265, https://doi.org/10.1016/j.enggeo.2004.01.007.

Aleotti, P., Chowdhury, R. (1999): Landslide hazard assessment: summary review and new perspectives. Bull Eng Environ 58, 21–44, https://doi.org/10.1007/s100640050066.

Anonymus (1889-1941): Gedenkenbuch der Gemeinde Gross Aupa, I. Teil.

Anonymus (1897a): Bericht über Hochwasserschäden. In: Der Wanderer im Riesengebirge 17(9), Hirschberg, 130–131.

Anonymus (1897b): Der Wolkenbruch im Riesengebirge. In: Der Wanderer im Riesengebirge 17(9), 131–132.

Biafiore, M., Braca, G., De Blasio, A., Martone, M., Onorati, G., Tranfaglia, G. (2002): Il monitoraggio ambientale dei territori campani a rischio di frane e di alluvioni: lo sviluppo della rete idropluviometrica del Servizio Idrografico e Mareografico Nazionale. Unpublished report.

Bíl, M., Andrášik, R., Zahradníček, P., Kubeček, J., Sedoník, J., Štěpánek, P. (2016): Total water content thresholds for shallow landslides, Outer Western Carpathians. Landslides 13, 337–347, https://doi.org/10.1007/s10346-015-0570-9.

Blahůt, J. (2008): Mapa náchylnosti terénu Krkonoš ke vzniku lavin vytvořená pomocí nástrojů GIS a statisticko – pravděpodobnostních metod. Opera Corcontica 45, 35–44.

Brand, E. W. (1989): Occurrence and significance of landslides in Southeast Asia, In: Brabb, E. E., Harrod, B. L. (eds) Landslides: extent and economic significance. A. A. Balkema, Rotterdam, 303–324.

Campbell, R. H. (1975): Soil slips, debris flows, and rainstorms in the Santa Monica Mountains and vicinity, southern California. In: US Geological Survey Professional Paper 851, Washington DC: U.S. Government Printing Office 51, https://doi.org/10.3133/pp851.

Chaloupský J., Teisseyre, H. (1968): Geology of the West Sudeten. Guide to Excursion 8 AC, International geolog. congress, XXIII session, Prague 1968, 17–18, Praha.

Corominas, J., Moya, J. (1996): Historical landslides in the Eastern Pyrenees and their relation to rainy events. In: Irigaray, C., Fernandez, T. (eds) Landslides. Chacon J. Rotterdam, A. A. Balkema, 125–132.
Corominas, J., Moya, J. (1999): Reconstructing recent landslide activity in relation to rainfall in the Llobreat River basin, Eastern Pyrenees, Spain. Geomorphology 30, 79–93, https://doi.org/10.1016/S0169-555X(99)00046-X.
Corominas, J. (2000): Landslide and climate. Keynote lecture. In: Proc. 8th Int. Symp. On Landslides (Bromhead E, Dixon N, Ibsen ML, eds.). In: Guzzeti, F. et al. (eds.): Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorology and Atmospheric Physics 98(3–4), 239–267.
Crosta, G. (1998): Regionalization of rainfall thresholds: an aid to landslide hazard evaluation. Environ Geol 35(2), 131–145, https://doi.org/10.1007/s002540050300.
Crosta, G. B., Frattini, P. (2001): Rainfall threshold for triggering soil slips and debris flow. In: Proc. 2nd EGS Plinius Conf. on Mediterranean Storms. In: Danišík, M., Migoń, P., Kuhlemann, J., Evans, Nj., Dunkl, I., Frissch, W. (2010): Thermochronological constraints on the long-term erosional history of the Karkonosze Mts., Central Europe. Geomorphology 117, 78–89.
Crozier, M. J., Eyles, R. J. (1980): Assessing the probability of rapid mass movement. In: Technical Groups, Proceedings of 3rd Australia–New Zealand Conference on Geomechanics. New Zealand, Wellington: 247–251
Czerwiński, J. (1967): Osuwisko w doline Lomniczki w Karkonoszach. In: Opera Corcontica.
Dahal, R. K., Hasegawa, S. (2008): Representative rainfall thresholds for landslides in the Nepal Himalaya. Geomorphology 100(3–4), 429–443, https://doi.org/10.1016/j.geomorph.2008.01.014.
Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Masuda, T., Nishino, K. (2009): Failure characteristics of rainfall-induced shallow landslides in granitic terrains of Shikoku Island of Japan. Environ Geol 56(7), 1295–1310, https://doi.org/10.1007/s00254-008-1228-x.
Danišík, M., Migoń, P., Kuhleman, J., Evans, NJ., Dunkl, I. Frisch, W. (2010): Thermochronological constraints on the long-term erosional history of the Karkonosze Mts., Central Europe. Geomorphology 117, 78–89, https://doi.org/10.1016/j.geomorph.2009.11.010.
Demuth, J. (1901): Der politische Bezirk Trautenau.
D'odorico, P., Fagherazzi, S. (2003): A probabilistic model of rainfall-triggered shallow landslides in hollows: a long-term analysis. Water Resour Res 39(9), https://doi.org/10.1029/2002WR001595.
Dvořák, I. J., Kociánová, M., Pírková, L. (2004): Příklad využití technologií GPS a GIS při studiu kryoplanačních teras na Luční a Studniční hoře. Opera Corcontica 41, 18–24.
De Quervain, M.R. et al. (1981): Avalanche atlas. UNESCO, International Commission on Snow and Ice of the International Association of Hydrological Sciences. Paris: 1–265.
De Vita, P., Reichenbach, P. (1998): Rainfall – triggered landslides: Reference list. Environmental Geology 35(2–3). Springer Verlag, 219–233.
Endo, T. (1970): Probable distribution of the amount of rainfall causing landslides. Annual report, Hokkaido Branch, Govern. Forest Experiment Station, Sapporo, 123–136.
Engel, Z. (1997): Pleistocénní zalednění české části Krkonoš. Przyroda Sudetów Zachodnich 6, 223–234.
Engel, Z. (2003): Současný stav poznatků o pleistocénním zalednění české části Krkonoš. Geografie – Sborník ČGS 102, 288–302
Engel, Z. (2007): Late Pleistocene glaciations in the Krkonoše Mts. In: Goudie AS. Kalvoda, J. (eds) Geomorphological Variations. P3 K. Praha, 269–285.
Engel, Z., Nývlt, D., Křížek, M., Treml, V., Jankovská, V., Lisá, L., (2010). Sedimentary evidence of landscape and climate history since the end of MIS 3 in the Krkonoše Mountains, Czech Republic. Quat Sci Rev 29(7–8), 913–927, https://doi.org/10.1016/j.quascirev.2009.12.008.
Engel, Z., Braucher, R., Traczyk, A., Léanni, L., Asterteam (2014): ^{10}Be exposure age chronology of the last glaciation in the Krkonoše Mountains, Central Europe. Geomorphology 206, 107–121, https://doi.org/10.1016/j.geomorph.2013.10.003.
Frattini, P., Crosta, G., Sosio, R. (2009): Approaches for defining thresholds and return periods for rainfall triggered shallow landslides. Hydrol Process 23(10), 1444–1460, https://doi.org/10.1002/hyp.7269.
Giannecchini, R., Galanti, Y., Avanzi, G. D. (2012): Critical rainfall thresholds for triggering shallow landslides in the Serchio River Valley (Tuscany, Italy). Nat Hazards Earth Syst Sci 12(3), 829–842, https://doi.org/10.5194/nhess-12-829-2012.
Gil, E., Dlugosz, M. (2006): Threshold values of rainfalls triggering selected deep-seated landslides in the Polish flysch Carpathians. Stud Geomorphol Carpatho-Balc 11, 21–43.
Glade, T. (2000): Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical "antecedent daily rainfall model". Pure Appl Geophys 157(6), 1059–1079, https://doi.org/10.1007/s000240050017.
Glade, T., Crozier, M. J. (2005): The nature of landslide hazard impact. In: Glade, T., Anderson, M., Crozier, M. J. (eds): Landslide hazard and risk. Wiley, Chichester, 73–74, https://doi.org/10.1002/9780470012659.ch2.
Godt, J. W., Baum, R. L., Chleborad, A. F. (2006): Rainfall characteristics for shallow landsliding in Seattle, Washington, USA. Earth Surf Proc Land 31(1), 97–110, https://doi.org/10.1002/esp.1237.
Guzetti, F., Cardinali, M., Reichenbach, P., Cippola, F., Sebastiani, C., Gali, M., Salati, P. (2004): Landslides triggered by 23 November 2000 rainfall event in the Imperia Province, Liguria, Italy. Engineering Geology 73, 229–245, https://doi.org/10.1016/j.enggeo.2004.01.006.
Guzzetti, F., Peruccacci, S., Rossi, M., Stark, CP. (2007): Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorology and Atmospheric Physics 98(3–4), 239–267, https://doi.org/10.1007/s00703-007-0262-7.
Heyerdahl, H., Harbitz, C. B., Domaas, U., Sandersen, F., Tronstad, K., Nowacki, F., Engen, A., Kjestad, O., Devoli, G., Buezo, S. G., Diaz, M. R., Hernandez, W. (2003): Rainfall induced lahars in volcanic debris in Nicaragua and El Salvador: practical mitigation. In: Proceedings of international conference on fast slope

movements – prediction for risk mitigation, IC- FSM2003, Naples. Patron Pub, 275–282.
Horník, S., Buzek, L., Mičian, Ľ., Pech, J., Trnka, P. (1986): Fyzická geografie II. Státní pedagodigické nakladatelství Praha, 319.
Hraško, J., Linkeš, V., Němeček, J., Novák, P., Šály, R., Šurina, B. (1991): Morfogenetický klasifikační systém pôd ČSFR, VÚPÚ Bratislava, 106.
Cháb, J., Stránik, Z., Eliáš, M. (2009): Geologická stavba. In: Hrnčiarová, T., Mackovičin, P., Zvara. I. et al. Atlas krajiny České republiky. Praha, Ministerstvo životního prostředí ČR, Průhonice: Výzkumný ústav Silva Taroucy pro krajinu a okrasné zahradnictví, v.v.i., 106–107.
Janeček, M., Bečvář, M., Bohuslávek, J., Dufková, J., Dumbrovský, J., Dostál, T., Hula, J., Jakubíková, A., Kadlec, V., Krása, J., Kubátová, E., Novotný, I., Podhrázská, J., Tippl, M., Toman, F., Vopravil, J., Vrána, K. (2007): Ochrana zemědelské půdy před erozí, Metodika. (Protection of agricultural land against erosion, the methodology). Výzkumný ústav meliorací a ochrany půdy, v.v.i., Praha, 76 (in Czech).
Jitrasek, F. J. (1915): Volks – und Heimatskunde des politischen Bezirkes Hohenelbe I. Hohenelbe.
Juras, R., Pavlásek, J., Blahůt, J., Bašta, P., Klose, Z., Vítěslav, M., Balek, J. (2013): New tools for avalanche forecasting in the Krkonoše Mountains. International Snow Science Workshop Grenoble – Chamonix Mont – Blanc.
Kim, SK., Hong, WP., Kim, YM. (1991): Prediction of rainfall-triggered landslides in Korea. In: Bell, DH. (ed) Landslides. A. A. Balkema. Rotterdam, 989–994.
Kociánová, M., Spusta, V. (2000): Vliv lavinové aktivity na kolísání horní hranice lesa v Krkonoších. Opera Corcontica 37, 473–480.
Kociánová, M., Špatenková, I., Tondrová, A., Dvořák, I. A., Pilous, V. (2004): Základní a smíšené laviny ve vztahu k přemisťování svahovin a dynamice vegetace. Opera Corcontica 41, 86–99.
Kohler, M. A., Linsley, R. K. (1951): Predicting the runoff from storm rainfall. Weather Bureau, US Department of Commerce, Research Paper, 34, Washington.
Knippel, E. W. (1850): Rytina. In: Pilous, V. (1973): Strukturní mury v Krkonoších, I. Část. Opera Corcontica 10, 15–69, Praha.
Králík, F., Sekyra, J. (1969): Geomorfologický přehled Krkonoš, In: Fanta, J. (ed) Příroda Krkonošského národního parku, SZN, Praha, 59–87.
Krejčí, O., Baroň, I., Bíl, M., Jurová, Z, Bárta, J., Hubatka, F., Kašpárek, M., Kirchner, K., Stach, J. (2002): Some examples of deep-seated landslides in the Flysch Belt of Western Carpathians, In: Rybář, Stemberk, Wagner (eds.) Landslides. A. A., Balkema Publishers, Lisse, 373–380.
Křížek, M. (2007): Periglacial landforms above the alpine timberline in the High Sudetes, In: Goudie, A. S., Kalvoa, J. (eds) Geomorphological Variations, P3K, Praha, 313–337.
Křížek, M., Uxa, T. (2013): Morphology, sorting and microclimates of relict sorted polygons, Krkonoše Mountains, Czech Republic. Permafrost Perigl Proc 24, 313–321, https://doi.org/10.1002/ppp.1789.
Lysenko, V. (2007): Morfotektonická analýza Krkonoše pomocí metod dálkového průzkum Země. Opera Corcontica 44(1), 37–42.
Marchi, L., Arattano, M., Deganutti, A. M. (2002): Ten years of debris-flow monitoring in the Moscardo Torrent (Italian Alps). Geomorphology 46, 1–17, https://doi.org/10.1016/S0169-555X(01)00162-3.
Marques, R., Zezere, J., Trigo, R., Gaspar, J., Trigo, I. (2008): Rainfall patterns and critical values associated with landslides in Povoac¸a˜o County (Sa˜o Miguel Island, Azores): relationships with the North Atlantic Oscillation. Hydrol Process 22(4), 478–494, https://doi.org/10.1002/hyp.6879.
Migoń, P. (1992): Tektoniczne formy rzeźby na północznym stoku Karkonoszy. Opera Corcontica 29, 5–24.
Migoń, P. (1999): The role of preglacial relief in the development of mountain glaciation in the Sudetes, with the special reference to the Karkonosze Mountains. Z Geomorp NF Supp 113, 33–44.
Migoń, P., Pilous, V. (2007): Geomorfologie. In: Flousek, J. et al. (eds) Krkonoše – příroda, historie, život. Baset. Praha, 103–124.
Mishra, S. K., Singh, V. P. (2003): Soil conservation service curve number (SCS-CN) methodology. Kluwer Academic Publisher, Dordrech.
Myslil, V. (1997): Zpráva o geologickém průzkumu sesuvů v okrese Trutnov. Geomedia s.r.o. Archív Geofond, Praha, 13.
NCDC NOAA: National Centers for Environmental Information [online]: Available on: https://www7.ncdc.noaa.gov/CDO/cdodata.cmd.
Nilsen, T. H., Taylor, F. A., Brabb, E. E. (1976): Recent landslides in Alamanda County, California (1940–71). US Geological Survey Bul, 1398.
Novotný, J. (2000): Analýza vlivu vody na stabilitu svahu v jílových horninách. Autoreferát disertační práce k získání vědecké hodnosti kandidáta věd, MS, ÚSMH AV ČR, Praha, 27.
Ouvrier, H. (1933): Beiträge zur Morphologie des Hohen Riesengebirges. Veröff Schles Ges Erdkunde. Breslau, 88.
Pánek, T., Brázdil, R., Klimeš, J., Smolková, V., Hradecký, J., Zahradníček, P. (2011): Rainfall induced landslide event of May 2010 in the eastern part of the Czech Republic. Landslides 8(4), 507–516, https://doi.org/10.1007/s10346-011-0268-6.
Partsch, J. (1894): Die Vergletscherung des Riesengebirges zur Eiszeit. Forschungen zur deutschen Landse- und Volkskunde 8(2): 103–194.
Pasuto, A., Silvano, S. (1998): Rainfall as a triggering factor of shallow mass movements, A case study in the Dolomites, Italy. Environmental Geolog 35, 184–189, https://doi.org/10.1007/s002540050304.
Pilous, V. (1973): Strukturní mury v Krkonoších, I. Část. Opera Corcontica 10, 15–69, Praha.
Pilous, V. (1975): Strukturní mury v Krkonoších, II. Část. Opera Corcontica 12, 7–50, Praha.
Pilous, V. (1977): Strukturní mury v Krkonoších, III. Část. Opera Corcontica 14, 7–94, Praha.
Pilous, V. (2016): Krkonoše Mountains: A Case Study of Polygenetic Relief. In: Pánek, T., Hradecký, J. (eds.) Landscapes and Landforms of the Czech Republic. World Geomorphological Landscapes, https://doi.org/10.1007/978-3-319-27537-6_15.
Polemio, M., Sdao, F. (1999): The role of rainfall in the landslide hazard: the case of the Avigliano urban area (Southern Apennines, Italy). Eng Geol 53(3–4), 297–309, https://doi.org/10.1016/S0013-7952(98)00083-0.
Quitt, E. (1971): Klimatické oblasti Československa. Studia Geographica, 16, Geografický ústav ČSAV, Brno.

Rybář, J., Novotný, J. (2005). Vliv klimatogenních faktorů na stabilitu přirozeních antropogenních svahů. Zpravodaj Hnědé uhlí 3, 13–28.
Saito, H., Nakayama, D., Matsuyama, H. (2010): Relationship between the initiation of a shallow landslide and rainfall intensity–duration thresholds in Japan. Geomorphology 118(1–2), 167–175, https://doi.org/10.1016/j.geomorph.2009.12.016.
Sekyra, J., Sekyra, Z. (1995): Recent cryogenic processes. In: Soukopová, L. et al. Artic – alpine tundra in the Krkonoše. The Sudetes. Opera Corcontica 32, 5–88.
Schneider, G. (1897): Die Hochwasserschäden im Riesengebirge von Ende des fünfzehnten Jahrhunderts bis Juli 1897.
Schuster, R. L., Wieczorek, G. F. (2002): Landslide triggers and types. In: Rybář, J., Stemberk, J., Wagner, P. (eds.): Proceedings of the first European conference on landslides, Prague Czech Republic, June 24–26 2002, A. A. Balkema, 59–75.
Smolíková, J., Blahůt, J., Vilímek, V. (2016): Analysis of rainfall preceding debris flows on the Smědavská hora Mt., Jizerské hory Mts., Czech Republic. Landslides 13(4), 683–696, https://doi.org/10.1007/s10346-015-0601-6.
Sokol, F., Vavřík, K. (1971): Vznik a asanace svohových sesuvů v Hrubém Jeseníku. Ochrana přírody 26(4), 84–86.
Spusta, V., Kociánová, M. (1998): Lavinový katastr české části Krkonoš. Opera Concortica 35, 3–205.
Spusta Sen, V., Spusta Jun, V., Kociánová, M. (2003): Snow avalanche cadastre and winter situation in the Czech part of Krkonoše 1998/99–2002/03 (in Czech). Opera Corcontica 40, 5–86.
Spusta Sen, V., Spusta JUN, V., Kociánová, M. (2006): Snow avalanche cadastre of Czech part of Krkonoše 2003/04 at 2005/06 (in Czech). Opera Corcontica 43, 81–93.
Suk, P., Klimánek, M. (2011): Creation of the snow avalanche susceptibility map of the Krkonoše Mountains using GIS. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brinensis 59(28), 237–246.
Steinhart, M. (2010): Application of rainfall-runoff model Boussmo. Diploma thesis, Czech Agriculture University in Prague (in Czech).
Šourek, J. (1977): Monografie obce Pec pod Sněžkou. MS dep. in muz. Trutnov, s.d.
Šebesta, J., Treml, V. (1976): Glacigenní a nivační modelace údolí a údolních uzávěrů Krkonoš. Opera Concortica 13, 7–44.
Tauber, G. (app. 1850): Rytina. In: Pilous, V. (1973): Strukturní mury v Krkonoších, I. Část. Opera Corcontica 10, 15–69.
Terlien, M. T. J. (1998): The determination of statistical and deterministic hydrological landslide – triggering thresholds. Environmental Geology 35(2–3), 124–130.
Tittel, F. A. (1830): Rytina. In: Pilous, V. (1973): Strukturní mury v Krkonoších, I. Část. Opera Corcontica 10, 15–69.
Treml, V., Křížek, M., Engel, M. (2010): Classification of patterned ground based on morphometry and site characteristics: A case study from the High Sudetes. Central Europe. Permafrost Perigl Proc 21, 67–77, https://doi.org/10.1002/ppp.671.
Turner, A. K., Schuster, R. L. (eds) (1996): Landslide – Investigation and mitigation. Special report 247, Transportation research board. Washington D.C., 674.
Thach, N. N., Xuan, N. T., My, N. Q., Quynh, P. V., Minh, N. D., Hoa, D. B., Bao, D. V., Dan, N. V., Thuy, T. V., Hien, N. T. (2002): Application of remote sensing and geographical information system for research and forecast of natural hazards in Hoa Binh Province. National University Hanoi, Hanoi.
Vrba, M., Spusta, V. (1975): Snow avalanche cadastre of Krkonoše (in Czech). Opera Corcontica 12, 65–90.
Vrba, M., Spusta, V. (1991): Snow avalanche cadastre of Krkonoše (in Czech). Opera Corcontica 28, 47–58.
Wieczorek, G. F., Glade, T. (2005): Climatic factors influencing occurrence of debris flows. In: Jakob, M., Hungr, O. (eds) Debris flow hazards and related phenomena. Springer, Berlin, 325–362, https://doi.org/10.1007/3-540-27129-5_14.
Záruba, Q., Mencl, V. (1982): Landslides and their control. 2nd edition. Academia, Praha, 324.
Zezere, J. L., Trigo, R. M., Trigo, I. F. (2005): Shallow and deep landslides induced by rainfall in the Lisbon region (Portugal): assessment of relationships with the North Atlantic Oscillation. Nat Hazards Earth Syst Sci 5, 331–344, https://doi.org/10.5194/nhess-5-331-2005.
Zhao, Y., Wei, F., Yang, H., Jiang, Y. (2011): Discussion on using antecedent precipitation index to supplement relative soil moisture data series. Procedia Environmental Sciences 10, 1489–1495, https://doi.org/10.1016/j.proenv.2011.09.237.

Maps of estates in Bohemia as an example of an undervalued historical geographic source – research survey and examples of utilization

Martina Tůmová*

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

ABSTRACT

As an interdisciplinary discipline, historical geography uses a diverse aggregate of historical sources, from traditional written through tangible up to pictorial, i.e., iconographic. From the pictorial sources, we can separate an individual group of cartographic materials – old maps and plans, which are one of the most important sources of information in current historical geographic research. The paper briefly summarizes the development of historical geography and then addresses the presentation of the source materials in historical geography. It draws attention to classification of the sources concerning their form and purpose and then focusses on historiographic sources of comparative and individual character. In addition to traditional comparative cartographic works (the Müller map, the military surveys and the stable cadaster), the paper emphasizes the individual maps of smaller territorial units – maps of estates – as sources that originated until the mid-19th century and as still relatively insufficiently utilized material. Through their analysis and interpretation, it is possible to obtain relatively detailed knowledge on the historical landscape of the pre-industrial period, thus filling blank spots in the research on past landscapes. The paper documents the testimonial value of these maps with an example of two aristocratic estates, those of Jilemnice and Nové Hrady, and a comparison of several specific features (depiction of the settlement, land cover, and map legends).

KEYWORDS

historical geography; cartographic sources; maps of estates; 18th and 19th centuries

Received: 26 April 2018
Accepted: 29 October 2018
Published online: 18 December 2018

Tůmová, M. (2018): Maps of estates in Bohemia as an example of an undervalued historical geographic source – research survey and examples of utilization.
AUC Geographica 53(2), 238–251
https://doi.org/10.14712/23361980.2018.22

1. Introduction

"Geography must never be separated from the history," says the famous sentence of the French lawyer and enlightener Louis-René de Caradeuc de la Chalotais from the 18th century, a sentence frequently cited by geographers and historians but in particular by historical geographers (Semotanová 2007: 197). The interconnection of chronological and spatial arrangement of phenomena is one of the main tasks of historical geography (Schenk 2011). Historical geography is referred to as an interdisciplinary discipline bordering on the natural and social sciences (Semotanová 2006) or as "geography of the past" (Matthews, Herbert 2008; Baker 2003). Similarly to the other scientific disciplines, it is undergoing a diverse development in its thematic focus. It is also undergoing transformations not only of methodologies but also in the definition of itself (for various concepts between island and continental historical geographies, see Heffernan 2009; Baker 1987; Denecke 2005; Schenk 2011; in the Czech Lands Kašpar 1990; Semotanová 2006; Semotanová, Chromý 2012; Ježek 1983; Nováček 2008; Kučera 2007). Although the roots of historical geography can be traced back to ancient times, the birth of modern historical geography is dated in the 19th century (e.g., works focussing on geographical determinism or topography, such as those of Fratišek Palacký or V. V. Tomek). In the 1930s, Henry Clifford Darby and his students transitioned research in historical geography to new concepts – from the reconstruction of landscapes in the form of description or typologization through quantitative approaches and evaluations of the phenomena and processes in the 1960s–1980s up to post-modernist methods of searching for the roots and causes of the current issues in landscapes and society. Historical geography taken this course up to now (Baker 1987, 2003). The range of studied themes has greatly extended, and the interdisciplinary approaches are applied through a variety of tools to address research issues (Chromý 2001).

Although the definition of the discipline, its history and the theoretical methodological issues of historical geographic knowledge receive great attention, the materials as sources of information and knowledge for historical geographic study stand alone. This paper therefore attempts to evaluate the material basis of historical geography, note its classification, the variety of forms, its variability over time and the necessity to search for new, insufficiently studied and unexploited resources. Its broad potential will be introduced and discussed in an example of the most widely used sources of historical geography – old maps and plans, particularly from the 18th century. In addition to traditional sources of a comparative nature, the article notes a still very little appreciated type of individual source (maps of estates), which can significantly enrich the current research on historical landscapes before the mid-19th century. The essay thus aims to answer the following questions: What types of historical sources are applied in current historical geographic research? Why are the cartographic sources significant? How can they be classified and applied? Which old maps can be used for reconstructions of historical landscapes at the various stages of development of the society (pre-industrial, industrial, and post-industrial)? What are maps of estates? From what point of view can we use them?

2. Sources of historical geography

Historical geography uses a diverse aggregation of sources. They include historical resources and resources of other related disciplines (particularly from technical and other natural sciences) (Semotanová 2006). Many papers have been written on historical geographic sources, their classification and methods of application (e.g. Hroch 1985; Butlin 1993; Morrissey, Nally, Strohmayer, Whelan 2014). Sources of information on the past can be classified according to various criteria (e.g., their form or content). Most frequently, however, the sources are simply classified according to their external features – tangible, written and illustrative, i.e., iconographic, from which the group of cartographic materials is separated (Fig. 1).

The tangible sources are most frequently utilized by archaeologists, architects, urbanists and geographers. They include all tangible artefacts (mostly archaeological findings) and traces of the past, primarily the surviving relics in the landscape. Such artefacts and traces can be applied to the study of the change in reliefs (remains of extinct sites, roads or ponds) (Pavelková, Frajer, Havlíček et al. 2016), transformed composed landscapes and areas (parks, gardens, and alleys) (Šantrůčková 2012), agricultural and industrial objects and other small architectural elements (e.g., brickworks, iron mills, glassworks, mills, crosses, and chapels) (Chodějovská, Semotanová, Šimůnek 2015).

Written materials are the main source of information for historians. They include not only official sources, sources of a narrative or literary nature, legal documents, official books, statistical documents, chronicles, diaries, correspondence, belles-letters, but also topographical and natural-scientific descriptions of territories and landscapes until the year 1800. In many cases, they supplement the other types of sources; they form textual parts of cartographic materials (the first or the second military survey or the stable cadaster) and express what cannot be captured by "pictures" (Chodějovská, Semotanová, Šimůnek 2015).

The third group of sources includes pictorial, i.e., iconographic sources such as paintings, city views, postcards, photographs, aerial photography and modern media such as film and video. These sources can provide visual information on how the landscape or

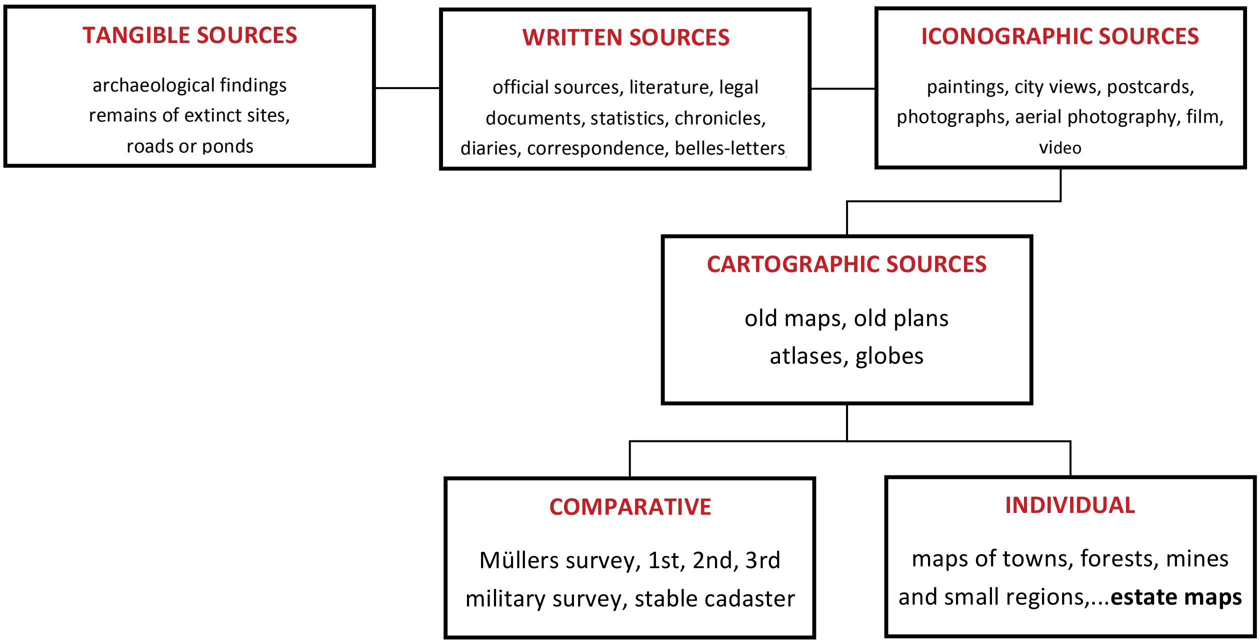

Fig. 1 Historical sources of historical geography (author's simple clasification of historical geographic sources with a focus on position of estate maps).

the particular buildings, phenomena and processes appeared in the past and how they were visually captured by their creators and actors.

Work with historical geographic sources requires an interdisciplinary approach that applies knowledge of history, geography, auxiliary sciences of history and other disciplines. The sources must be subjected to meticulous external and internal criticism and subsequent interpretation based on the source in the context of time and place. External criticism consists of describing and examining the authenticity of the source – an analysis of the date and place of origin, the authors, the material, and the techniques of production used. Internal criticism considers the content of the work in particular in relation to generally known facts (e.g., ideological and political situation, or accuracy and comparison with other parts of that time) and the person of the author (e.g., his social status, education, and nationality) (Chodějovská 2014; Semotanová 1994). Each work (either intentionally or unintentionally created) contains a large portion of subjectivity that can be detected only through a critical approach, knowledge and comparison with various sources (for criticism, see e.g. Harley 2001).

2.1 Cartographic sources, their classification and potentials for application

The iconographic sources also frequently include cartographic materials, i.e., old maps and plans (potentially also globes and atlases as collections of map compendia) classified more as fine arts between the 18th and mid-19th century (e.g. Woodward 2007), which has been one of the most applied group of sources in historical geography over the past few decades.

Old maps can be classified according to many aspects. One criterion is the scale, i.e., the size and detail of the depicted territory. According to the geographic approach (which differs from the geodetic approach), we divide the maps into maps of large scale (greater than 200 000), medium scale (1 : 200 000 to 1 : 1 000 000) and small scales (less than 1 : 1 000 000). From this point of view, maps and plans can also be distinguished; a large-scale map is considered a plan that captures such a small area that there is no distortion in terms of cartographic use. Additionally, the maps can be classified according to their content – what was the purpose of their creation and what information were they to convey with their content – for example, the displayed territory, the means of creation, the form of the record, the number of map sheets, the limitation of the map field, the time period, the concept of expression of reality or other criteria (Bláha 2017). The most useful classification of the cartographic materials for the study of historical landscapes in historical geography is the comparative and individual classification. In most cases, this division also coincides with the emergence and development of the map creation – from works originating as individual initiatives to works organized officially (by the state). The division also characterizes one of the essential features of Modern Age cartography – the absence of accurate geodetic materials for creation of maps until the mid-19th century – hence the emergence of highly individualized works of individual authors (an exception is, e.g., the first military survey) (Chodějovská, Semotanová, Šimůnek 2015). Comparative cartographic sources thus include maps that cover larger areas, particularly entire lands and states in the same depiction, scale and in particular the map key, thus allowing broader research in the vertical and horizontal plane. These works include the first maps of Jan Kryštof Müller, in particular the military surveys, the stable cadaster, collections of aerial photographs, newer maps of larger territorial units (e.g., districts) and of course the most recent state map. The individual sources are represented by maps of smaller territories – for example, parts of the lands, estates, the particular enterprises, and towns – that emerged as initiatives of individuals, usually for purely practical needs (Semotanová 2006).

Cartographic sources are applied in historical geography particularly by probes into particular areas from particular periods as the old maps were gradually found, publicized and interpreted. Interest in studying old maps started to grow in the early 18th century due to librarians and archivists who found them in collections and due to the interest of collectors who were concerned with the actual map content, its description and clarification for the study of geomorphology, archaeology, landscapes or towns (Roubík 1961).

The use of (primarily) cartographic sources, or old maps and plans, can be divided into two main directions. The first is concerned with the study and reconstruction of the cultural landscape, specifically its individual parts concerning the typology and area, for example, composed landscape complexes (Šantrůčková 2014), the individual landscape features (Havlíček 2013) or stages of development of the countryside as a whole, e.g., industrial or post-industrial (urban and rural) (Kolejka, Klimanek, Hrádek, Kirchner 2017). This direction therefore encompasses a broad period from the Middle Ages to the present, but it often reconstructs the period only through individual probes without a major effort to evaluate the long-term development. The second direction is concerned with the study of long-lasting transformations of the landscape, research of historical land use and land cover and interpretation of the motion forces that brought about these changes (e.g. Hrnčiarová, Mackovčin, Zvara et al. 2009; Bičík et al. 2015; Lipský 2000) (Chodějovská, Semotanová, Šimůnek 2015).

Interest in processing and utilizing cartographic materials as the main source of information for the reconstruction and interpretation of the historical landscape within the mentioned directions remains, however, limited to lands mapped by accurate geodetic surveys – the military surveys (except the first military survey, which was made without an accurate geodetic base but is more studied), the stable cadaster and the later period from the 19th century onwards (Kain, Baigent 1992; Hrnčiarová, Mackovčin, Zvara et al. 2009). These maps of so-called comparative character are not only topographically accurate but also very user-available and comprehensible. Due to their properties, they allow deeper comparative research concerning interpretation of long-term development trends. They include maps and plans that are made available via website map portals. These maps and plans are also useful for the application of geographic information systems and other tools of modern cartography and geoinformatics (for example, processing of 3D terrain models). The stable cadaster, a unique work created in a detailed scale for the entire territory of Cisleithania, is particularly useful for research on long-lasting changes in land use and land cover. Information from both of its record types (maps and written sources) creates the oldest layer of data in a database created for the territory of the Czech Lands. Its scope allows comparing the processes of land use on the Central European scale (Gabrovec, Petek 2002; Krausmann 2001; the LUCC Czechia database; see e.g. Bičík et al. 2012; Bičík et al. 2015). The stable cadaster served as a base for the creation of the second and third military surveys, which are commonly used for the reconstruction of landscapes at the time of transformation from the pre-industrial to industrial ages (Demek et al. 2007; Vichrová, Čada 2005; Havlíček 2013). The subsequent cadastral and topographic nationwide maps represent the core of modern geographical research. From the wide range of research, we must mention the application of comparative cartographic works in landscape ecology (Trpáková 2013), works dedicated to researching natural hazards (Raška Emmer 2014), tourism (Fialová, Steyerová, Semotanová 2015), and urban processes (Ouředníček et al. 2008–2014) or the mentioned research on land use and land cover (LUCC in Selected Regions in the World 2001–2015).

However, available cartographic materials, which can be used in research on historical landscapes, also come from an older period. Works that more-or-less depict conditions of landscapes and their individual features emerged around the 16th century. They can thus document the character of the pre-industrial landscape, i.e., the landscape until the mid-19th century, which had not yet been hit by the wave of industrialization. Knowledge of the conditions of the Modern-Age landscape is fragmentary and often varied. The literature on the history of this period provides more-or-less general characteristics of the individual smaller territories (Löw, Míchal 2003; Semotanová 2006), but landscape probes documenting various types of landscapes are sporadic. The landscape was more described (using predominantly written materials) than reconstructed. Research on significantly fragmented sources of varied character (pictorial, i.e., cartographic) is relatively difficult. It demands meticulous archive research, a critical approach and knowledge of many disciplines including linguistics and palaeography (Semotanová 2010; Chodějovská 2012; Chodějovská, Semotanová, Šimůnek 2015). Old maps of the early Modern Age (from the 16th until the 18th century) thus serve as one of the essential materials, which, together with other iconographic materials, allow the complex study of still insufficient research on the pre-industrial landscape. These materials can simultaneously deepen our knowledge of the condition and reconstruction of past landscapes and of the evaluation of long-lasting transformations of the landscape (research into long-lasting land use and land cover) or support debates on the relationships, protection and approaches to the landscape and landscape heritage until the present and into the future (Marcucci 2000; Antrop 2005; Kučera, Kučerová, Chromý 2008). To date, attention has been given only to comparative works that include maps created by Jan Kryštof Müller (a map of Moravia and Bohemia), Jan Wolfgang Wieland (a map of Silesia) (Kuchař 1959; Novák 1951) and, most importantly, the first military survey, which is the oldest map processed by the application of geographic information systems (Brůna, Uhlířová 2001; Mikšovský, Zimová 2006; Cajthaml 2012). These maps have already been subjected to analyses of their content, which provides a rich source of information on the landscape primarily due to their military purpose (Brůna, Buchta, Uhlířová 2002). The other publications addressed the rich map symbols used in this work and, more recently, with the accompanying written records, have even been editorially processed (Chodějovská 2012; Kudrnovská 1985; Kuchař 1967; Šimůnek, Antoš, Havrlant 2014; Tůmová 2015).

The period of the early Modern Age is particularly rich in the so-called individual maps. Historical descriptive essays and cartographic works focussed on cartometric analyses of selected old maps utilize individual maps from the early Modern Age (Chrást 2015; Drápela 2003; Semotanová 2001). The oldest maps of Bohemia, Moravia and Silesia created by Klaudyán, Fabricius, Arentin, Vogt, Comenius or other persons do not provide detailed or topographically accurate information on the landscape (Mucha 1992; Kuchař 1959). This role, however, can be supplemented by regional works, primarily the maps of estates.

3. Maps of estates as an undervalued historical geographic source

Maps of estates are individual maps of smaller territorial units. They originated not only in the Czech Lands but also worldwide in relatively large numbers (Bendall 1992). They were created beginning in the 16th century, primarily for economic reasons. With growing demands for higher financial gains from the land (increase in taxes) and hence transformations of the economy of estates, the registration of land ownership became crucial. As a means of depicting the situation and relatively high level of survey activities (although the available knowledge was far from completely utilized in the creation of the maps of estates), maps were thus the main tool to achieve new demands (Honl 1940). They initially emerged in the form of simple drafts and plans that have frequently not survived. They captured disputes over borders of plots and estates, parcelling of the noble estates, discussions concerning the direction of roads and collection of duties and charges from the plots or other properties, which were then resolved by the provincial court (Roubík 1961). From the 18th century, these works are preserved as large maps of entire large and important estates. Given their magnitude, the 18th century is often referred to as the century of "general maps of the landlords".

The arrival of the maps was accompanied by several persons who then affected their further development. The nobility (usually the church or aristocracy) played the role of customers and ordering parties. They thus determined their content – the area of the mapped territory, the elements to be emphasized in the map drawing (the types of enterprises) and also the embellishment and additional features in the corners of the maps. The map authors then created the work according to the specification. They were official state or manorial surveyors or other employees of the estates. The official state surveyors were trained predecessors of today's surveyors who worked for the State Registration Office. They were hired by the nobles and paid for the surveying of their estates, particularly on large estates. The manorial surveyors worked directly on the estate as did the regular staff, but the surveyors' occupation was only surveying. However, the maps were frequently composed by other employees, particularly foresters, as special assignments. They had the benefit of personal knowledge of the territory when drawing the map.

Similarly to other types of maps, the maps of estates can be divided into several types from various perspectives – particularly the form of creation, the age, the size of the mapped territory and the function and content that they primarily depicted.

In most cases, the maps of estates are manuscripts, with the exception of large and rich estates on which they were frequently printed (using lithography), particularly at the beginning of the 19th century as a result of the rapid development of mapping. Manuscript maps also include painted, or pictorial, maps, but the classification has only been artificially created for a group of works typologically bordering on a landscape painting and a map. In the past, the difference, which we observe in these works today, was not perceived (Šimůnek 2012). We can find extant maps of estates on various scales, capturing variously sized areas. They were usually made on relatively large scales. In the 18th century, overview maps of the entire estates (they provided an overall record of the business on the estates), but also maps of individual parts of the estate, the farmsteads and the accompanying grounds, farms or forest districts, became increasingly popular. With respect to processing, we can encounter maps with primarily documentary function or conversely representative function with application of further iconographic features (parergons and illustrations). Concerning the content, we encounter maps of waterworks, forest and mining maps, road maps or maps of towns and individual residence.

3.1 Study of maps of estates in Czechia and abroad

Maps of estates and similar smaller territorial units were produced by various companies throughout the world. The extant maps are deposited in many memory institutions – map collections, archives, libraries and museums. They primarily belong to resources of the state archives, particularly at the regional level. The condition of preservation of these maps is very difficult to grasp. It is clear that a large number of them are hidden in the previously mentioned institutions; in the literature, we can trace references to maps of smaller territorial units in Britain, Germany, Italy and various parts of North America, up to Jamaica (Buisseret 1996). Maps of estates in the Czech Lands are mostly preserved in state regional archives in which resources of this type belong according to the law and the archive hierarchy. They are assigned to archive collections of the manor farms, family archives of the aristocratic families, central administrations or head offices of their estates, both for individual maps and larger map sets. However, there are no clearly arranged inventories in the Czech or foreign milieus (only basic overview lexicons can be used – e.g., the World Directory of Map Collections; Roubík 1955; Semotanová, Šimůnek 2000); in most cases, they lack archival processing, so they are difficult to reach. The situation has recently been improved by digitalization of cartographic materials and creations of website portals, which open the maps to the public on the Internet (e.g., the Chartae-antiquae website portal). The only information on the condition of preserved maps from the individual territories then comes from the particular papers on the particular themes (see an example in Table 1).

Tab. 1 Representation of the maps of estates in state regional archives of Bohemia.

The State Regional Archive of Bohemia	The number of archive resources with a larger amount of extant maps of estates	The number of estates (archive funds) with extant overview maps of estates from the years 1750–1850
Prague	37	26 (18)
Litoměřice	25	9 (9)
Pilsen	27	11 (8)
Třeboň	28	23 (23)
Zámrsk	30	14 (12)

Note: Given the absence of processing of several archive collections, it can be assumed that research into the particular cartographic sources can never be considered entirely complete.

Foreign research focusses on the early Modern Age cartographic materials of an individual nature from various partial perspectives, which are summarized in the third volume of the History of Cartography dedicated to maps from the Renaissance period. With its content, the History reliably reflects the trend of the Modern Age cartographic production and the contemporary potentialities and methods of their research (Woodward 2007). The individual maps from the mediaeval period are preceded by maps of estates particularly from the Austrian territory (Stercken 2006). The most intensively examined maps from the period of the early Modern Age are maps of British estates or counties (many papers are dedicated to the maps of Christopher Saxton; Bower 2011) and smaller territorial units in Germany and Italy (Bendall 1992; Buisseret 1996; Ravenhill, Rowe 2000; Harvey 2010). However, these territorial units primarily focus on the popularizing of individual maps and their basic descriptions. Additional frequent subjects of the essays are questions concerning the emergence of the maps and their potential comparison with state cartographic production – particularly in the period from the mid-19th century, when maps were built on more-accurate geodetic foundations (Wolfart 2008; Česnulevičius 2013). The creation of maps of estates and other smaller territorial units tends to be connected with actual land ownership. The importance of maps for registration of the property is the theme of essays by David Fletcher and Laura Federzoni (Fletscher 1998; Federzoni 2013); however, the meaning of the landscape and its development remains overshadowed in these works. The potential comparison of this type of map with other materials has been noted by Steve Boyle, who compared a map of a Scottish estate from 1769 with local archaeological research data (Boyle 2009). These individual maps, however, are important not only for this knowledge. According to Thomas Horst, they are in fact the only works preceding modern mapping of the 19th century, which can be used for research on the pre-industrial landscape as a whole (Horst 2014: 375). However, they remain undervalued from this perspective and, as is clear from the mentioned examples, they are only occasionally used.

The interest in maps of estates in the Czech territory arrived in the 1950s when some of the agricultural and forestry archive collections were publicized. The first works presented only basic general descriptions of the maps, with information on their visual features and on how they might have been initiated (Roubík 1961; Honl 1960; Kalný 1990; Semotanová 2001). Over the past few years, papers on these maps have concentrated on large areas, which were mapped by significant surveyors – e.g., Šimon Podolský z Podolí, Ondřej Bernhard Klauser, Johann Glocksperger, and Ignác Hötzel (Černý 1955; Tlapák 1957; Brejnik 1997; Uhlířová 2006; Marečková 2010). These maps are sporadic, but due to the important positions of their creators, we can obtain more information on the emergence of these maps. Other partial essays address basic characteristics and archival descriptions of these documents (not only overview maps but also maps of individual aristocratic estates and farms) or their analyses concerning iconography and artistic processing (Chodějovská 2009; Šimůnek 2012; Severa 2015; Valenta 2016). This direction is also reflected in individual student works and other works that have until now concentrated only on the particular examples of maps of estates and their description (Čermáková 2011; Ourodová-Hronková 2011). The so-called polyfunctionality of this type of map – i.e., the emphasis on the multi-purpose role of a map (in addition to the documentary role, particularly the representative role through iconographic materials – e.g., the early Modern Age maps were frequently accompanied by views of important towns) is one of the main directions of research on individual cartographic works (Šimůnek 2012).

Maps of estates are maps of large scales. Given the level of cartographic processing, they are preserved as relatively detailed but at the same time schematic illustrations of smaller territorial units with economic and administrative data. Their aim was to register the aristocrats' property and represent it. They also inform about the thinking and progress of the given society through artistic and technical elaborations. They have rich potential for historical geographic research. Information that they contain can be divided into two levels – the content of the map drawing and the marginal notes.

The content of the map drawings reflects the condition and visual features of the cultural landscape. With the help of these drawings, it is possible to reconstruct the landscape at the time when the map was produced. By comparing them with maps from different periods, we can monitor developments and evaluate the various types of transformations and their dynamism. Maps of estates can capture different types of landscapes – their chronological aspects (pre-industrial and industrial landscapes) or concerning their character (rural, urban). Similarly to the maps created from the mid-19th century onwards, the content of these maps is very suitable for reconstruction of the land use and land cover. They allow monitoring the depiction and development of the landscape structure of the estates – the individual partial landscape features, road networks, watercourses, business objects or other small architectural features. Their added value is the use of regional toponyms – the frequently extinct names of the given regions.

Marginal notes and structural elements – map decorations, other supplementary data (tables, legends, graphic illustrations of the scales and map roses) can greatly enrich the content of the map. They reflect the given time and its society – the authors, customers or users of the maps. They allow us to gain information on the given estate concerning its functioning, its main dominants, important sites, the economy and important activities that provided the local population with sustenance. The map decorations also represented the estate and in particular its owners. Thus, they do not always reflect the reality.

3.2 Utilization of maps of estates on the example of model territories of Jilemnice and Nové Hrady

As mentioned above, estate maps of various types are preserved in relatively large numbers in archives and other memory institutions (not only in Czechia). Therefore, working with these sources cannot be accomplished without an initial, thorough archival search – finding, gathering and selecting the map of the estate according to the types and the predetermined objectives of the research. Within the framework of the dissertation research, 261 archive collections deposited in the state regional archives in Bohemia were processed. These collections were selected based on a search of available tools – the archive fund database of the Archival Administration of the Ministry of the Interior, the internal databases of individual state regional archives made available to researchers, the Guide to Archival Funds available for all state regional archives from the 1950s to the 1970s, and the previous Roubík map listing. Based on the study of the individual inventories of these archive funds, a typological definition (only overview maps of estates), the time perspective (the century from 1750 to 1850 because of the greatest preservation of the estate maps), and a database of the main archive collections containing maps of the estate from the territory of Bohemia (the number of important collections that include maps of estates is given in Table 1).

The uniqueness of the maps of estates, their differences and similarities can be documented by mutual comparison of maps of this type from various territories. Reflecting the author's main research interest, it is possible to present two important works from the area of eastern Bohemia. Significant and very similar but to some extent also very different maps can be found on the estates of Jilemnice and Nové Hrady. Both of these territories were mapped on the initiatives of the aristocrats in the second half of the 18th century, when manorial land surveying become important in Bohemia. Both estates were at that time in the hands of important and wealthy families (the Harrach and Harbuval Chamaré families), and the role of the maps was not only documentary but also largely representative. The main difference is the position of both territories on different latitudes – in the hilly regions around Jilemnice and the lowland of Nové Hrady (Fig. 2).

The Jilemnice estate spreads across the Giant Mountains and their foothills, historically situated in the Bydžov region. First mentioned in the 14th century, the estate had mediaeval roots. Throughout various modifications – associated with Branná yard – it remained within its borders until the land reform in the mid-19th century. The estate was attractive because it was owned by important noble families, primarily the Harrach from the end of the 17th century; the data on it survived many modifications and various types of archival materials (Severa, Kamenická 2016). The main reason for the interest was its

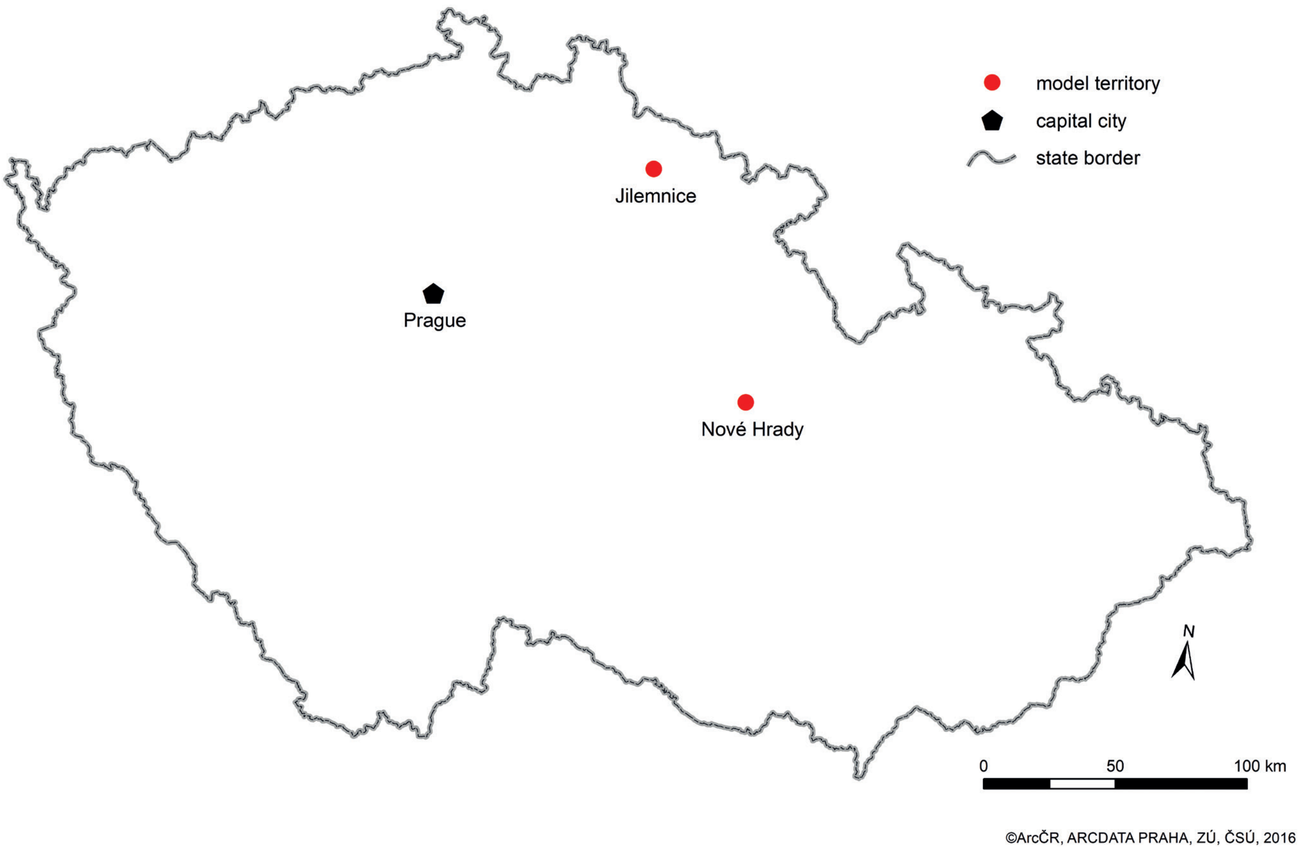

Fig. 2 Location of the Jilemnice and Nové Hrady estates on the territory of the Czechia. Author: Iveta Lžíčařová.

large businesses – the glassworks and textile production, which spread particularly in the hilly regions. The Giant Mountains was the highest mountain range in Bohemia, and the borderland was also a mysterious region that attracted many authors, including cartographers and surveyors. From the several extant cartographic works, we must focus on the map of the Jilemnice estate, the Branná entailed estate and the Žďár farmstead from 1765. This map is an overview of the entire estate, with dimensions of 199.5 × 355 cm, that was created by the Jilemnice estate administrator Jan Antonín Graupar. The map is a magnificent work not only for its size but also particularly for the artistic design and the information content. It provides a detailed overview of the estate landscape, its structure and the individual landscape features. The decoration provides information on the everyday life of the population, particularly their livelihoods.

Similarly to the Jilemnice estate, the Nové Hrady estate has a long history. It was located in the historical Chrudim region and developed from a village called Boží Dům and the accompanying Nový Hrad Gothic castle. Before the extinction of the patrimonial administration system, it incorporated a small town (Proseč) and 22 villages. The estate, located in a typically lowland area, prospered particularly due to the linen production, which spread across the estate in the mid-18th century under its owner, the Harbuval Chamaré family. At least two cartographic works have survived from the activity of these aristocrats – a map of the Potštejn estate from 1772 (which was already studied in the 1950s) and a map of the Nové Hrady estate from 1775, which remains a valuable but inadequately recognized gem of early Modern Age cartography. The latter depicts the individual locations, types of land cover and parcels and provides a detailed list of the toponyms (Severa, Kamenická 2016).

By comparing these two important maps, we can obtain an overview of not only the landscapes but also the societies of both eastern Bohemian estates in the second half of the 18th century. We can also gain an awareness of what information we can find on maps of smaller territorial units of this type and how they can be exploited. We can compare these two maps on both mentioned aspects – the content of the map drawings and the marginal notes – and focus in detail on the individual features. Table 2 presents a simple analysis of map legends, basis identification and statistical information on maps and their content (chosen by the random selection method).

From the perspective of the first content level (see above), we can focus on reconstruction of the manor landscape. Both maps capture the manors in an oblique view. They capture the individual houses in the built-up areas and the accompanying plots. The buildings are not depicted accurately; on both maps, we can observe considerable plainness in the drawings of the houses, usually simple symbols of the

Tab. 2 East Bohemian maps of estates – two model probes.

Information on the map	Map of the Jilemnice estate	Map of the Nové Hrady estate
Name of the map	Mappa der Hoch Reichs gräfflich harrachichen Herrschaft Starkenbach und Majorat Branna dann Gueth Zdiar	Mappa der in Königreich Böheim Chrudimer Creißes situirten in Anno 1775 geometrice ausgemessen Herrschaft Neuschloß,…
Author	Jan Antonín Graupar	Franz Xaver Thiel
Year of issue	1765	1775
Scale	approx. 1 : 12 000	approx. 1 : 5 313
Size of the map	199.5 × 355 cm	220 × 135 cm
Estate owners – the aristocrats	Harrach	Harbuval and Chamaré
The number of towns and villages on the estate	34	39
Depiction of the manors	oblique view	oblique view
Legend	villages, farmsteads, glassworks, ironworks and iron mills, textile manufactures, breweries, other manorial buildings, ponds, springs, manor forests, manor meadows, other map symbols of the individual features	types of plots, boundary points (trees, stones), other map symbols of the various features
Types of grounds	church, aristocratic, serf, municipal, defunct courts	dominican, rustic, farmsteads, forest districts, church
Map embellishment	city views, parerga (business facilities, surveying equipment, Sněžka, hunting, family coat of arms), decorative borders	parergon (the family coat of arms)
Point of interest	chalet economy	toponyms, plot numbering

building with a reddish-brown roof. However, important buildings – churches, palaces, town halls or important business objects (farmsteads with the typical square layout) – are distinguished in both maps. By comparison with newer maps, we can observe the layout and territorial development of the given manor. The comparison is facilitated by the areas of public spaces – the village and town squares, which practically did not change layouts in the history of the manor. The specific feature of the Nové Hrady estate is the accurate segmentation of individual plots not only in the immediate vicinity of the manors (Fig. 3) but across the entire estate. All plots are numbered, which might suggest the existence of a written addendum to the map that has not survived. The map of the Jilemnice estate contains many decorations including several vedute – town views that provide closer details of the dominants of individual manors. Concerning the depiction of the manor structure, it is necessary to mention the Giant Mountains chalet economy, whose reconstruction is facilitated by the map of the estate. In a simple drawing, it depicts all chalets that existed at the time of the map's conception. It also refers to them in the legend, in which they are described in more detail (Severa, Kamenická 2016; Valenta 2016; maps of the Jilemnice and Nové Hrady estates).

Meticulous depiction of forest, field and grassy areas in the maps of estates can be used in research on land use and land cover. Both maps capture various types of land cover: fields, meadows, forests and, in the case of the Jilemnice estate, the clearings and forest stands in the hilly region of the Giant Mountains. These areas represent symbols of stylized trees, slightly varying according to the type and density of the vegetation (Fig. 4).

However, the legends of the maps of estates offer not only these map symbols. Due to the practical but also representative use of the map, they include many other data. The map of the Jilemnice estate contains a legend divided into four cartouches, three of which contain only the explanations of the numerical and literal references given on the map itself (settlement, farmsteads, ponds, springs, huts, glassworks, breweries, textile manufactory and others). The branding key then adds the last part of the legend, which, in addition to the coloured markings of the boundaries of the estate of individual plots within the estate, includes a large number of individual buildings – primarily farm buildings (e.g., mills, saws, oil plants, bleachers, and barns). The map of the Nové Hrady estate is not so detailed in this respect; we find a sign for mills, roads, bridges and ponds. In contrast to the Jilemnice map, however, it distinguishes a large number of boundary stones and trees (mapped primarily on the western border with the Litomyšl estate and the eastern border of the Rychmburk estate).

The Jilemnice estate map was clearly used not only from the point of view of the map symbols and the extent of the map legends but also for the very careful recording of all of the economic objects within the aristocratic manor. The map also incorporates a large number of illustrations and decorative elements (so-called parergon) that serve to represent the owners and the overall promotion of the prosperous territory of the Giant Mountains. The business activity on the estate is documented by parerga pastorage,

Fig. 3 View of the town of Proseč on the map of the Nové Hrady estate from 1775.

Fig. 4 The Giant Mountains region with a vicarage and pub in Rokytnice on the map of the Jilemnice estate from 1765.

wood flooding, textile manufactory, glassworks, mining, which are complemented by several landscape sceneries. The map of the Nové Hrady estate, except for a coat of arms, does not contain any decorative elements. By the careful view of individual plots and border points, it demonstrates its registration character and, most importantly, its practical function (Severa, Kamenická 2016; maps of the Jilemnice and Nové Hrady estates).

4. Conclusion

Historical geography uses a diverse aggregation of historical sources that can be classified as tangible, written and pictorial, i.e., iconographic. The most important resource is a group of cartographic materials – old maps and plans of a comparative or individual nature that have an indisputable importance in historical geographic research. They facilitated the reconstruction of the landscape over a certain period in the past. By mutual comparison, they facilitate evaluation of the developments and transformations in the landscape. Through their artistic elaboration, they document the level of cartographic knowledge and technological advancement, also exhibiting the relationships within the society. In addition to the known comparable cartographic works (the Müller map, the military surveys and the stable cadaster), there exist individual maps that can significantly move forward the research of historic landscapes deeper into the past. Blank spots in the research on the pre-industrial landscape (until the mid-19th century) can be filled by surveys of individual maps of smaller territorial units – particularly maps of estates, which bring detailed information on smaller regions that mostly cannot be obtained from other historical sources. Although they did not survive for all estates or territories of similar administrative character, their great documentary value is undeniable at least for their age. Based on the example of the model territories of the eastern Bohemian estates of Jilemnice and Nové Hrady, it is clear that maps of estates document the past landscapes and lives of people by virtue of their content and marginal information. Via a comparison, we can obtain valuable knowledge on the manors of the estates, the areas and types of land cover at various altitudes, and the business focusses and sources of the people's livelihood. Each of the maps also has its own specific features – the Jilemnice map illustrates the chalet economy, whereas the Nové Hrady map has very rich toponyms. The maps of estates thus represent a significant contribution to the regional geography or regional history. However, they can also serve the needs of territorial planning or reclamation and revitalization of the landscape. They should therefore not be neglected in any historical geographic research.

Acknowledgements

The paper was written with support of the Charles University Grant Agency project No. 580816 "The research on early modern cartographical manuscripts" and the Czech Science Foundation project No. P410/12/G113 "Research Centre for Historical Geography".

References

Antrop, M. (2005): Why landscapes of the past are important for the future. Landscape and Urban Planning 70, 21–34, https://doi.org/10.1016/j.landurbplan.2003.10.002.
Baker, A. R. H. (1987): Některé aspekty britské historické geografie v letech 1966–1986. Historická geografie 26, 25–43.
Baker, A. R. H. (2003): Geography and History. Bridging and Divide. Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9780511615818.
Bendall, S. (1992): Maps, Land and Society: A History with a Carto-Bibliography of Cambridgeshire Estate Maps, 1600–1836. Cambridge, Cambridge University Press.
Bičík, I. a kol. (2012): Vývoj využití ploch v Česku. Praha, ČGS.
Bičík, I. et al. (2015): Land Use Changes in the Czech Republic 1845–2010. Cham, Springer, https://doi.org/10.1007/978-3-319-17671-0.
Bláha, J. D. (2017): Vybrané okruhy z geografické kartografie. Ústí nad Labem, Univerzita J. E. Purkyně v Ústí nad Labem.
Bower, D. (2011): Saxton's Maps of England and Wales: The Accuracy of Anglia and Britannia and Their Relationship to Each Other and to the County Maps. Imago Mundi 63(2), 180–200, https://doi.org/10.1080/03085694.2011.568704
Boyle, S. (2009): Mapping Landscapes of the Improvement Period: Surveys of North Lochtayside, 1769 and 2000. Scottish Geographical Journal 125(1), 43–60, https://doi.org/10.1080/14702540902873915.
Brejník, M. (1997): Proměny krajiny sídel v průběhu jednoho století (1740–1840). Historická geografie, 29, 21–50.
Brůna, V., Uhlířová, L. (2001): Staré mapy v prostředí GIS. Studia Oecologica 12, 107–109.
Brůna, V., Buchta, I., Uhlířová, L. (2002): Identifikace historické sítě prvků ekologické stability krajiny na mapách vojenských mapování. Laboratoř geoinformatiky UJEP, Ústí nad Labem.
Buisseret, D. (ed.) (1996): Rural Images. Estate Maps in the Old and New Worlds. Chicago, Univesity of Chicago Press.
Butlin, R. A. (1993): Historical Geography. Through the Gates of Space and Time. London, Arnold.
Cajthaml, J. (2012): Tvorba souvislé mapy I. vojenského mapování Habsburské monarchie – testovací oblast Ústecký kraj. Geodetický a kartografický obzor 59(101), 212-219.
Čermáková, J. (2011): Malovaná mapa frýdlantského panství od Samuela Globice z Bučína. Bakalářská

práce. Liberec, Technická univerzita v Liberci, Fakulta přírodovědně-humanitní a pedagogická, katedra historie.
Černý, V. (1955): Mapy velkostatků v zemědělsko-lesnických archivech. Kartografický přehled 9(2), 91–93.
Česnulevičius, A. (2013): M. K. Radvila map in Middle Europe mapping context of the XVII century. Geografija 49(2), 145–153.
Demek, J., Havlíček, M., Mackovčin, P., Slavík, P. (2007): Landscape changes in the Czech Republic 1763–2005: An Assessment Based on Historical Maps and GIS Technology. Nova Acta Leopoldina NF 94(349), 187–200.
Denecke, D. (2005): Wege der Historischen Geographie und Kulturlandschaftforschung. Ausgewählte Beiträge. Stuttgart, Franz Steiner Verlag.
Drápela, M. (2003): Komenského mapa Moravy z roku 1624 – nejstarší vydání. Historická geografie 32, 21–30.
Federzoni, L. (2013): Egnazio Danti's map of the Palata estate and the lands of the sixteenth-century Bolognese aristocracy: practical and symbolic dimensions of estate surveying in the Renaissance. Journal of Historical Geography 42, 110–123, https://doi.org/10.1016/j.jhg.2013.07.001.
Fialová, D., Steyerová, M., Semotanová, E. (2015): Vltavské ostrovy v Praze. Praha, ČGS.
Fletscher, D. (1998): Map or Terrier? The Example of Christ Church, Oxford, Estate Management, 1600–1840. Transactions of the Institute of British Geographers 23(2), 21–237.
Gabrovec, M., Petek, R. (2002): A methodology for acessing the change in Land Use in Slovenia from the viewpoint of sustanaible development. In: Bičík, I., Chromý, P., Jančák, V., Janů, H. (eds): Land use/Land cover changes in the period of globalization. Proceedings of the IGU-LUCC international konference, Prague 2001, 168–179.
Havlíček, M. (2013): Význam starých map pro studium změn krajiny v okrese Hodonín. Disertační práce, Přírodovědecká fakulta, Brno, Masarykova univerzita.
Harley, J. B. (2001): The New Nature of Maps. Baltimore, The Johns Hopkins University Press.
Harvey, P. D. A. (2010): Manors and Maps in Rural England, from the Tenth Century to the Seventeenth. Farnham, Ashgate, Variorum Collected Studies Series.
Heffernan, M. (2009): Historical Geography. In: Johnston, R. J., Gregory, D., Pratt, G., Watts, M., Whatmore, S. (eds): The Dictionary of Human Geography. Oxford, Wiley-Blackwell, 332–335.
Honl, I. (1960): K rozvoji map velkostatků v 18. a na začátku 19. století. Archivní časopis 10, 221–227.
Horst, T. (2014): Kartographie und Grundstückseigentum in der Frühen Neuzeit. Zeitschrift für Geodäsie, Geoinformation und Landmanagement 139(6), 369–376.
Hrnčiarová, T., Mackovčin, P., Zvara, I. a kol. (2010): Atlas krajiny České republiky. Praha-Průhonice, MŽP – VÚKOZ.
Hroch, M. (1985): Úvod do studia dějepisu. Praha, Státní pedagogické nakladatelství.
Chodějovská, E. (2009): Příspěvek k ikonografii hospodářských budov na schlikovských panstvích Jičínska. In: Navrátil, I. (ed.): Šlechtické rody a jejich sídla v Českém ráji: sborník referátů z vědecké konference konané ve dnech 24.–25. dubna v Turnově. Z Českého ráje a Podkrkonoší, Suplementum 13, Semily, Státní oblastní archiv v Litoměřicích pro Pekařovu společnost Českého ráje v Turnově, 332–346.
Chodějovská, E. (2012): I. vojenské mapování jako pramen pro poznání krajiny 2. poloviny 18. století, in: E. Chodějovská – R. Šimůnek (red.): Krajina jako historické jeviště: k poctě Evy Semotanové, Praha, 159–184.
Chodějovská, E. (2012): Studying the Early Modern Landscape in the Czech Republic. Historická geografie 38(1), 63–99.
Chodějovská, E. (2014): Historiografie a její prameny. Geografické rozhledy 24(1), 14–15.
Chodějovská, E., Semotanová, E., Šimůnek, R. (2015): Historické krajiny Čech: Třeboňsko – Broumovsko – Praha. Praha, Historický ústav.
Chrást, J. (2015): O starých mapách českých zemí. Fabriciova a Komenského mapa Moravy. Praha, Academia.
Chromý, P. (2001): Výzvy pro českou historickou geografii. Historická geografie 31, 87–108.
Jeleček, L. (1983): Nástin vývoje československé historické geografie. Historická geografie 22, 11–51.
Kain, R. J. P., Baigent, E. (1992): The Cadastral Map in the Service of the State: A History of Property Mapping. Chicago, The University of Chicago Press.
Kalný, A. (1990): Mapy a plány třeboňského panství a velkostatku. (Rozbor fondového souboru). Sborník archivních prací 40(2), 385–395.
Kolejka, J., Klimanek, M., Hrádek, M., Kirchner, K. (2017): Czech Post-industrial Landscapes in the Border Zone with Austria: Identification, Typology nad Value. Mitteilungen der Österreichischen Geographischen Gesellschaft 159(1), 221–242.
Kašpar, J. (1990): Vybrané kapitoly z historické geografie českých zemí a nauky o mapách. Praha, SPN.
Krausmann, F. (2001): Land use and industrial modernization: an empirical analysis of human influence on the functioning of ecosystems in Austria 1830–1995. Land Use Policy 18(1), 17–26, https://doi.org/10.1016/S0264-8377(00)00042-9.
Kučera, Z. (2007): Historická geografie mezi geografií a historiografií. Historická geografie 34, 9–19.
Kučera, Z., Kučerová-Kuldová, S., Chromý, P. (2008): Landscape heritage between areal preservation and areal development – the case of Czechia. Geographia Polonica 81(2), 5–23.
Kudrnovská, O. (1985): Josefské mapování českých zemí a jeho topografický popis, Historické geografie 24, 55–81.
Kuchař, K. (1959): Mapy českých zemí do poloviny 18. století. Vývoj mapového zobrazení území Československé republiky I. Praha, Ústřední správa geodézie a kartografie.
Kuchař, K. (1967): Mapové prameny ke geografii ČSR, Acta Universitatis Carolinae Geographica 2(1), 57–97.
Land Use/Cover Changes in Selected Regions in the World – Volume I–XI (2001–2015). IGU Commission on LUCC.
Lipský, Z. (2000): Sledování změn v kulturní krajině. Kostelec nad Černými lesy, ÚAE LF ČZU.
Loiseaux, O. (ed.) (2000): World directory of Map collections. München, K.G. Saur, https://doi.org/10.1515/9783110950434.
Löw, J., Míchal, I. (2003): Krajinný ráz. Kostelec nad Černými lesy, Lesnická práce s.r.o.
Mapová sbírka Chartae-antiquae, http://www.chartae-antiquae.cz/cs/ (2. 4. 2018).

Marcucci, D. J. (2000): Landscape history as a planning tool. Lanscape and Urban planning 40(1–2), 67–81, https://doi.org/10.1016/S0169-2046(00)00054-2.
Marečková, J. (2010): Soubor rukopisných map plzeňského panství v období raabizace koncem 18. století. Historická geografie 36(2), 197–243.
Mathews, J. A., Herbert, D. (2008): Geography: a very short introduction. Oxford, Oxford University Press, https://doi.org/10.1093/actrade/9780199211289.001.0001.
Mikšovský, M., Zimová, R. (2006): Müllerovo mapování a první vojenské mapování českých zemí (se zřetelem k digitalizaci a centrální evidenci map v Česku), in: Šimůnek, R. (ed.): Historická krajina a mapové bohatství Česka. Historická geografie – Supplementum I, 14-25.
Morrissey J., Nally, D., Strohmayer, U., Whelan, Y. (2014): Key Concepts in Historical Geography. Los Angeles, Sage Publications, https://doi.org/10.4135/9781473920538.
Mucha, L. (1992): Mikuláš Klaudyán a jeho mapa Čech z roku 1518. Geografické rozhledy 2–3, 33–34.
Nováček, A. (2008): Historische Geographie in Tschechien. Ein Bericht über den aktuellen Stand der Forschung. Siedlungsforschung. Archäologie – Geschichte – Geographie 26, 293–304.
Novák, V. (1951): Wielandovy mapy Slezska. Slezský sborník 49, 289–299, 284–503.
Ourodová-Hronková, L. (2011): Mapa třeboňského panství. In: Gaži, M. (red.): Rožmberkové. Rod českých velmožů a jeho cesta dějinami. Národní památkový ústav, Praha, 154–155.
Ouředníček, M. a kol. (2008–2014): Suburbanizace.cz. Univerzita Karlova v Praze, Přírodovědecká fakulta, katedra sociální geografie a regionálního rozvoje, Praha, Urrlab.
Pavelková, R., Frajer, J., Havlíček, M. et al. (2016): Historical ponds of the Czech Republic: an example of the interpretation of historic maps. Journal of Maps 12, Supp1, 551–559, https://doi.org/10.1080/17445647.2016.1203830.
Raška, P., Emmer, A. (2014): The 1916 catastrophic flood following the Bílá Desná dam failure: The role of historical data sources in the reconstruction of its geomorphologic and landscape effects. Geomorphology 226, 135–147, https://doi.org/10.1016/j.geomorph.2014.08.002.
Ravenhill, M. R., Rowe, M. M. (2000): Early Devon Maps: Map of Lands and Estates in Devon before 1700. Exeter, Friends of Devon's Archives.
Roubík, F. (1955): Soupis map českých zemí II. Praha, ČSAV.
Roubík, F. (1961): Rukopisné mapy od 16. do poloviny 18. století ve Státním ústředním archivu v Praze. Sborník archivních prací 11, 138–186.
Semotanová, E. (1994): Kartografie v historické práci. Praha, Historický ústav.
Semotanová, E., Šimůnek, R. (2000): Lexikon mapových archivů a sbírek České republiky. Praha, Historický ústav.
Semotanová, E. (2001): Mapy Čech, Moravy a Slezska v zrcadle staletí. Praha, Libri.
Semotanová, E. (2006): Historická geografie českých zemí. Praha, Historický ústav.
Semotanová, E. (2007): Dějiny, současnost a perspektivy rekonstrukčních map. Historická geografie 34, 197–215.
Semotanová, E. (2010): Raně novověká krajina českých zemí. In: Bůžek, V. a kol.: Společnost českých zemí v raném novověku, Praha, 19–53.
Semotanová, E., Chromý, P. (2012): Development and current trends of czech historical geography. Historická geografie 38(1), 9–34.
Severa, M. (2015): Zemský měřič František Xaver Preitsch a jeho mapy městeček a vsí rychmburského panství. In: Kudrnová, J., Polehla, P., Severa, M., Vojtíšková, J. (eds.): Ad honorem VN: k 65. narozeninám Věry Němečkové. Pavel Mervart, Hradec Králové, 247–282.
Severa, M., Kamenická, M. (2016): Východočeská panství na mapách. Katalog výstavy. Zámrsk, Státní oblastní archiv v Zámrsku.
Schenk, W. (2004): „Städtelandschaft" als Begriff in der Historischen Geographie und Anthropogeographie. In: Gräf, H. T., Keller, K. (Hrsg.): Städtelandschaft – Réseau Urbain – Urban Network. Städte im regionalen Kontext in Spätmittleralter und früher Neuzeit. Städteforschung A/62, Böhlau Verlag, Köln, 25–46.
Schenk, W. (2011): Historische Geographie. Darmstadt, WBG.
Stercken, M. (2006): Kartographie von Herrschaft im Mittelalter. Rheinische Vierteljährblätter 70, 134–154.
Šantrůčková, M. (2012): The Principles and Development of Landscape Parks in the Czech Republic and Their Study from the Perspective of Historical Geography. Historická geografie 38(1), 99–118.
Šantrůčková, M. (2014): Krajinářská tvorba Jana Rudolfa Černína: vznik a vývoj parků v Krásném Dvoře, Jemčině, Petrohradě a Chudenicích. Praha, Karolinum.
Šimůnek, R. (2012): Velkoplošný obraz krajiny jako reprezentativní dekorace interiéru: mapa třeboňského panství z roku 1684. In: Chodějovská, E., Šimůnek, R., Adam, D. (eds.): Krajina jako historické jeviště: k poctě Evy Semotanové. Historický ústav, Praha, 81–104.
Šimůnek, R., Antoš, F., Havrlant, J. (2014): Třeboňsko na I. vojenském mapování, Historická geografie 40(2), 153–204.
Tlapák, J. (1957): Podblanicko na starých mapách velkostatků. Benešov, Sborník vlastivědných prací z Podblanicka, 117–127.
Trpáková, I. (2013): Krajina ve světle starých pramenů. Kostelec nad Černými lesy, Lesnická práce s.r.o.
Tůmová, M. (2015): Město Opočno v prvním vojenském mapování. Východočeské listy historické 34, 37-52.
Uhlířová, L. (2006): Kulturní krajina Česka na mapách panství z první poloviny 18. století – na příkladech panství Kosmonosy a Tloskov. Historická geografie, Supplementum I, 78–82.
Valenta, A. (2016): Grauparova mapa velkostatku Jilemnice. Vrchlabí, Správa Krkonošského národního parku.
Vichrová, M., Čada, V. (2005): Kartografické vyjadřovací prostředky a interpretace obsahu map druhého vojenského mapování. Historické mapy. Zborník referátov z vedeckej konferencie, Bratislava, 246–255.
Wolfart, P. D. (2008): Mapping the early modern state: the work of Ignaz Ambros Amman, 1782–1812. Journal of Historical Geography 34, 1–23, https://doi.org/10.1016/j.jhg.2007.07.021.
Woodward, D. (ed.) (2007): The History of Cartography (3, 1–2). Cartography in the European Renaissance. Chicago, University of Chicago Press.

Sources

State Regional Archive in Zámrsk, archival fund Nové Hrady Manor, map of the estate from 1775, non-processed resource. Scan from State Regional Archive in Zámrsk.

State Regional Archive in Zámrsk, archival fund The Harrach Central Administration, map No. 2. Scan from VÚGTK, Zdiby.