ACTA UNIVERSITATIS CAROLINAE

AUC GEOGRAPHICA

54
2/2019

CHARLES UNIVERSITY • KAROLINUM PRESS

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

Territorial Impact Assessment – European context and the case of Czechia

Štěpán Nosek*

Charles University, Faculty of Science, Department of Social Geography and Regional Development, Czechia
* Corresponding author: stepan.nosek@gmail.com

ABSTRACT
The aim of this paper is to critically review recent EU level debates on territorial impact assessment, which serves as a tool to improve the understanding of uneven territorial impacts of the EU sectoral policies. The paper also seeks to elicit (1) which European countries employ territorial impact assessment when designing various national policies and (2) how this tool is used in different governance environments. Particular attention is paid to the case of Czechia. The paper elaborates upon the state-of-the-art tools used on a national level and analyses the motivation of actors on regional and local levels to use such tools in their decision-making process. The research shows that EU member states that employ this tool can be categorized into four groups, for instance, depending on whether they enrich other impact assessments tools by territorial aspects, or whether they focus coherence of regional development strategies and sectoral strategies. Territorial impact assessment is not implemented in Czechia, although recent years have witnessed an upsurge of interest concerning territorial aspects of investments.

KEYWORDS
Territirorial Impact Assessment; European Union; Czechia

Received: 27 April 2019
Accepted: 4 July 2019
Published online: 20 August 2019

Nosek, Š. (2019): Territorial Impact Assessment – European context and the case of Czechia. AUC Geographica 54(2), 117–128
https://doi.org/10.14712/23361980.2019.1

1. Introduction

For a long time, there has been a quest to understand spatial or territorial consequences of public policies (Nijkamp and Van Pelt 1989). Some public policies are focused directly on increasing the development potential of certain regions (regional policy). Other public policies are not directly focused on regions, but their impact is often greater compared to explicit regional policies. For decades, there have been attempts to develop tools which would enable to assess or to measure a spatial impact of public policies.

In the 1970s, first tools emerged in the USA in order to assess the impact of policies or projects with a focus on environmental and economic impacts (e.g. Caldwell 1988; Bond and Pope 2012; Francis 1975). However, geographers have been traditionally interested in more complex tools assessing all aspects of territorial development. Discussions at the EU level regarding the territorial dimension and territorial cohesion paved the way for more complex tools enabling a holistic approach in the evaluation of impact on territorial development (e.g. Medeiros 2017).

According to Fischer et al. (2015), territorial impact assessment (TIA) is a distinctly EU policy assessment closely connected to the EU territorial cohesion agenda and "does not make much sense in, e.g. the USA where people are used to move to areas of greater economic opportunity" (p. 9). The Territorial Agenda of the European Union 2020 (European Commission 2011) states that territorial cohesion "enables equal opportunities for citizens and enterprises, wherever they are located" (p. 3). Following the addition of territorial cohesion among strategic goals of the European Union in 2009, it seems to be even more important to evaluate potentially uneven territorial impacts of particular projects, strategies, or policies.

Territorial Impact Assessment is a tool which appeared in a practical debate in 1989 with the launch of the European Spatial Development Perspective (ESDP; Zonneveld and Waterhout 2009).

The aims of this paper are twofold: firstly, to critically review recent developments regarding impact assessment tools at the EU level and, secondly, to analyse awareness about these tools and the current state of their implementation in one EU member state – in Czechia. Czechia was selected for a case study due to the fact that the author of this paper has followed the discussion and subsequent application of territorial impact assessment in this country over the past five years. Therefore, the paper seeks primarily to answer these questions:

(Q1) How do European countries deal with the challenge of evaluating territorial impacts of their policies?

(Q2) What is the current state of Territorial Impact Assessment implementation in Czechia?

The paper unfolds as follows: Chapter 2 presents an overview of different types of impact assessment with special attention to TIA. Chapter 3 analyses the current state of the art in selected European countries in using TIA or similar tools. Chapter 4 focuses on Czechia and explains how TIA or similar tools are employed. Chapter 5 summarizes the main conclusions and presents several suggestions for further development of TIA in the Czech context.

From a methodological standpoint, the paper is based mostly on content analysis of available documents dealing with the topic of policy evaluation and (territorial) impact assessment in particular countries. This source of information has been complemented with questionnaires sent to academics and public officials in several countries. The ambition of the study was to elaborate on the current situation in TIA use in all EU countries. Thus, respondents were asked an open question, i.e. to describe use of TIA in particular EU member states. Combination of those two sources (content analysis and questionnaires) enabled to divide EU member states into four categories (see Chapter 3 and Table 2). Nonetheless, relevant information has not been found in case of particular countries.

The analysis concerning Czechia is based on this author's experience gained at the Czech Ministry of Regional Development and at the International Advisory Centre for Municipalities, which implemented impact assessment techniques in the city of Litoměřice.

2. The Role of Impact Assessment Tools in Policy Coordination

It is only natural that since the addition of territorial cohesion among the European Union's goals, it has become even more pressing to assess territorial impacts of sectoral policies that do not explicitly follow regional policy goals – no matter what the goals of particular regional policies in different national contexts are, the Territorial Impact Assessment methodology has been developed as a tool to measure potential territorial effects of sectoral policies (Golobič and Marot 2011).

TIA might be understood as a new addition to a vast array of impact assessment tools that have been applied since the 1970s. Impact assessment, which is mostly used and embedded in legislation of many countries all over the world focuses primarily on economic and environmental dimensions (see Table 1). However, proper evaluation of impacts on territorial development should include other dimensions as well. Stutz and Warf (2012) state that, "in conventional usage development is a synonym for economic growth" (p. 268). Nevertheless, territorial development should be understood in a more complex way, Medeiros (2017) concludes that territorial development should follow five main goals: economic competitiveness, social cohesion, environmental

Tab. 1 Types of Impact Assessment.

	Impact assessment focused on			Complex Impact Assessments
	Economic dimension	Environmental dimension	Social dimension	
Examples	Cost-benefit analysis Regulatory Impact Assessment	Environmental Impact Assessment Strategic Impact Assessment	Poverty Impact Assessment Health Impact Assessment	Territorial Impact Assessment Impact Assessment (European Union)

Source: Author

sustainability, sound processes of territorial governance, and efficient processes of spatial planning or territorial articulation. TIA is then supposed to encompass all those dimensions.

2.1 Impact assessment – definition and origins

The purpose of impact assessment is to assess potential impacts of certain actions before they are implemented (Hayes 2017). Impact assessment is thus supposed to evaluate what might happen as a direct consequence of a particular project, strategy, or a policy. Today, there are more than 40 types of IA discussed in literature (Morrison-Saunders et al. 2014). The most frequently used ones are Environmental Impact Assessments, Strategic Environmental Impact Assessment, or Regulatory Impact Assessments, which are embedded in law in most of countries.

The first impact assessments focused on environmental impacts: the first legally binding IA was the Environmental Impact Statement (EIS), which came into force on January 1, 1970 in the USA (Bond and Pope 2012). This tool was aimed at federal actions that might significantly affect the quality of human environment (Francis 1975). At first, the tool was used for large projects such as dams or nuclear power plants but it was also used for assessment of smaller projects in urban environment relatively soon (Francis 1975).

The system of impact analysis in the United States was strengthened in the 1980s, when the Urban and Community Impact Analysis (UCIA) was developed as a consequence of Jimmy Carter's presidential campaign promise to develop the first explicit urban policy. The main aim of the impact assessment program was to ensure that there will be no actions contradicting the urban policy (Hack and Langendorf 1980).

Glickman (1980) summarizes the methodology as being focused on assessing policies rather than projects and being focused on places rather than people UCIA was being focused on assessing macroeconomic impacts. Evers (2011) understands the Urban Impact Analysis as a clear predecessor to the TIA, since "it sought to assess the impact of non-urban policies (regional, local or even international) on urban areas" (p. 8).

The implementation of the Urban Impact Analysis sparked a debate concerning potential unintended spatial effects of public policies. With the exception of the United States, there were no impact analyses or assessments embedded in the law in other advanced countries. However, according to Nijkamp and Van Pelt (1989), the OECD collected a number of case studies of urban impact analyses in Sweden, Canada, or France.

In the late 1980s, the term strategic environmental assessment was coined in an interim report to the European Commission (Fungisland Tetlow and Hanusch 2012). Strategic environmental assessment of policies and strategies has been clearly developed upon the practice of Environmental Impact Assessment (EIA) focused on the level of projects. However, whereas EIA is primarily concerned with how a certain proposed development could affect the environment, Strategic Environmental Assessment (SEA) was also supposed to analyse potential effects of alternative developments early in the decision-making process (Fungisland Tetlow and Hanusch 2012).

2.2 Territorial Impact Assessment of projects, policies and future trends

Since the addition of territorial cohesion among the goals of the European Union in 2009, the importance of evaluating potential territorial impacts of policies, strategies or particular projects became yet more relevant, at least in the European context. Nevertheless, policy-evaluation tools were developed already in 1980s as an integral part of the Structural Funds intervention process (Medeiros 2017). Between 1995 and 1999, the European Commission (EC) produced a detailed evaluation methodology under the MEANS programme, which *inter alia* clearly distinguished between "results" and "impacts". The first one connected to direct measurable consequences, the latter focused on long-term consequences. Medeiros (2017) analysed "the MEANS collection" and concluded that "the notion of territorial impacts is entirely absent" and that "no concrete references are made to the need for a more holistic impact assessment" (p. 149).

Another set of documents dealing with TIA were the outcomes stemming from the EVALSED programme concluded in 2008 (updated in 2013). According to Medeiros (2017), EVALSED took territorial dimension of policy evaluation more into account.

The European Commission released its first Impact Assessment (IA) guidelines only in 2005. Surprisingly, the guidelines did not include any reference to "territorial impacts"; instead they have focused on the assessment of economic, social, and environmental impacts in a way that is complementary to SEA, and EIA (Tscherning, König, Birthe, Helming and Sieber 2007). At the same time, during public consultation to the IA procedure, the EC published guidelines to assess regional and local impacts, which could be considered as the first EC attempt to establish a TIA procedure (Medeiros 2017).

As a result of all these programs and activities, the awareness about TIA, at least on the European level, was growing. The European Commission document "Investing in Europe's Future – Fifth Report on Economic Social and Territorial Cohesion" included a statement that "both policies with and without an explicit spatial dimension could benefit from an assessment of territorial impact" (European Commission 2010, p. 195).

However, at the beginning (in the 1990s and the 2000s), TIA was conceived rather as an ex-post assessment tool of the EU policies such as the Common Agricultural Policy or Transport and Trans-European Network Policies and was based on quantitative models (Fischer et al. 2015).

The European Spatial Development Perspective Action Plan adopted in Tampere in 1999 defined the development of a TIA methodology as one of key tasks for the European Spatial Planning Observation Network (ESPON) (Fischer et al. 2015). ESPON enabled to involve experts in the field of TIA and subsequently to support numerous research projects. Thus, it was the ESPON programme (established in 2002) that triggered the development of first rigorous TIA methodologies. According to Medeiros (2017), the ESPON programme funded more than 20 analyses dedicated to the identification of impact of EU sectoral policies, but only a few of them could be regarded as TIA tools. Fischer et al. (2015) divide ESPON methodologies into several period which demonstrates gradual shift of TIA design.

First period (2004–2006) was characterized by use of quantitative models and ex-post assessment of particular European policies (e.g. Common Agricultural Policy). During this period for instance STIMA (Spatial Telecommunications IMpact Assessment) methodology has been established (see ESPON 2004).

Methodologies prepared in second period (2008–2010) were still based on quantitative models, but focused mostly on ex-ante assessment. This was the case for instance for TEQUILA tool/model which included all dimensions and components of the concept of territorial cohesion (Medeiros 2017), but the results were difficult to read (Medeiros 2017). Specific was ARTS methodology which admitted use of qualitative models and more participatory approach.

Third period (since 2012) is connected especially with EATIA methodology which brings significant simplification of TIA procedure and use of qualitative methods. EATIA is still based on ex-ante assessment. Another step in a quest for simplification was the introduction of the TIA Quick Check Tool. This interactive web-based tool was introduced by ESPON in 2015 and enables to measure territorial impacts at the level of the NUTS 3 regions. The TIA Quick Check Tool Is based on the ESPON ARTS methodology and was prepared by the Austrian Institute for Spatial Planning. However, according to Medeiros (2017), the TIA quick check tool is too simplified and does not include all crucial components of territorial impact evaluation. Medeiros (2017) calls for a more robust (and relevant) evaluation technique, despite the fact that such techniques demand more resources and time.

Overall, one can argue that TIA moved from a tool to be used for assessment of territorial impact of projects (Austria, Germany) to a tool used for evaluation of potential territorial impacts of policies. There have been even efforts to analyse territorial impacts of future trends (for instance Böhme and Lüer 2017).

Broadly speaking, the methodology of TIA is similar to those of EIA and SEA, or other impact assessment tools. All of these methodologies "explore the causal links between proposed actions and impacts" (Perdicoulis et al. 2016, p. 42). EIA, SEA, or Cost Benefit Analyses (CBA) methodologies can be considered as prescriptive regarding which procedures should be followed and which aspects should be addressed (Fischer et al., 2015). TIA, on the other hand, is supposed to be more holistic, and should also take into account governance arrangements. According to Evers (2011), the emergence of TIA methodology has also been caused by the need "to gain information on policy effects within an increasingly fragmented public sector" (p. 76).

Medeiros (2017) sees TIA as a tool with a significant potential to replace the EIA and SEA, since all existing TIA methodologies take into account the environmental dimension. However, it might be a challenging task, given the path-dependency of public policies. Consequently, when compared to its predecessors focusing on impacts in a narrow sense (EIA, SEA, CBA), TIA enables to evaluate a potential impact in a more complex way (Evers 2011).

Medeiros (2014) summarizes briefly all impact assessment tools and concludes that TIA is "the most difficult and complete IA procedure, since it needs to take into consideration all the aspects of territorial development (socio-economic, environmental, governance and spatial organization)" (p. 198). However, according to Medeiros (2015), the territorial development does not equal the sum of socio-economic and environmental impacts. Medeiros (2018 and 2019) also suggests to apply TIA to evaluate spatial planning processes and (Medeiros 2019) and sectoral policies, for instance cross-border cooperation (Medeiros 2018).

According to Medeiros (2015), the economic dimension of policy evaluation has always been omnipresent because of the "strong position of economists in this particular scientific field of policy evaluation". Medeiros (2015, p. 9) also explains that complex IA tools started to be tested after some of the EU strategic documents explicitly recognized the "territorial dimension as a key element of policy intervention". Medeiros (2015, p. 9) suggests that there is a mainstream vision that policies should contribute primarily to the economic development. Nevertheless, he calls for a more holistic approach that would take into account several other dimensions of development such as social cohesion, environmental sustainability, territorial governance and spatial planning. Similarly, Zonneveld and Waterhout (2009) argue that TIA "has to focus on the impact of policy and contextual developments on the spatial organization and spatial position of spaces and places" (p. 3).

3. TIA implementation in EU member states

TIA at the EU level certainly represents a valuable tool that helps to understand unintended territorial impacts of EU policies or new legislature. Nevertheless, such a tool at the EU-level cannot reflect all specificities at the national, regional, or local levels. Therefore, TIA at the EU level should be followed by national TIAs (Zonneveld and Waterhout 2009). However, according to Medeiros (2017), no EU member state or other country has encoded a mandatory use of TIA. Consequently, the TIA procedures could be labelled as an "EU experiment" (Medeiros 2017).

In European countries, different techniques are employed to ensure that sectoral (non-regional) interventions are in line with the principles of regional policy (no matter how defined). Obviously, numerous factors, including the governance system, administrative cultures, and political settings, influence the way (territorial) impact assessment is implemented (Meuleman 2015). Meuleman (2015) suggests that the governance approach may vary: it can be more market-driven in Anglo-Saxon countries, more network-driven in northwest Europe, and more hierarchical in Central, Eastern, and Southern Europe. Each governance system implies certain weaknesses, such as prioritization of efficiency in a market-driven model, the risk of losing focus in a network-driven approach, or rigidity in the hierarchical model (Meuleman 2015).

Within the EATIA project, there were some, albeit limited, efforts to disseminate the European TIA methodology to particular EU member states (Slovenia, Portugal, the United Kingdom) (Fischer et al. 2015). One of the conclusions stemming from the dissemination of results was a suggestion that the TIA concept is flexible enough to be adaptable to different policy-making traditions (Fischer et al. 2015). Although Slovenian (e.g. Golobic and Marot 2011) and Portuguese (e.g. Medeiros 2015) academics contributed significantly to the research in the field of TIA, both in Slovenia and Portugal TIA procedures are still rather unknown concept (Marot 2015; Medeiros 2015).

Regarding TIA implementation EU countries could be distinguished into several categories which are not mutually exclusive. First category consists of German speaking countries (chapter 3.1). Second category represents countries where focus on compliance with regional or sustainable development strategy is required when sectoral policies are designed (chapter 3.2). Countries in third category (chapter 3.3.) enrich traditional impact assessment tools by territorial aspects. Countries in fourth category (chapter 3.4) elaborate Territorial Impact Assessment ad-hoc in order to assess impact assessment of large projects, or European directives/policies.

3.1 TIA as a traditional tool in German speaking countries

Tools similar to TIA are used traditionally mostly in German-speaking countries. The roots of TIA can be traced back to tools used traditionally in Austria, Germany, and Switzerland. Even the fact that the term territorial impact assessment is directly translated from its German equivalent demonstrates that the roots of the tool can be traced in the German-speaking world (Othengrafen and Cornett 2013; Healy 2001). In Austria and Germany, TIA is used extensively for ex-ante evaluation of major projects.

Particularly in Austria, the use of TIA can be traced back to 1959, thus even predating the introduction of EIA (Healy 2001). In German-speaking countries, competences of different levels of public administration (national, regional, and local) are set out very clearly (Tosics et al. 2010). This holds true for spatial planning as well. According to Tosics et al. (2010), the federal systems of Germanic countries result in a strong regional level of planning.

Particular authors even declare that TIA is embedded in Austrian, German, or Swiss Law (e.g. Zonneveld and Waterhout 2009). This was, nevertheless, dismissed by Dallhammer (2016) who argues that "Raumverträglichkeitsprüfung" (which is a tool sharing similarities with TIA) is proceeded on voluntary basis and only in particular federal states (e.g. Carinthia).

Paradoxically, Austria does not have a very strong explicit regional policy (Polvevari and Michie 2011). The Austrian Conference on Spatial Planning has been established only in 1971 in order to enhance cross-sectoral coordination. The body comprises representatives of the state (Bund), federal states (Länder), and municipalities. Each ten years the Austrian Spatial Development Concept is adopted to coordinate federal policies with spatial impact and policies implemented by Länder (Polvevari and Michie 2011). Similarly, in Germany, sectoral policies are coordinated through the Regional Joint Task (Polvevari and Michie 2011).

Germany evaluates ex-ante impacts of transport infrastructure or large retail projects "to verify whether these are in line with the aims and objectives of official planning policies" (Othengrafen and Cornett 2013, p. 13). The process (called Raumordnungsverfahren) usually takes around three months. According to Healy (2001), the German procedure is focused on testing the conformity of new projects and new regional plans with existing plans. Bundesländers (sixteen units) are in charge of this process.

3.2 TIA as tool to align territorial priorities with sectoral and regional strategies

Such strategy is relevant for instance for Switzerland, where Sustainable Development Strategy 2016–2019 states that the implementation of the document is coordinated by the interdepartmental Sustainable Development Committee, which is supposed to ensure the incorporation of sustainable development principles into sectoral policies (Swiss Federal Council 2016).

Similar strategies are employed in Ireland, where the compliance of strategies at both local and regional levels with the Irish National Spatial Strategy is a key requirement (Hague 2010), and in Estonia (Ministry of Regional Development 2018) where compliance with the most important elements of the Estonian Regional Development Strategy for 2014–2020 is assessed while preparing new regulations, strategic policy documents, and grant schemes supported by the EU Structural Funds.

In Poland specific document entitled Territorial dimension of strategic documents have been prepared in order to bridge the gap between territorial and sectoral strategic documents (Szlachta, Zaucha and Komornicki 2017). Similar tool with focus on operational programmes have been invoked in Czechia (see chapter 4).

3.3 TIA as an enrichment of traditionally used tools

This strategy is relevant for Finland, Belgium, or Denmark. Those countries use standard EIA, SEA, or Regulatory Impact Assessment (RIA) procedures and "upgrade" them, at least in some cases in order to assess wider development processes. Finnish EIA aims to evaluate not only the impact on human health and environmental aspects (soil, water, air, climate etc.), but also the impact on "the urban structure, buildings, landscape, townscape and cultural heritage". (Ministry of Environment Finland 2009). Similarly, in Wallonia the EIA procedure was adjusted in order to encompass certain aspects of TIA (Hague 2001; Böhme and Lüer 2017). Denmark held discussions regarding a holistic approach within the EIA procedure currently, as well (Kørnøv, Christensen and Nielsen 2005). In another cases the RIA is "upgraded". This is the case for Belgium and its Flemish Region, where RIA includes the Sustainable Development Impact Assessment (OECD 2010; Huge and Waas 2011).

The United Kingdom has specific impact assessment called rural proofing, which evaluates potential impacts on rural areas (Hague 2010). Rural proofing evaluates impacts on services, infrastructure, business, employment, environment etc. (Department for Environment, Food and Rural Affairs 2017). In 2011, IA Toolkit was published by the Department for Business, Innovation and Skills (2011). In the British context, IA is based mostly on a cost-benefit analysis. The use of impact assessments in the United Kingdom has been discussed in connection with Brexit, but no impact assessment of Brexit has been conducted so far. This was later criticized by Fischer and Sykes (2017), who suggested that impact assessment taking into account not just economic impact but a wider scale of impact should be conducted in case of Brexit, because it is clear that it is going to have uneven regional impacts.

3.4 Ad-hoc use of TIA

Such TIAs were applied in Netherlands, Poland, or Lithuania. In the Netherlands, there is no standard TIA procedure. Large projects are usually assessed via a cost-benefit analysis. Currently, there are efforts to include a social dimension via a social cost-benefit analysis (Evers 2016). Nonetheless, the Netherlands experienced an ad-hoc implementation of TIA in connection with possible regional implications of the EU's policies. According to Evers (2011), the Dutch Ministry of Housing, Spatial Planning and the Environment requested the Environmental Assessment Agency to conduct TIA as a reaction to the European territorial cohesion policy and to the energy-climate package. Evers (2011) stresses that "given the two very different points of the policy process, the studies are very different in character in terms of methodology and conclusions" (p. 98). The TIA focusing on the European territorial cohesion policy was affected by not having a clear definition of territorial cohesion. In contrast, the energy-climate package had clearly stipulated goals and it was thus ambition of TIA to find policy actions in order to meet targets. Ad hoc impact assessments are undertaken as well for large projects (Poland), or after implementation of the EU programming period (Lithuania).

4. Summary

Table 2 shows that there is a rather limited use of Territorial Impact Assessment or similar tools across European countries. Nevertheless, according to several studies (e.g. Dabrowski 2012; Raagmaa and Stead 2015), a certain trend towards Europeanization of public policies can be observed especially in Central and Eastern Europe, especially in the fields of spatial planning and regional policy. This might be seen, for instance, in Partnership Agreements signed between the European Commission and particular member states, which include an analysis of tools to be used by the member states to enhance the territorial approach

Tab. 2 Typology of TIA use in EU member states and Switzerland.

Type of TIA use	Description	Countries
TIA encoded in the law	TIA used extensively for ex-ante impact assessment of large infrastructure projects	Germany, Austria
TIA as tool to align of territorial priorities with sectoral and regional strategies	Compliance of policies, plans, or strategies compared with main regional development strategies	Estonia, Ireland, Switzerland, Poland
TIA as an enrichment of traditionally used tools	Territorial aspects more or less incorporated into existing evaluation techniques	Belgium, Finland, Greece, Denmark, Great Britain
Ad-hoc TIA	TIA used "when needed" such as before implementation of European directives and policies into national contexts	Netherlands
No TIA procedure	Mostly Central and European countries declare no use of standard TIA procedures	Slovakia, Slovenia, Hungary, Lithuania, Latvia, Romania, Bulgaria, Croatia, Czechia, Portugal
No relevant information founded	During desk research no relevant information regarding TIA in seven EU countries has been found	Sweden, Malta, Cyprus, Spain, Italy, France, Luxembourg

Source: Author

in implementation of the EU cohesion policy (Nosek 2017).

The position of (spatial) planning in Central and Eastern European countries in the 1990s was difficult, mostly because planning of any kind was associated with the communist regime by both policy-makers and population in general (Raagmaa and Stead 2015).

There is no doubt that an EU membership pushes member states to adopt changes in the strategic planning process (Dabrowski 2012), including an application of more sophisticated methods in impact assessment. Therefore, it can be expected that some sort of mapping of territorial impact will become more embedded in spatial planning processes in the near future.

Namely, some Central and Eastern European countries plan to apply impact assessments tools over the next few years. This is the case, for instance, in Romania, which expects that the methodology for measuring territorial impact of sectoral policies will be approved by the Government following the approval of the Territorial Development Strategy (Ministry of Regional Development 2018).

5. The case of Czechia

According to the law defining responsibilities of particular Czech ministries, the Ministry of Regional Development is obliged to coordinate other ministries in order to implement regional policy. Such setting might seem favourable for pushing sectoral ministries to reflect regional specifics.

Nonetheless, the actual promotion of regional aspects into sectoral strategies and policy-making process cannot be taken for granted. Such promotion demands numerous discussions and eventually also political support – which is in Czech context frequently lacking. The Ministry of Regional Development is often considered less influential in comparison with Ministry of Industry and Trade or Ministry of Environment. As a consequence, certain ministries see the territorial dimension concept as a threat and intervention into its competencies.

Czechia is one of the countries which do not use TIA methodology at the national level. However, an interest in the tool has been growing in recent years, similarly to other Central and Eastern European countries and certain measures have been taken to adjust sectoral policies in order to reflect territorial specificities.

5.1 TIA at the policy level in 2014–2020 programming period

According to the Partnership Agreement between the European Commission and Czechia concluded in 2014, the territorial dimension is in the Czech context understood as a possibility to concentrate resources from the European Structural and Investments Funds (ESIF) in specific types of territories supporting both competitiveness and moderating territorial disparities (social, economic, environmental).

In Czechia, a relatively comprehensive approach to incorporate the territorial dimension into the programmes within the period 2014–2020 was developed. In particular, a special document called the National Document for Territorial Dimension (NDTD; Ministry of Regional Development 2015) was prepared with the aim to identify, which specific goals should be preferred in specific types of territories.

The incentive for its existence came implicitly from the European Commission, which pushed for more territorially sensitive implementation of European Structural and Investment Funds. This document also served as a "substitute" for Czech NUTS II regions which implemented its own regional operational programmes in 2007–2013 programming period.

Overall, eight types of specific territories have been identified in the NDTD (e.g. for urban areas, for rural areas, for socially excluded localities, for solving labour market problems, for enhancement of entrepreneurship etc.).

NDTD could be understood as a kind of ex-ante TIA. This document indicates which measures will most likely have uneven territorial impacts, similarly to concepts described in the previous chapter and applied by ESPON or by particular countries (e.g. Austria or Germany). However, in contrast to ESPON models, in Czechia, no data analysis has been performed in order to decide which specific goals should have the territorial dimension.

The NDTD also specifies the volume of funds that the state will distribute via the territorial dimension. Resources used for addressing the territorial dimension include funds distributed through tools designed to support urban development (i.e. Integrated Territorial Investments and Integrated Territorial Development Plans) and tools aimed at rural development (Community-led Local Development). The total allocation is expected to be approximately 8 billion EUR, which is almost one third of the Czech allocation from the EU European Structural and Investment Funds.

The very existence of the NDTD might be seen as progress in promoting the territorial dimension into sectoral policy making. However, the NDTD was the first document of its kind implemented in Czechia and suffers from several weaknesses. The NDTD was arguably created too late, since it was finalized well after the finalization of operational programmes. Territorial dimension was thus not implemented into the operational programmes, but it was described ex-post, based on discussion with responsible ministries/managing authorities.

The NDTD does not cover national grant schemes and currently, there is no alternative document stipulating the territorial dimension within national grant schemes or national sectoral policies. For example, even the Regional Development Strategy 2014–2020 does not stipulate how sectoral policies are supposed to reflect territorial specificities. The Czech Regional Development Strategy only identifies micro-regions of 57 municipalities with extended powers as "economically weak" and without any further specification states that these micro-regions should be taken into account within sectoral policies.

Each year, progress evaluation in implementation of the NDTD is to be prepared. This evaluation should be primarily based on an analysis of particular calls for proposal and on an analysis of allocation of funds. So far, evaluation has been prepared for years 2015, 2016, and 2017. According to the evaluations for the period 2015–2017, the calls were in most cases targeted at territories specified by the NDTD. Analyses of spatial distribution of allocation of funds have been limited by poor quality of available databases, in which a large proportion of projects did not include information concerning their localization at the municipal level. Nonetheless, available data suggest that the share of the Cohesion policy expected by NDTD to be spend on the territorial dimension has been achieved.

The analysis of individual calls for proposals raised several important questions. First, the pivotal question was which calls should be declared as taking into account the territorial dimension. The NDTD did not include any rigorous methodology in order to assess calls with regard to its territorial dimension. Eventually, four types of possible scenarios have been identified by the Ministry of Regional Development:

- The call is focused exclusively on a specific territory (for instance, on a selected NUTS 2 region).

Tab. 3 Territorial dimension in operational programmes in Czechia in 2014–2020.

Operational Programme	Level of Territorial Dimension	Description of Territorial Dimension
Integrated Regional OP	High	Specific calls for certain type of regions (e.g. regions with socially excluded localities); Funds in certain calls distributed in accordance with regional needs.
OP Entrepreneurship and Innovation for Competitiveness	High	Regions with high unemployment rate or economically weak regions benefited in selection process.
OP Transport	Medium	Projects contributing to regional development priorities supported primarily.
Rural Development Programme	Medium	Most of the activities within Community-led Local Development financed via Rural Development Programme.
OP Research, Innovation and Education	Medium	Local and regional action plans in education define local/regional priorities in elementary schools and high schools.
OP Environment	Low	In certain cases specific territorial dimension applied – for instance localities specified in flood plans.
OP Employment	Low	Except of labour market policy.
OP Prague – Growth pole	Not Relevant	Focused only on the Capital City of Prague.

Source: Author

– The call explicitly declares the distribution of funds among a specific type of regions/municipalities.
– The call sets up the criteria favouring particular regions (e.g. those with an above-average rate of unemployment).
– The call is focused on addressing specific problems that are concentrated in certain regions/territories (i.e. implicit territorial dimension).

When applying such criteria, it appeared that most calls that were expected to have a territorial dimension indeed incorporated it in one way or another into the calls' designs.

The Ministry of Regional Development itself manages in 2014–2020 programming period Integrated Regional Development Programme, which is seen as a flagship programme for regions and cities. Certain important aspects of regional development were however implemented by other (sectoral) ministries, as well.

Table 3 summarizes operational programmes implemented in 2014–2020 programming period and territorial dimension expected by National Document for Territorial Dimension.

5.2 Territorial dimension expectations in post-2020 period

Discussions regarding implementation of post-2020 programming period are already underway. Following significant growth of Czech economy in recent years, large decrease of European funds allocation in Czechia is expected in post-2020 period. Thus, the Regional Development Strategy post-2020 aims to define the territorial dimension within particular national sectoral policies in addition to the territorial dimension embedded in operational programmes supported via the European Structural and Investment Funds.

Discussion with bodies responsible for sectoral policies again shows scepticism of sectoral ministries regarding the role of Ministry of Regional Development. Promotion of regionally sensitive policies is often understood as a unnecessary intervention which put artificial element into the policy-making process.

Regional Development Strategy post-2020 aims to elaborate only topics with identified territorial dimension and is supposed to be more specific with comparison to its predecessor. Thus, key questions appearing during most of the discussions are: (1) which topics have territorial dimension (i.e. needs territorially sensitive solutions), and for those which have, (2) how it can be defined in particular fields? Apparently, the debate is often affected by conflicting motivations of particular stakeholders. While representatives of rural areas, cities and regions have a tendency to promote territoriality as much as possible, ministerial officers tend to declare that certain fields (and respective policies) are spatially blind and no significant regional adjustments are needed.

It might be argued that "territorial" stakeholders aim to ensure allocation of funds in future policies implementing the strategy. In their understanding, more territoriality equals higher chances to get funded. Ministries, on other hand, aim to avoid, from their point of view, unnecessary intervention into their policies. In such context, the role of Ministry of Regional Development might be seen as of a mediator.

5.3 TIA at the Project level

Until 2016, there was no certified TIA methodology in Czechia at the level of projects. Finally, in 2016, the Ministry of Regional Development published project impact assessment methodology prepared in cooperation with regional policy experts. Following the publication of this methodology, the Ministry of Regional Development plans to start a project (to be implemented in 2019–2020) aimed at raising awareness at the methodology and to train public officials who will be implementing the methodology on all hierarchical levels – national, regional, and municipal. There is also a plan to develop a software tool to guide potential users through the methodology.

This TIA methodology has been inspired by similar documents based on the initiative of the European Commission (especially EC 2009), which includes a list of potential impacts that should be taken into account when evaluating possible impacts of EU policies. Therefore, the Czech methodology designed at the project level has been developed on the basis of the EU methodology, which, however, targeted the level of policies.

The TIA procedure is divided into two steps. In the first step, the evaluator must assess overall quality of the project proposal and decide whether the project will have a significant territorial impact or not. The potential impact is evaluated in the following fields: employment, environment, transport, market, quality of life, and public services. The evaluator must also assess what the level of impact is (average, important, and fundamental).

Project proposals that are considered as those with a potential socioeconomic impact are subject to further analysis. The subsequent procedure of the TIA differs based on the scale of the potential impact in the above-listed fields.

Prior to 2016, TIA or similar methodologies were only exceptionally applied in Czech regions and/or municipalities. One of these exceptions was the city of Litoměřice in Northern Bohemia, whose representatives decided to establish a procedure analogical to TIA in order to measure the possible impact of project proposals supported by the city from the economic, social, and environmental point of view. The assessment is undertaken by "Sustainable Development Strategic Team" consisting of the mayor or his/her deputy, members of the department of strategic development and the department of sustainable development, and members of other relevant departments (depending on type of the project). The resulting assessment is subsequently presented to the political representation and forms the basis for the decision-making process. This process is similar to the one employed in Austrian Carinthia, as described in Chapter 3. However, city representatives are not obliged to follow recommendations of this assessment.

6. Conclusions

Studies in the sphere of regional policy and regional development have emphasized uneven territorial impacts of non-regional public policies for a long time. Over the past two decades, the search for a sophisticated tool that would enable to analyse territorial impacts of proposed interventions has already found several potential solutions.

Nevertheless, the ambitions behind the search for a suitable TIA methodology seem to be conflicting. On the one hand, the tool is supposed to be sophisticated enough to take into account all potential territorial impacts, while, on the other hand, it should be simple enough to serve the purposes of public officials in the sphere of regional development at all levels of public administration. This is hardly realistic. Nevertheless, the paper suggests that the territorial impact assessment is a tool that is developing quite dynamically both at the EU level and at the level of member states.

The process of developing TIA methodologies is still under-researched. The aim of this paper was to contribute to research of TIA via an analysis of the extent and form to which the EU member states implement the TIA methodology within their policy-making processes. Specific attention was paid to the case of Czechia, which recently witnessed an upsurge in the interest of public officials in the TIA methodology at the national level.

The first question posed in this paper was focused on the current state of affairs in European countries regarding the use of Territorial Impact Assessment or a similar tool. It can be concluded that while the European Commission has been a clear leader in developing new (and in the past decade largely simplified) methods of territorial impact assessment, the EU member states have not paid similar attention to measuring territorial impacts despite the recent incorporation of territorial cohesion among the European Union goals. From author's overview follows Germany and Austria lead the way towards sound TIA implementation. Several other countries aim at introducing at least simple methods to measure uneven territorial impacts of public policies (Estonia, Romania). This process might be seen as another proof of Europeanization of public policies in Central and Eastern European countries (Dabrowski 2012; Raagmaa and Stead 2015).

The second question focused solely on the recent development of TIA in Czechia. In this country, in order to encompass the territorial dimension within the operational programmes for the EU 2014–2020 programming period, the National Document for Territorial Dimension has been approved by the government. It is the first document of its kind and, despite some shortcomings, it forced the sectoral ministries to take the territorial dimension into account by one way or another during implementation of operational programmes they are responsible for. Nevertheless, a large variety of interpretations of the term territorial dimension have been observed during the implementation phase of particular operational programmes.

Finally, at the project level, a significant achievement has been recently accomplished in Czechia, as in 2016 a TIA methodology at the project level has been published by the government. This came just in time to assist some more active cities and regions in their attempt to design an impact assessment approach to their own projects. Thus, the recently published TIA methodology can stimulate stakeholders at the local and regional level to deal seriously with uneven territorial impacts of various public interventions. The extent to which this potential to deliver more effective and efficient public policies will be used by local and regional actors remains to be seen

Consequently, a thorough analysis of advances made in the field in various countries should be subject to further research.

References

Bond A., Pope J. (2012): The state of the art of impact assessment in 2012. Impact Assessment and Project Appraisal 30(1), 1–4, https://doi.org/10.1080/14615517.2012.669140.

Böhme, K., Doucet, P., Komornicki, T., Zaucha, J., Świątek, D. (2011): How to strengthen the territorial dimension of Europe 2020 and the EU Cohesion Policy. Report based on Territorial Agenda 2020. Retrieved March 1, 2018, http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/challenges2020/2011_territorial_dimension_eu2020.pdf.

Böhme, K., Lüer, Ch. (2017): Assessing territorial impacts of future trends. In: Medeiros, E.: Uncovering the Territorial Dimension of European Union Cohesion Policy. Cohesion, Development, Impact Assessment and Cooperation, https://doi.org/10.4324/9781315169743-11.

Caldwell, L. K. (1988): Environmental Impact Analysis (EIA): Origins, Evolution, and Future Directions. Impact Assessment 6(3–4), 75–83, https://doi.org/10.1080/07349165.1988.9725648.

Dabrowski, M. (2012): Shallow or deep Europeanisation? The uneven impact of EU cohesion policy on the regional and local authorities in Poland. Environment and Planning C: Government and Policy 30(4), 730–745, https://doi.org/10.1068/c1164r.

Dallhammer, E. (2016): Personal communication, 9 February, 2016.

Department for Business, Innovation and Skills (2011): Impact Assessment Toolkit. Retrieved May 6, 2018, www.legislationline.org/documents/id/16803.
Department for Environment, Food and Rural Affairs (2017): Rural Proofing. Practical guidance to assess impacts of policies on rural areas. Retrieved May 5, 2018, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/600450/rural-proofing-guidance.pdf.
ESPON (2004): ESPON 2.1.1: Territorial Impact of EU Transport and TEN policies.
ESPON (2013): ESPON ARTS: Assessment of Regional and Territorial Sensitivity. Final Report.
European Commission (2009): Impact Assessment Guidelines. Retrieved April 9, 2018, http://ec.europa.eu/smart-regulation/impact/commission_guidelines/docs/iag_2009_en.pdf.
European Commission (2010): Fifth Report on Economic, Social and Territorial Cohesion – Investing in Europe's future. Retrieved April 8, 2018, http://ec.europa.eu/regional_policy/en/information/publications/reports/2010/fifth-report-on-economic-social-and-territorial-cohesion-investing-in-europe-s-future.
European Commission (2011): Territorial Agenda of the European Union 2020. Retrieved April 7, 2018, http://ec.europa.eu/regional_policy/en/information/publications/communications/2011/territorial-agenda-of-the-european-union-2020.
European Council of Town Planners and Committee on Spatial Development (2001): Proceedings of the one-day conference on Territorial impact assessment 26 October 2001 Louvain-la-Neuve, Belgium. Retrived March 14, 2018, http://www.ectp-ceu.eu/images/stories/download/cp20030515.pdf.
Evers, D. (2011): Territorial Impact Assessment: a critical examination of current practice. PBL Netherlands Environmental Assesment Agency.
Evers. D. (2016): Personal communication, 18, January, 2016.
Hayes, A. C. (2017): What is impact assessment? Some personal reflections C. P. Wolf (1933–2015), edited posthumously by Adrian C. Hayes. Impact Assessment and Project Appraisal 35(3), 186–199, https://doi.org/10.1080/14615517.2017.1322812.
Hague (2010): European Perspectives on Territorial Impact Assessment. A Background paper for the ESPON EATIA Project. Retrieved March 11, 2018, http://www.rtpi.org.uk/media/5994/European-Perspectives-on-Territorial-Impact-Assessment-April-2010.pdf.
Fischer, T. B. (2014): Impact Assessment: there can be strength in diversity! Impact Assessment and Project Appraisal 32(1), 9–10, https://doi.org/10.1080/14615517.2013.872844.
Fischer, T. B., Sykes, O., Gore, T., Marot, N., Golobič, M., Pinho, P., Waterhout, B., Perdicoulis, A. (2015): Territorial Impact Assessment of European Draft Directives – The Emergence of a New Policy Assessment Instrument. European Planning Studies 23(3), 433–451, https://doi.org/10.1080/09654313.2013.868292.
Fischer, T. B., Sykes, O. (2017): Impact Assessments: What Impact Assessments? And what is an Impact Assessment? Retrieved March 10, 2018, https://news.liverpool.ac.uk/2017/12/08/impact-assessments-what-impact-assessments-and-what-is-an-impact-assessment.
Francis, M. (1975): Urban Impact Assessment and Community Involvement: The Case of the John Fitzgerald Kennedy Library. Environment and Behavior 7(3), 373–404, https://doi.org/10.1177/001391657500700307.
Fungisland Tetlow, M., Hanusch, M. (2012): Strategic environmental assessment: the state of the art. Impact Assessment and Project Appraisal 30(1), 15–24, https://doi.org/10.1080/14615517.2012.666400.
Glickman, N. J. (1980): Urban Impact Analysis: Premises, Promises, Procedures, and Problems. Built Environment 6(2), 84–91.
Golobic, M., Marot, N. (2011): Territorial impact assessment: Integrating territorial aspects in sectoral policies. Evaluation and Program Planning 34, 163–173, https://doi.org/10.1016/j.evalprogplan.2011.02.009.
Hack, G., Langendorf, R. (1980): Lessons from Urban Impact Assessment. Environmental Impact Assessment Review 1(3), 251–266, https://doi.org/10.1016/S0195-9255(80)80103-X.
Healy, A. (2001): Keywork Phrase: Territorial Impact Assessment. Retrieved June 18, 2017, http://www.esprid.org/keyphrases/30.pdf.
Huge, J., Waas, T. (2011): Converging impact assessment discourses for sustainable development: the case of Flanders, Belgium. Environment, Development, and Sustainability 13(3), 607–626, https://doi.org/10.1007/s10668-010-9279-z.
Kørnøv, L., Christensen, P., Nielsen, E. H. (2005): Mission impossible: does environmental impact assessment in Denmark secure a holistic approach to the environment? Impact Assessment and Project Appraisal 23(4), 303–314, https://doi.org/10.3152/147154605781765427.
Marot, N. (2015): On the maximal shortest path in a connected component in V2V. Performance Evaluation 94, 25–42, https://doi.org/10.1016/j.peva.2015.09.003.
Medeiros, E. (2014): Assessing territorial impacts of the EU cohesion policy at the regional level: the case of Algarve. Impact Assessment and Project Appraisal 32(3), 198–212, https://doi.org/10.1080/14615517.2014.915134.
Medeiros, E. (2015): personal communication, October 8, 2015.
Medeiros, E. (2015): Territorial Impact Assessment and Public Policies: The Case of Portugal and the EU.
Medeiros, E. (2017): From simple to relevant TIA tools for European policies. In: Medeiros, E.: Uncovering The Territorial Dimension of European Union Cohesion Policy. Cohesion, Development, Impact Assessment and Cooperation, https://doi.org/10.4324/9781315169743.
Medeiros E. (2018) Focusing on Cross-Border Territorial Impacts. In: Medeiros E. (eds) European Territorial Cooperation. The Urban Book Series. Springer, Cham, https://doi.org/10.1007/978-3-319-74887-0.
Medeiros, E. (2019) Spatial Planning, Territorial Development and Territorial Impact Assessment. Journal of Planning Literature 34(2), 171–182, https://doi.org/10.1177/0885412219831375.
Meuleman, L. (2015): Owl meets beehive: how impact assessment and governance relate. Impact Assessment and Project Appraisal 33(1), 4–15, https://doi.org/10.1080/14615517.2014.956436.
Ministry of Environment Finland (2009): Act on Environmental Impact Assessment Procedure, Unofficial

translation. Retrieved August 24, 2017, https://www.finlex.fi/en/laki/kaannokset/1994/en19940468.pdf.
Ministry of Regional Development (2015): National document for Territorial Dimension.
Ministry of Regional Development (2018): Questionnaire among National Territorial Cohesion Contact Points.
Morrison-Saunders, A., Pope, J., Gunn, J. A. E., Bond, A., Retief, F. (2014): Strengthening impact assessment: a call for integration and focus. Impact Assessment and Project Appraisal 32(1), 2–8, https://doi.org/10.1080/14615517.2013.872841.
Nijkamp, P., van Pelt, M. (1983): Spatial Impact analysis for developing countries. A framework and a case study, Serie Research Memoranda 0014, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics.
Nosek, Š. (2017): Territorial cohesion storylines in 2014–2020 Cohesion Policy. European PlanningSstudies 25(12), 2157–2174, https://doi.org/10.1080/09654313.2017.1349079.
OECD (2010): Better regulation in Europe. Retrieved May 17, 2018, http://www.oecd.org/gov/regulatory-policy/betterregulationineuropebelgium.htm.
OECD (2011): Territorial Reviews Switzerland. Retrieved May 15, 2018, https://read.oecd-ilibrary.org/urban-rural-and-regional-development/oecd-territorial-reviews-switzerland-2011_9789264092723-en#page1.
Othengrafen, F., Cornett, A. P. (2013): A Critical Assessment of the Added Value of Territorial Cohesion. Refereed article No. 53, October, 2013, European Journal of Spatial Development, https://archive.nordregio.se/en/News/A-Critical-Assessment-of-the-Added-Value-of-Territorial-Cohesion/index.html.
Perdicoúlis, A., Batista, L., Pinho, P. (2016): Logical chains in territorial impact assessment. Environmental Impact Assessment Review 57, 46–52, https://doi.org/10.1016/j.eiar.2015.11.006.
Polverari, L., Michie, R. (2011): Complementarity or conflict? The (in)coherence of Cohesion policy. European Policy Research Center, No. 78, University of Strathclyde 40 George Street Glasgow G1 1QE, http://www.eprc-strath.eu/public/dam/jcr:23283c26.../EPRP%2078.pdf.
Raagmaa, G., Stead, D. (2015): Spatial Planning in the Baltic States: Impacts of European Policies IN Raagma, G., Stead, D. (eds.). Impacts of European Territorial Policies in the Baltic States. London: Routledge – Taylor & Francis Group, 1–9.
Stutz, F.P., Warf, B. (2012): The World Economy: Geography, Business, Development. Hoboken, NJ: Pearson Education.
Szlachta J., Zaucha, J., Kormonicki, T. (2017): Polish development policy and its territorial dimension. In.: J. Bradley, J.Zaucha (eds.): Territorial Cohesion: A missing link between economic growth and welfare. Lessons from the Baltic tiger, Gdansk, 49–72.
Swiss Federal Council (2016): Sustainable Development Strategy 2016–2019. Retrieved April 23, 2018, https://www.eda.admin.ch/agenda2030/en/home.html.
Tosics, I., Szemzö, H., Illés, D., Gerheis, A., Lalenis, K., Kalergis, D. (2010): National spatial planning policies and governance typology, www.plurel.org/images/D221.pdf.
Tscherning, K., König, H., Birthe, S., Helming, K., Sieber, S. (2007): Ex ante impact assessments (IA) in the European Commission: An overview. In: Helming, K., Marta-Pérez-Soba, M., Tabbush (eds.): Sustainability Impact Assessment of Land Use Changes. New York: Springer, 17–33, https://doi.org/10.1007/978-3-540-78648-1_3.
Zonneveld, W., Waterhout, B. (2009): EU Territorial Impact Assessment: Under what Conditions? Final Report, Delft University of Technology, Delft, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.615.647&rep=rep1&type=pdf.

Influence of neotectonics on land surface evolution in the upper part of the Blue Nile Basin (Ethiopia): findings from a DEM

Michal Kusák[1,*], Vít Vilímek[2], Jozef Minár[3]

[1] Institute of Rock Structure and Mechanics of the Czech Academy of Sciences, Czech Republic
[2] Charles University, Faculty of Science, Department of Physical Geography and Geoecology, Czech Republic
[3] Comenius University in Bratislava, Faculty of Natural Sciences, Department of Physical Geography and Geoecology, Slovak Republic
[*] Corresponding author: kusak@irsm.cas.cz

ABSTRACT

The morphometric analysis of lineaments, valleys and signs of erosion taken from a digital elevation model (DEM) made it possible to not only confirm most of the conclusions of the morphotectonic development of the Blue Nile Basin from the previously published results of structural, petrological, tectonic and geochronological analyses, but also to expand our knowledge by applying several new hypotheses. The relative age of the morpholineaments of particular directions was estimated from the character of topographic profiles. Faults, lineaments and valleys are predominantly oriented in a direction compatible to the published concepts of the tectonic development of the area. Overall, the most abundant NE-SW and NNE-SSW lines reflect a change of extension from a NW-SE to WNW-ESE direction during the Pliocene, in relation to the creation and development of the Main Ethiopian Rift (MER). This is confirmed by a more developed character of the valleys and less pronounced erosion activity of the NE-SW oriented valleys contrary to the deeper narrower NNE-SSW valleys characterised by downward and headward erosion in the second direction. The most pronounced morphological manifestations of the E-W extension of the MER and western Afar during the Quaternary are confined to the borders of the MER. The directions of the Pre-Neogene rift structures to the NW-SE and WNW-ESE are compatible with the oldest elements of the current landscape and with the relict fragments of the valley network on the SE boundary of the upper Blue Nile Basin, which could have been drained across current shoulders of the MER to the S and E before the Late Miocene.

KEYWORDS

landscape evolution; neotectonics; river piracy; lineaments; DEM; Main Ethiopian Rift; Ethiopian Highlands

Received: 17 April 2019
Accepted: 28 June 2019
Published online: 26 August 2019

Kusák, M., Vilímek, V., Minár, J. (2019): Influence of neotectonics on land surface evolution in the upper part of the Blue Nile Basin (Ethiopia): findings from a DEM. AUC Geographica 54(2), 129–151
https://doi.org/10.14712/23361980.2019.13

1. Introduction

Lineaments are linearly arranged elements of the landscape – linear sections of a valley, ridges or straight sections of slopes – should be considered as a potential zone of brittle fractures of bedrock and/or the concentration of erosion processes with an influence on the geomorphological evolution of an area (Abdullah et al. 2010). An analysis of the lineaments can give an insight into landscape evolution, thus providing information on tectonic activity over large areas, which is particularly useful for areas with limited field access (Ehlen 2004), like the Ethiopian Highlands. Recently, the lineaments are increasingly used not only as a surface expression of individual faults (fracture zones) but their statistics is used also to define/confirm the character (directions) of the palaeomorphotectonic stress fields (Minár and Sládek 2008; Koronovskya et al. 2014; Šilhavý et al. 2016).

In this paper we aim to understand relations between landforms and the tectonic structures of the Ethiopian Highlands, as the region is strongly influenced by tectonics (Kazmin 1975; Pik et al. 2003; Beyene and Abdelsalam 2005; Gani and Abdelsalam 2006; Gani et al. 2007, 2009; Wolela 2010). We focused on the upper part of the Blue Nile Basin.

The main objectives of this work are to: 1) analyse the morpholineaments, faults and valley networks of the Blue Nile Basin and their interrelationships; 2) study the relation between the morpholineaments and the geomorphological evolution, in order to determine which elements are potential zones of brittle fractures of bedrock and serve as concentrations for erosion processes; 3) identify areas with the most dynamic changes of valley networks.

Several works deal with the tectonic and volcanic history of the upper part of the Blue Nile Basin (Abebe et al. 1998; Chorowicz et al. 1998; Wolfenden et al. 2004; Pik et al. 1998, 2003; Kieffer et al. 2004; Gani et al. 2009); however, less attention has been paid to the related geomorphic development. Local morphotectonic and morphodynamic aspects have mainly been investigated (Ayalew and Yamagishi 2003; Ismail and Abdelsalam 2012; Kusák et al. 2016; Mäerker et al 2016; Kycl et al. 2017; Gani and Neupne 2018). Works by Gani and Abdelsalam (2006) and Gani et al. (2007) deal most comprehensively with the selected aspects of geomorphic development.

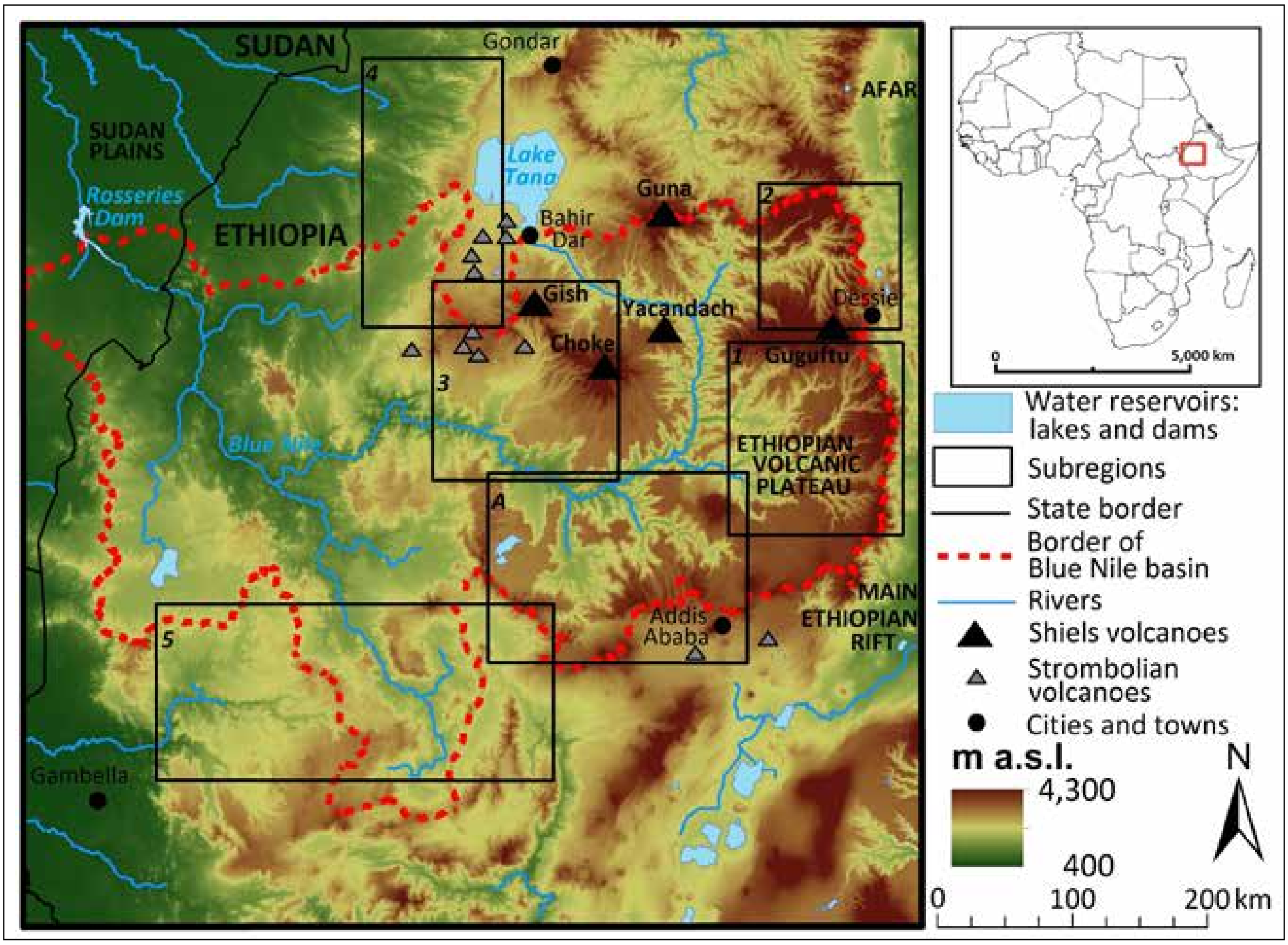

Fig. 1 Map of the Blue Nile basin and adjacent area, Ethiopian Highlands. Subregions particularly analyzed in the following text: 1 – the Jemma River; 2 – Beshlo Wenz River; 3 – Volcanoes Mt. Choke and Mt. Gish; 4 – Lake Tana; 5 – Didessa River and Baro River; A – Guder River and Muger River.

They refuse the older idea that present physiography of the elevated plateau exists since the Oligocene (e.g. Pik et al. 2003) and they argue that the rapid increase of incision rate of the Blue Nile initiates ca. 10 Ma ago. This paper presents evidence from satellite data that support young (Late Miocene – Quaternary) increasing morphotectonic activity of the territory, specifying directions and space impact of palaeomorphotectonic fields on the land surface development.

2. The study area: geomorphological and geological settings

The study area includes the Blue Nile Basin between Lake Tana and the Roseires water reservoir (175,400 km^2) and the adjacent area (Figure 1). The Blue Nile Basin consists of a large plateau that is inclined from the border of Afar and the MER (more than 3,000 m a.s.l.) to the W and NW (approximately 2,000 m a.s.l.). Huge Neogene shield volcanoes tower up to 2,000 m above the plateau, and smaller strombolian Quaternary volcanoes are concentrated to the SSW of Lake Tana reach several hundreds of meters at most (Kieffer et al. 2004). Pliocene and Quaternary tectono-thermal uplift of the rift shoulders and a swell centred on the Lake Tana basin are responsible for the recent elevation of the plateau (Chorowicz et al. 1998; Ismail and Abdelsalam 2012; MacGregor 2015; Gani and Neupne 2018). Uplift caused the rivers to cut into the bedrock and divide the area by canyons and deep valleys with steep slopes linked to faults in the eastern part of the study area (Gani et al. 2009). An older landscape area at a lower elevation and with gentle valley slopes is situated in the western part of the study area and the Blue Nile flows through the basin with a low vertical division (Figure 1).

The modern topography of the area formed Tertiary and Quaternary volcanism and tectonics. The uplift of the Ethiopian Highlands (during the last 29 Ma) and opening of the MER (during the last 18 Ma) has led to the formation of faults and cracks (Kazmin 1975; Pik et al. 2003; Beyene and Abdelsalam 2006; Gani et al. 2007; Gani et al. 2009; Wolela 2010; Ismail and Abdelsalam 2012).

The plateau in the eastern part of the study area is built by various types of Cenozoic volcanic rocks and several Cenozoic volcanoes tower above it. Erosion by the Blue Nile and its tributaries exposed Jurassic and Crateceous sedimentary complexes in the valleys incised into the plateau. The western part of the study area is built by a Precambrian crystalline complex and volcanics (Figure 2; Mangesha et al. 1996).

The Ethiopian Highlands are marked by a number of faults increasing from the west to the east, where they border the MER – part of the East African Rift (Chorowitz 2005). The oldest Permo-Trias and Crateceous-Palaeogene rift systems in the study area have a NW-SE orientation (Gani et al. 2009; MacGregor 2015). The main phase of extension in the MER started approximately 11 Ma (Ukstins et al. 2002) and the majority of fault activity occurred in the last 5–9 Ma, and particularly in the last 1–2 Ma (MacGregor 2015). The NW-SE Late Miocene extension formed NE(NNE)-trending faults (Wolfenden et al. 2004; Gani et al. 2009). E-W and NNE-SSW Quaternary extensions

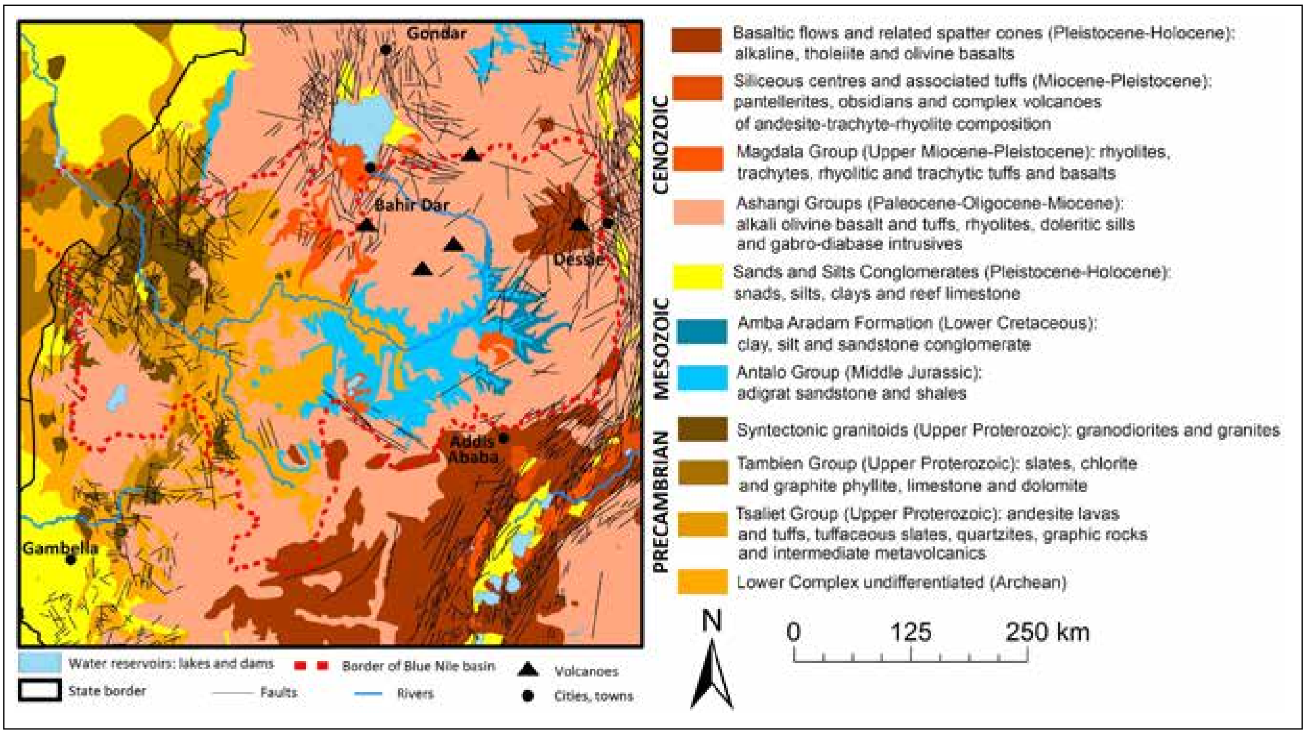

Fig. 2 Geological map of the Blue Nile basin and adjacent area, Ethiopian Highlands (scale 1 : 250,000; after Mangesha et al. 1996).

opening the MER are related to the development of E-trending transverse faults. N-ESE- and NW-trending extensional structures are connected with the NE-SW Quaternary extension in southern Afar and the E-W extension in western Afar (Gani et al. 2009). To the west of Addis Ababa, an elongated Yerer-Tullu Wellel volcano-tectonic lineament is probably an inherent structure, which has decoupled the Northern Ethiopian Plateau and the MER as a transtensional dominantly oblique E-W to ENE-WSW structure since the Late Miocene (Abebe et al. 1998).

3. Materials and methods

3.1 Digital Elevation Model (DEM) and satellite images

In this paper we use the SRTM DEM 1 Arc-Second Global version (resolution ≈ 30 m), which was released in September 2014. Shaded relief images with varying directions of illumination were created in ArcMap 10.2 (ESRI 2015) using the hillshade tool (Spatial Analyst Toolbox → Surface toolset). Various elevation angles of the illumination source (30°, 35°, 40° and 45°) and four different values of illumination azimuth 1) from the north (solar azimuth of 0°), 2) from the north-east (solar azimuth of 45°), 3) from the east (solar azimuth of 90°), and 4) from the south-east (solar azimuth of 135°), were used in order to assure the independency of the results from the direction of illumination. This procedure maximizes the visualization of a surface for better observation and graphical display, to allow for the extraction of lineaments by using shaded-relief process by changing the illumination azimuth (Smith and Clark 2005; Ahmed et al. 2017).

3.2 Analysis of linear features

In the selected areas of the Ethiopian Highlands we mapped the linearly arranged elements of the landscape, for example linear sections of valleys, ridges or straight sections of slopes. These linear landforms and structures were manually mapped from: a) digital elevation model (hillshade images of various elevation angles of the illumination source and solar azimuths), and b) satellite images of Google Earth (pixel size of 15 m, scale 1 : 100,000), and subsequent overlapping of the layers (according to Wladis 1999; Ehlen 2004).

Many of these long linear structures are parallel to each other and the processes forming the current landscape are concentrated there. In order to determine whether certain parallel linear structures are of the same age and were created by the same process (tectonic activity and erosion), the parallel mega-lineamentswere marked with the same colour (red, green, blue, etc.). The shortest length of the mega-lineamentswas 20 km. Shorter linear structures (length < 20 km) are called lineaments and thousands of them are seen on the Ethiopian Highlands. We mapped the linearly arranged elements of the landscape (the mega-lineamentsand lineaments) for an analysis of the main directions of the landscape forming processes. The faults in the Ethiopian Highlands were taken from the geological map of Ethiopia (scale 1 : 250,000; Mangesha et al. 1996).

A valley network model for the Ethiopian Highlands was extracted from SRTM DEM by following method of Jenson and Domingue (1988) and using ArcHydro tools in ArcGIS 10.2.: *Fill* → *Flow Direction* → *Flow accumulation* (ESRI 2015), i.e. the valley network was identified by analysing the flow accumulation after a filling function that was applied to remove sink-like artefacts caused by noise and inaccuracies in the original DEM dataset. The threshold of contributing area for the valley network generation was 1,000 pixels (0.9 km^2). For the analysis of the azimuth of stream channels, we converted the valley network raster into a vector format.

The azimuths of megalineaments, lineaments, and faults were determined in terms of the orientation of lines to the coordinate system and the azimuth of stream channels was also determined as the orientation of the stream channels (i.e. parts of valleys from the valley heads to the first point of the valley confluence; or parts of valleys between two valley connection points) to the coordinate system. The azimuths were illustrated by rose diagrams, which are divided into 72 intervals of 5° (360° in total). The numbers of lines in the rose diagrams were multiplied (weighted) by their length according to Belisario et al. (1999), Ciotoli et al. (2003) and Kusák et al. (2016).

Moreover, several types of profiles were created: 1) valley cross-sections which are in figures depicted by different colours (e.g. green, red) to determine the erosion intensity and (if possible) the relative age; 2) long profiles in the landscape to reveal the possibilities of river piracy; 3) profiles perpendicularly to the MER margin to reveal its block structures.

4. Results

4.1 Aggregated characteristics

In the eastern part of the Blue Nile Basin the old plateaus are strongly eroded by fluvial erosion and they recede in favour of expanding deep valleys. The fragments of the plateaus consist of Paleocene-Oligocene-Miocene basalts and tuffs and deeply cut valleys are also located there. In the western part of the Blue Nile Basin, Precambrian rocks are exposed (Figure 2).

A total of 161 mega-lineaments(with an average length of 60 km), 1,264 faults (with an average length of 21 km), 3,429 lineaments (with an average length of 10 km) and 81,137 valleys segments (with an average length of 2.3 km) were mapped and analysed in the upper part of the Blue Nile Basin. The results were then compared using azimuths and rose diagrams (Figure 3). The analysis revealed that the mega-lineaments(Figure 3A,B) and faults (Figure 3C, D) are strongly correlated and concentrated into two directions: azimuth 20° (NNE-SSW orientation; red lines in Figure 3A) and azimuth 40° (NE-SW orientation; green colour in Figure 3A). The third-most dominant direction being E-W (azimuth 90°) is clearly visible in the lineaments. It has not a significant reflection in mapped faults, however Abebe et al. (1998) and Gani et al. (2009) also mentioned important faults of this direction. Smaller concentrations of linear features were also identified in the following directions: ENE-WSW (azimuth 80°; blue lines), WNW-ESE (azimuth 115°; brown lines), NW-SE (azimuth 145°; orange lines), NNW-SSE (azimuth 170°; violet lines) and N-S (azimuth 0°, black lines). Not all these directions are clearly reflected in aggregated faults but can have a fault reflection at a local level.

4.2 Local scale analysis

The sub-regions 1–5 from Figure 1 were chosen in order to represent various types of landscapes in the upper part of the Blue Nile River Basin. Sub-regions 1 (covering the area of 18,800 km^2) and 2 (11,500 km^2) are on the boundary of the Ethiopian Highlands close to the MER. Sub-region 3 (20,500 km^2) is a typical area

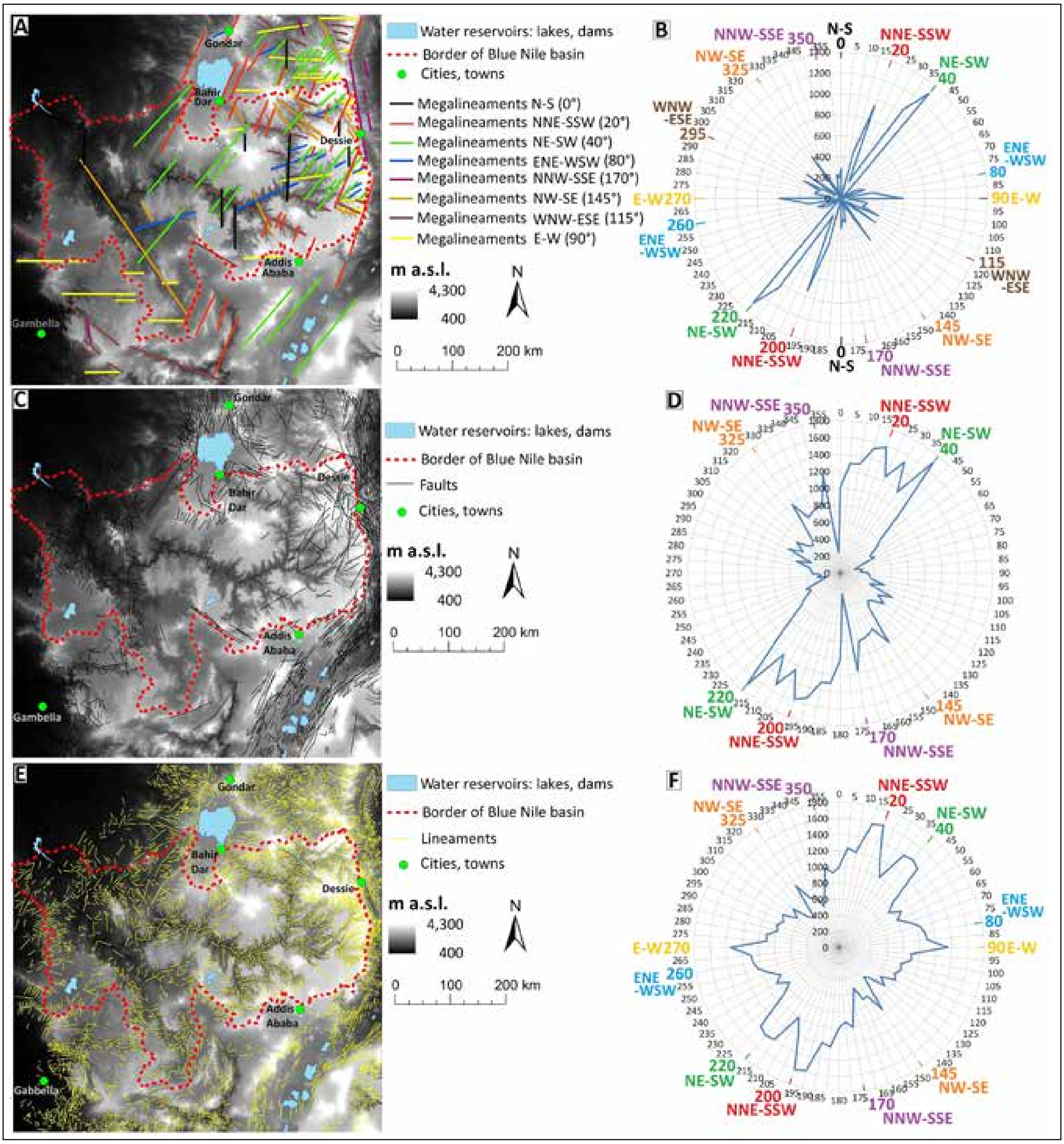

Fig. 3 A) Map of megalineaments; B) rose diagram of megalineaments; C) map of faults; D) rose diagram of fault azimuths; E) map of lineaments; F) rose diagram of lineament azimuths.

adjacent to the central volcano, while sub-region 4 (21,000 km^2) represents the Lake Tana area. Sub-region 5 (35,000 km^2) in the southwest represents a highly-eroded part (i.e. the lowest part) of the Ethiopian Highlands.

It is evident that the NE-SW direction (green lines) is important in the sense of the valley evolution. This is true across all of the sub-regions, regardless of the type of relief, and could represent an old structural predisposition, which influenced the valley network evolution in the Ethiopian Highlands.

4.2.1 The Jemma River

The local scale analysis of sub-region 1 in the catchment of the Jemma River revealed that apart from the importance of the NE-SW orientation (azimuth 40°;

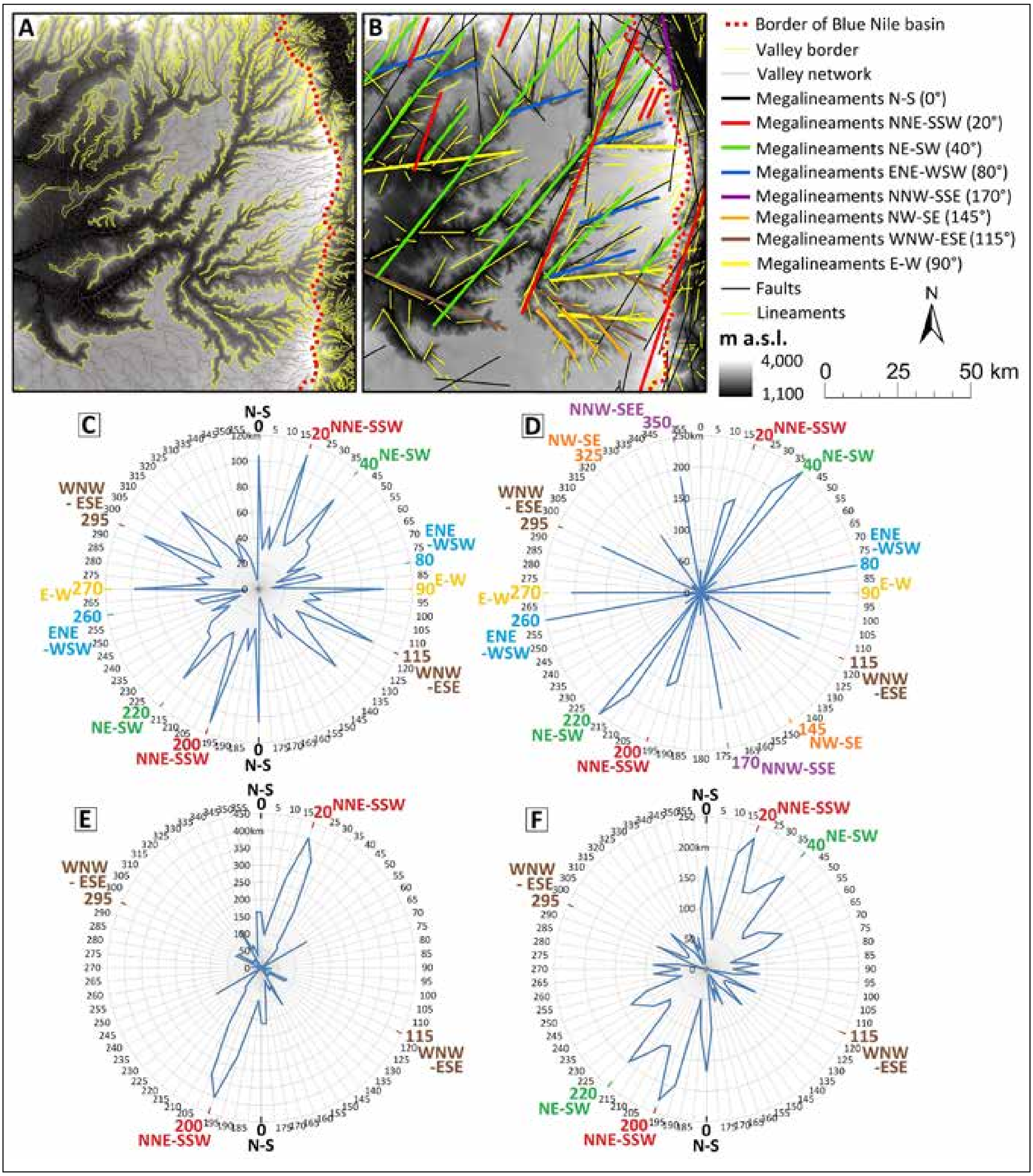

Fig. 4 The sub-region 1 (Jemma River Basin sub-region): A) map of the valley network and valley border; B) map of megalineaments, faults and lineaments; C) rose diagram of valley network azimuths; D) rose diagram of mega-lineamentsazimuths; E) rose diagram of fault azimuths; F) rose diagram of lineament azimuths.

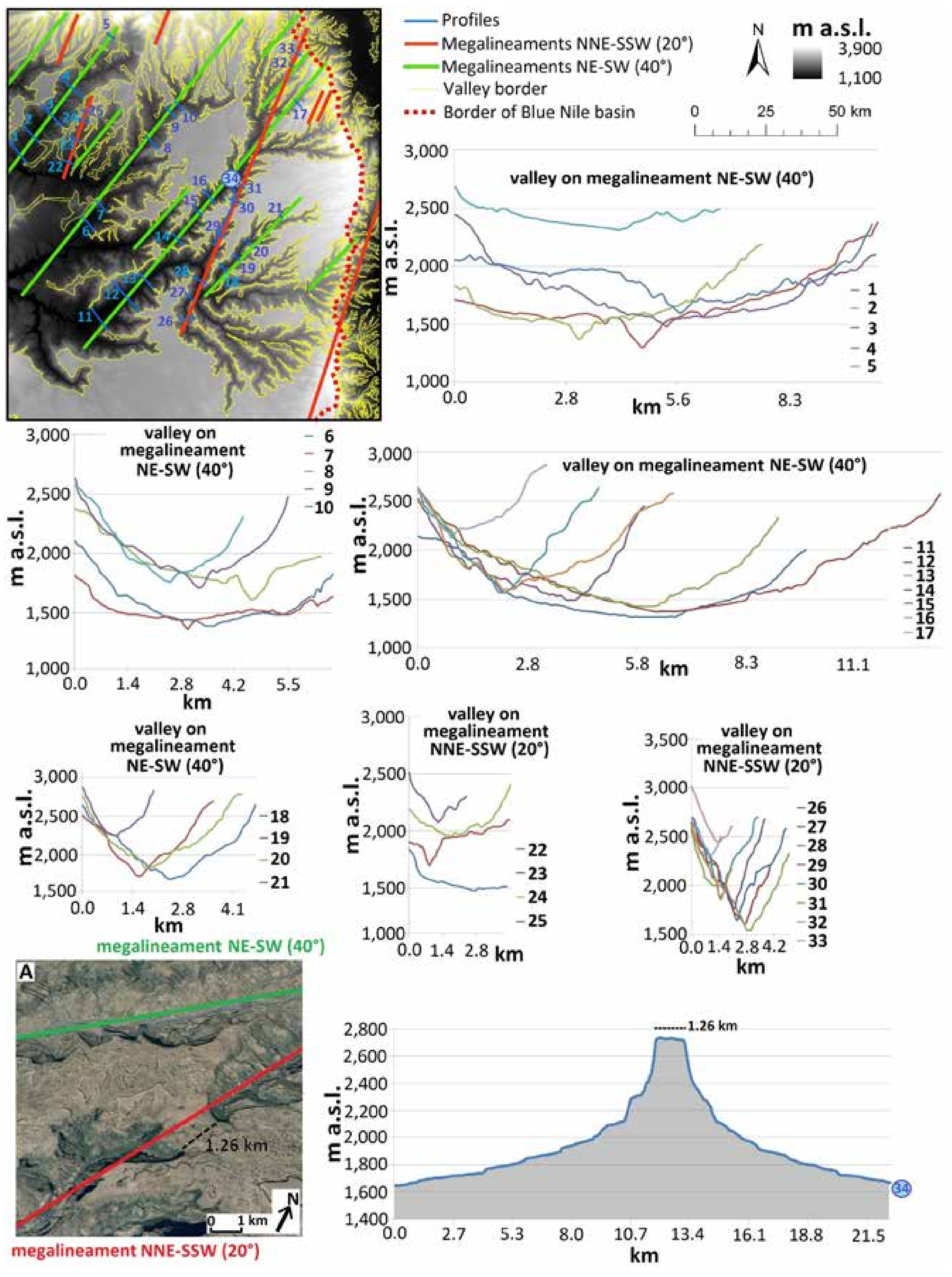

Fig. 5 Profiles in sub-region 1 (Jemma River Basin). Profiles 1–21 are valley cross-sections across green lines and profiles 22–33 are valley cross-sections across red lines; A – longitudinal profile No. 34 along a single red line (a site with a tendency for river piracy).

green line), the NNE-SSW orientation (azimuth 20°; red line) plays a more significant role in the valley network evolution and lineament depiction. The rose diagram for the faults suggests tectonic predisposition. Both directions (green and red) have a similar azimuth (40° and 20°, respectively); however, they are sharply bounded and distinguished in the rose diagrams (Figure 4). The shapes of the red line valley cross-sections suggest more intensive downward erosion compared to the green line valley cross-sections, which suggests that the red line direction is younger (Figure 5). One site with a tendency to river piracy was also positively identified along one of the red lines (Figure 5A). The headward erosion is strong

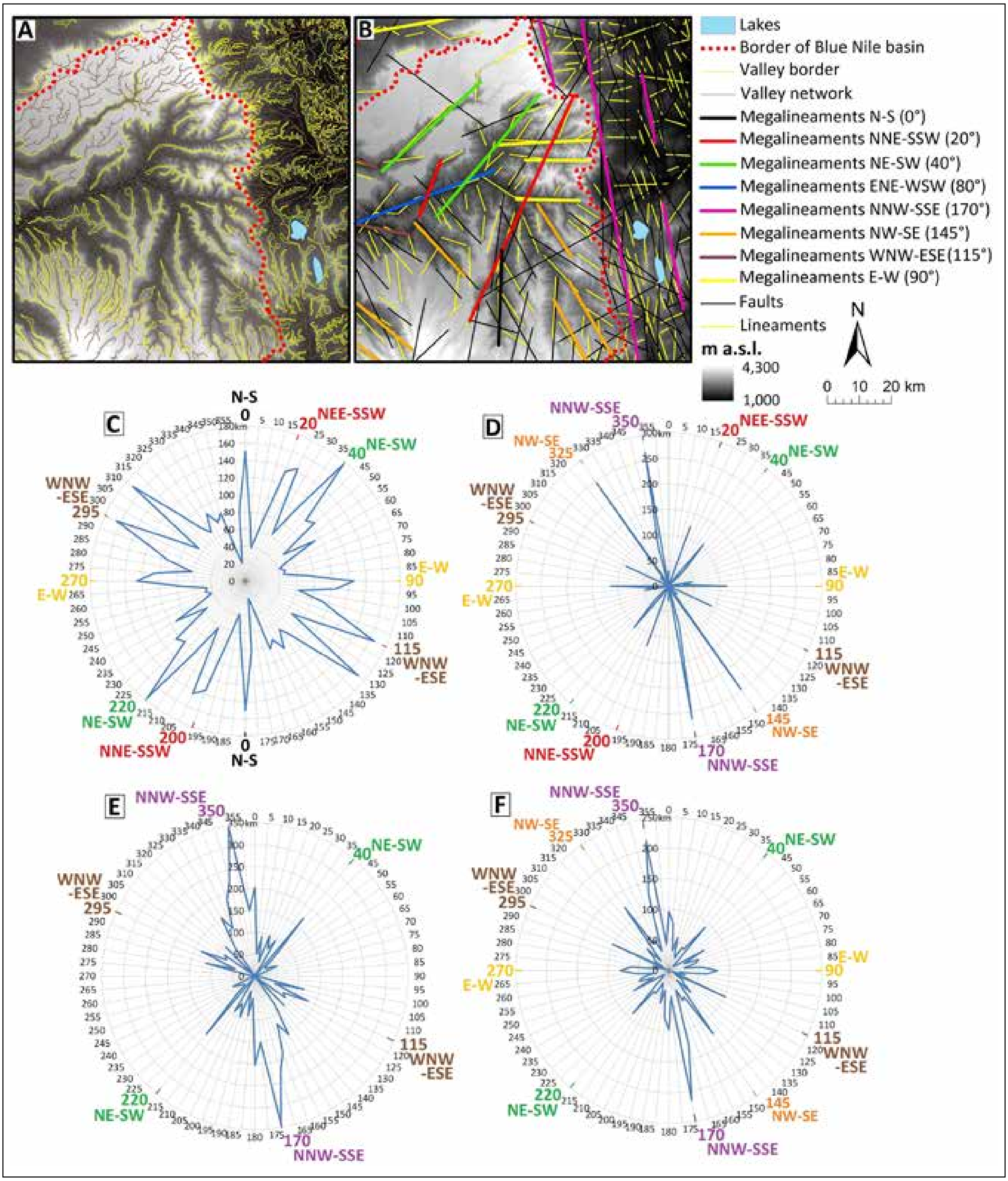

Fig. 6 Sub-region 2 (Beshlo Wenz River Basin sub-region): A) map of the valley network and valley border; B) map of megalineaments, faults and lineaments; C) rose diagram of valley network azimuths; D) rose diagram of mega-lineamentsazimuths; E) rose diagram of fault azimuths; F) rose diagram of lineament azimuths.

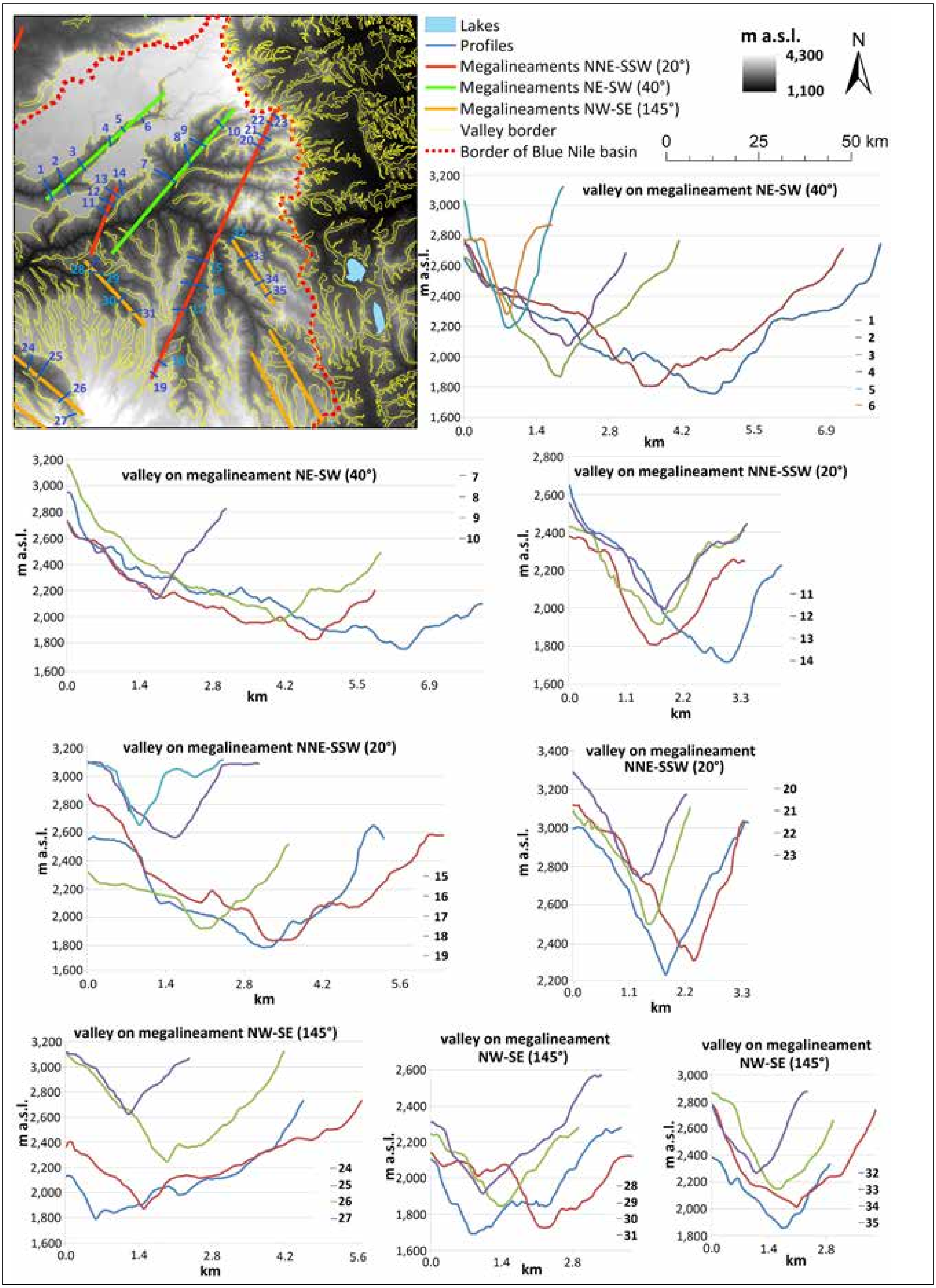

Fig. 7 Profiles in the Beshlo Wenz River Basin sub-region. Note: Profiles 1–10 are cross-sectional profiles directed across the valley marked by green lines; and 11–23 are cross-sectional profiles directed across the valley marked by red lines. Profiles 24–35 are cross-sectional profiles directed across the valley marked by an orange line.

from both sides of the two different valleys and river piracy can be expected to happen soon because the remnant of the original relief created by the basalt layer is only 1.3 km wide.

It is clear that the shape of the cross-sectional profile is influenced by several processes (or predispositions), including intensity and type of erosion, discharge and lithology; nevertheless, the set of profiles will reveal the general trend in valley evolution (if the lithology is homogeneous), which can be compared to other sub-regions. Valley cross-sections along the green lines (profiles 1–21 in Figure 5) are wider and more open compared to the profiles along the red lines (especially profiles 26–33), which are more narrow and tighter. This is regardless of the position of the valley in the river network system i.e. profiles of one direction are similar in the upper, middle and lower parts of the valley system.

4.2.2 Beshlo Wenz River

The valley network predisposition of the landscape evolution in the Beshlo Wenz Basin (sub-region 2) is similar to the Jemma River Basin. However, the faults do not have an azimuth of 20°, but the NNW-SSE orientation (azimuth 170°; violet colour on the graphs) is well pronounced (Figure 6). The Afar area adjacent to the Blue Nile Basin, is also included in this sub-region. The Rift has a NNW-SSE direction herein. Border faults (violet colour) are reflected in lineaments and megalineaments. The rose diagram for the valley network is rather complex because the western part is similar to sub-region 1 (Jemma River Basin), while the eastern part bears significant features of the slope facing the MER.

In accordance with sub-region 1, the valleys in the Beshlo Wenz catchment lying along green lines, are again more open and wide, while those running along the red lines are markedly incised (Figure 7). The only exceptions are profiles 3 through 6, and 10 (closest to the rift, Figure 7), which are also very narrow like those connected to the red lines. The area next to the border of the MER is under very intensive young evolution (strong downward erosion) regardless of the type of valley (red or green lines).

Several tectonic blocks evolved on the east-facing slope of the rift, which are asymmetrical in their cross-sectional profile and retain small lakes (Figure 8). The drainage in a NNW-SSE orientation is modified by strong headward erosion along the steep, east-facing fault slope. In our fieldwork, we identified the case of

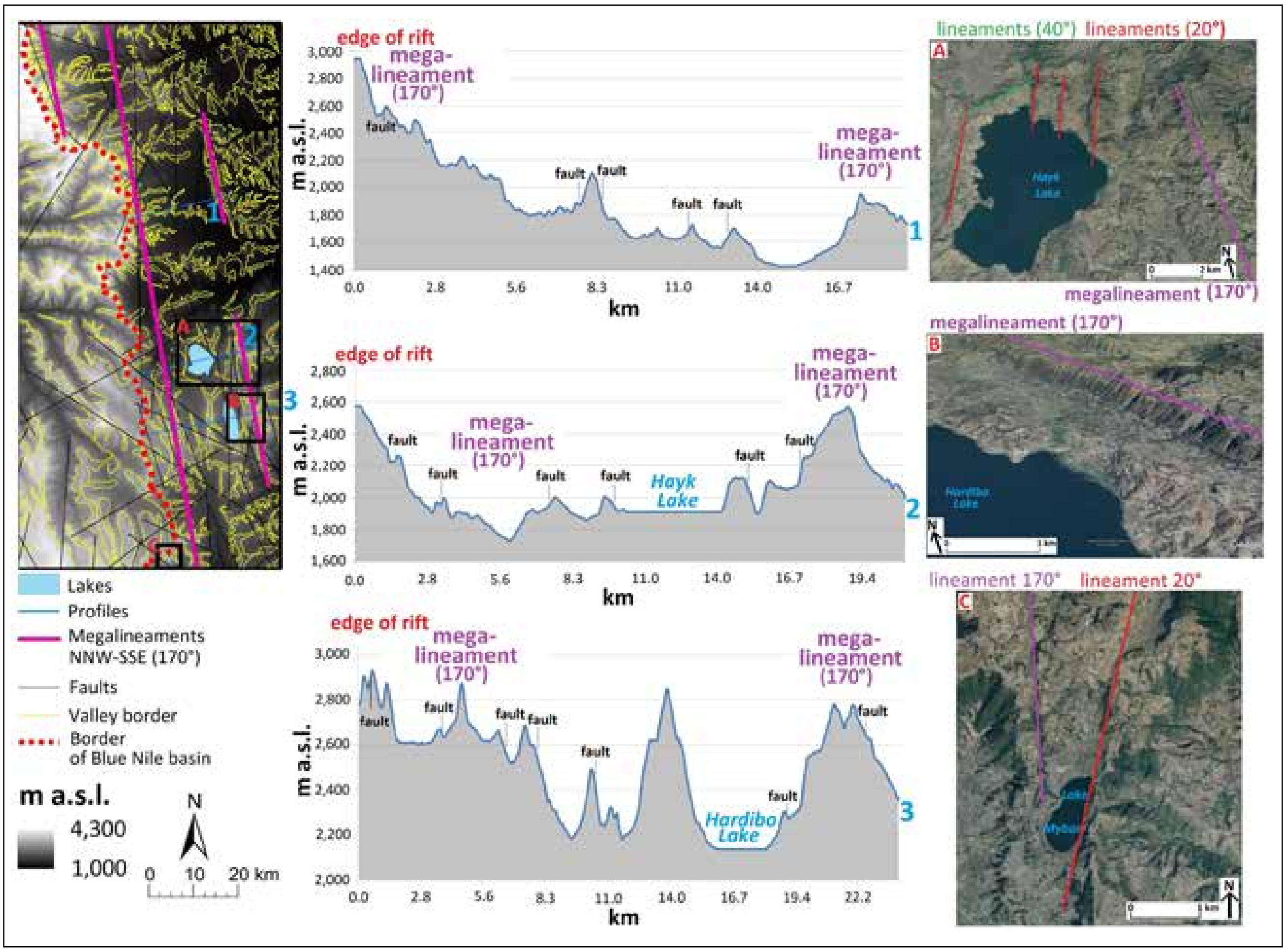

Fig. 8 Cross-sectional profiles oriented towards the MER. Tectonic blocks, which are asymmetrical in the cross-sectional profiles evolved here and retain Lake Hayk (A), Lake Hardibo (B) and Lake Mybar (C).

Lake Mybar (Figure 8C) – the lake is drained to the east nowadays because a small river reached the lake shore due to very intensive headward erosion and the former drainage to SSE was abandoned.

In this locality the tectonic predisposition is represented by faults with an azimuth of 20°. This is also the area with highest frequency of faults. Nevertheless, 30 km to the north the tectonic predisposition is changing in connection with the main structures of the rift. Lake Hardibo (Figure 8B) is predisposed by tectonics with an azimuth of 170° (NNW-SSE orientation) depicted by a violet colour in Figure 6 and 8, while in the surrounding of the adjacent Lake Hayk (Figure 8A) both NNE-SSW and NNW-SSE trending faults were identified (red and violet lines, respectively).

4.2.3 Volcanoes Mt. Choke and Mt. Gish

The situation around the volcanoes of Mt. Choke (4,100 m a.s.l.) and Mt. Gish (2,890 m a.s.l.) is rather

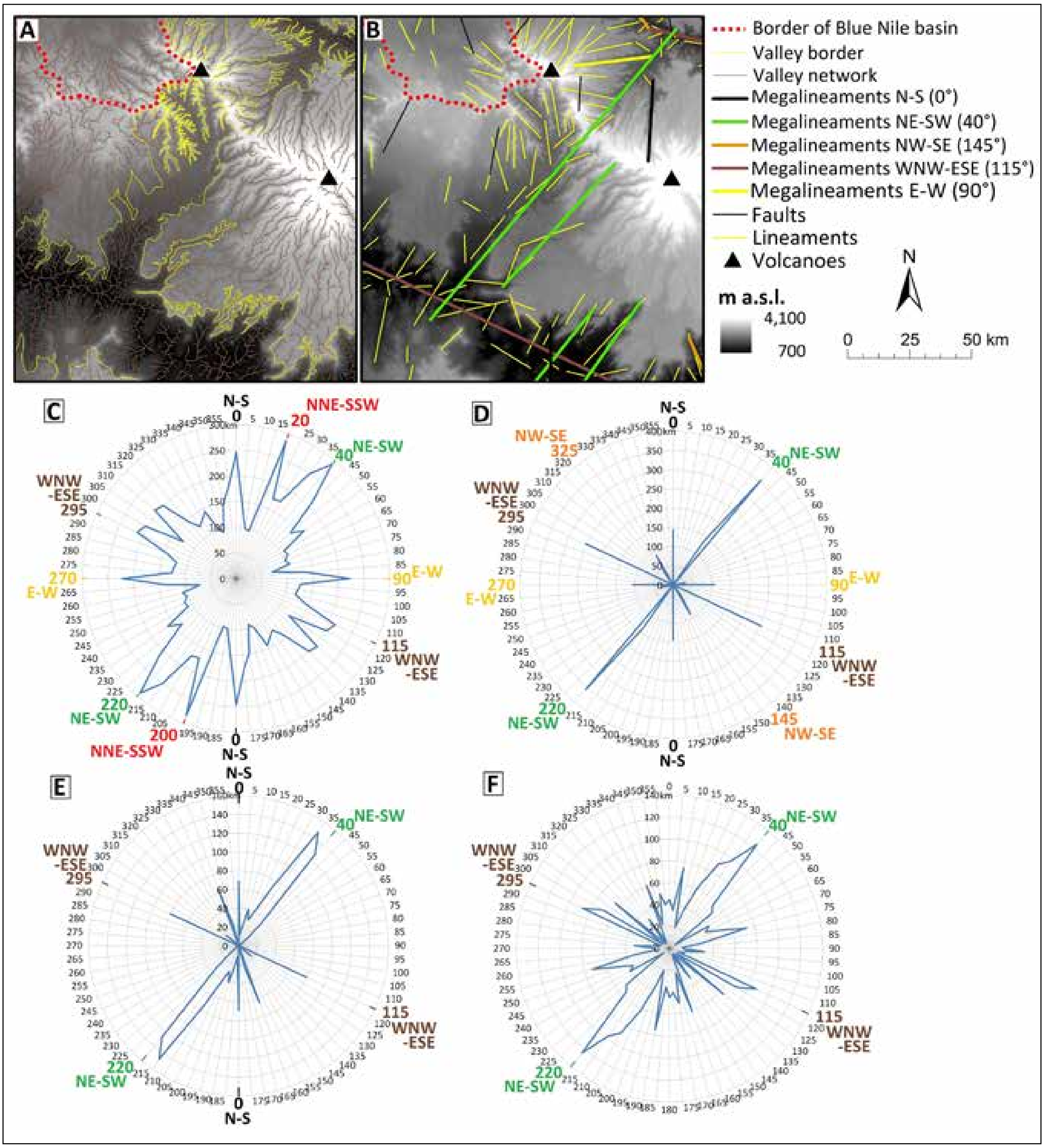

Fig. 9 Sub-region 3 (Mt. Choke and Mt. Gish volcano sub-region): A) map of the valley network and valley border; B) map of megalineaments, faults and lineaments; C) rose diagram of valley network azimuths; D) rose diagram of mega-lineaments azimuths; E) rose diagram of fault azimuths; F) rose diagram of lineament azimuths.

simple. The valley system reflects the position of the volcanoes in this sub-region (radial network with preferences towards certain directions); nevertheless, all other rose diagrams show NE-SW trending faults, lineaments and mega-lineamentswithout any influence of the younger NNE-SSW tectonics (Figure 10).

The long valley system between the volcanoes of Mt. Choke and Mt. Gish is probably a part of an older river network, which is pronounced in this part of the Ethiopian Highlands due to the NE-SW trending mega-lineaments identified in all of the sub-regions (Figure 10). Position on the boundary between the LT and HT provinces of flood basalts (Pik et al. 1998; Kieffer et al. 2004) could also play a role. However, parallel to this NE-SW oriented valley, a younger, deeper and narrower valley has been created on the lower slopes of Mt. Choke, which were dated by Kieffer et al. (2004) to an age of 22.4 Ma. This direction

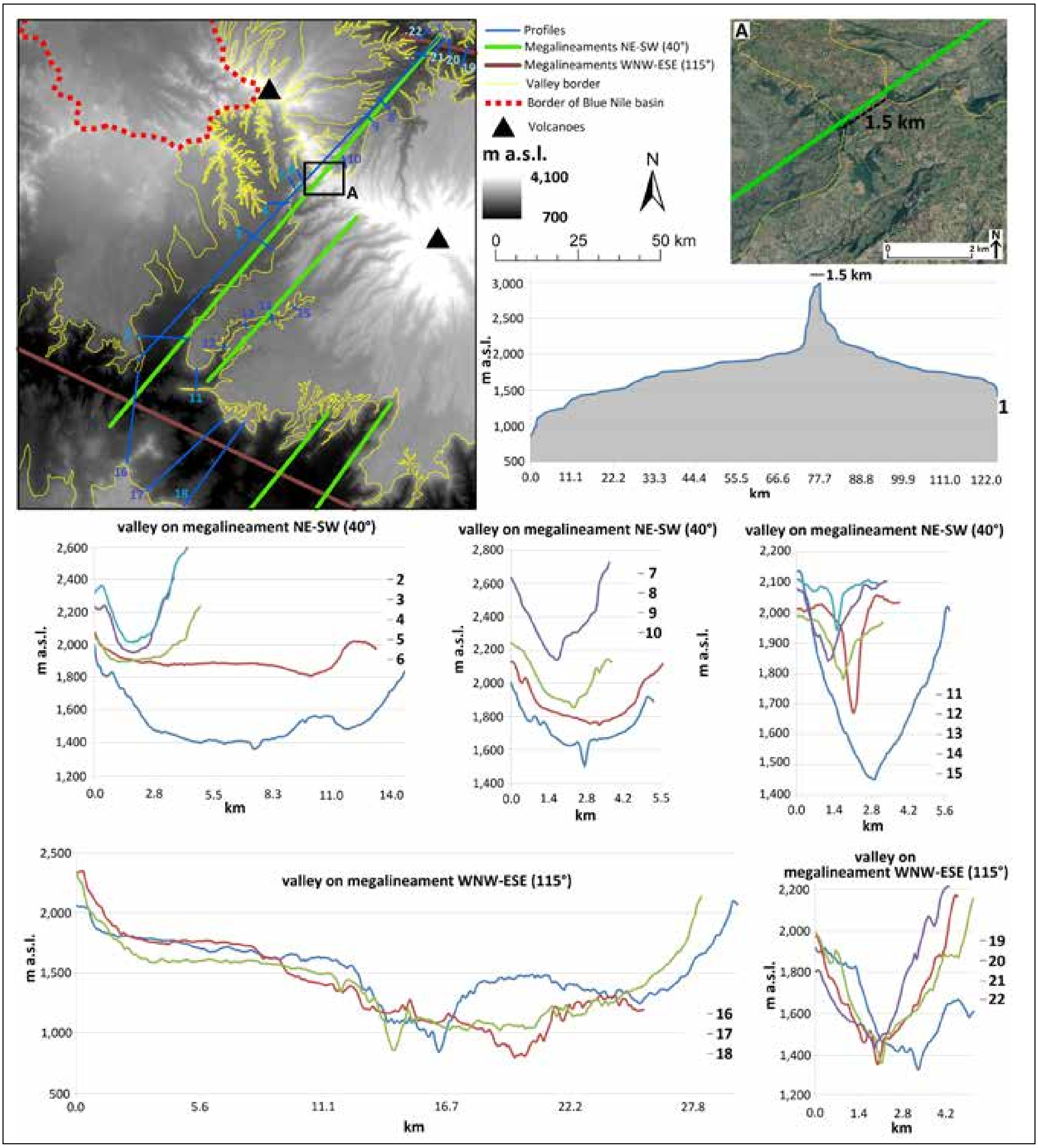

Fig. 10 Profiles in the Mt. Choke and Mt. Gish volcano sub-region. Note: Profile 1 along the single green line (an area with a tendency for river piracy); Profiles 2–15 are cross-sectional profiles directed across the valley and marked by green lines; Profiles 16–22 are cross-sectional profiles directed across the valley and marked by brown lines.

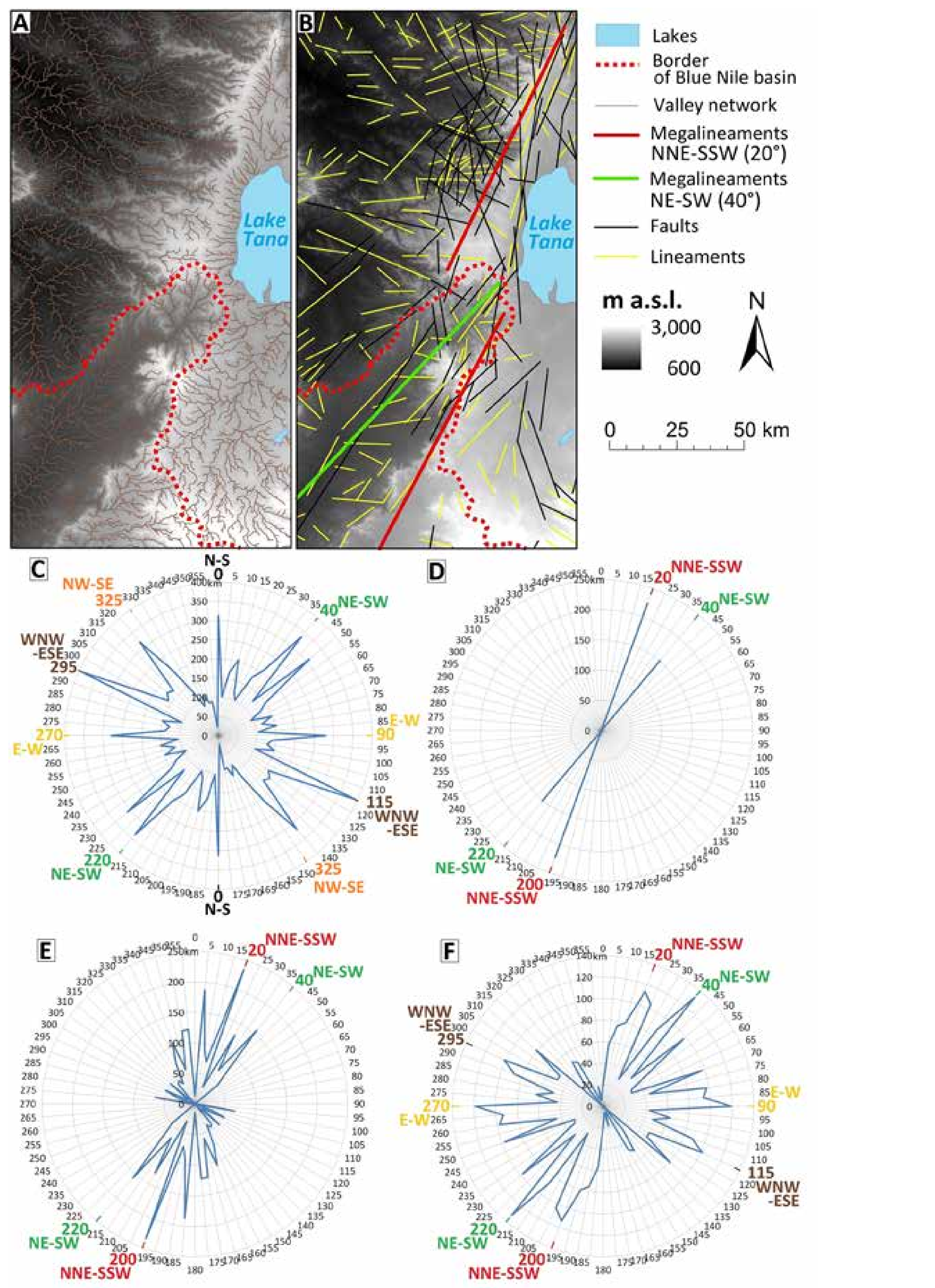

Fig. 11 Sub-region 4 (Lake Tana sub-region): A) map of the valley network and valley border; B) map of megalineaments, faults and lineaments; C) rose diagram of valley network azimuths; D) rose diagram of mega-lineamentsazimuths; E) rose diagram of fault azimuths (E); rose diagram of lineament azimuths (F).

seems to be dominant in the area of Mt. Choke and Mt. Gish, even though the river network should have a radial character around the volcanos. If we consider the fact that the NE-SW valleys (green lines) also cut 10.7 Ma old basalts, then they have to be younger than the basalts itself.

Another example of a tendency for river piracy was identified here (Figure 10A). Less than 1.5 km of the original landscape was left before the rivers from the SW or NE came across the old lava flows. Here we can presume expansion of a SW valley due to the profile (Fig. 10, profile 1).

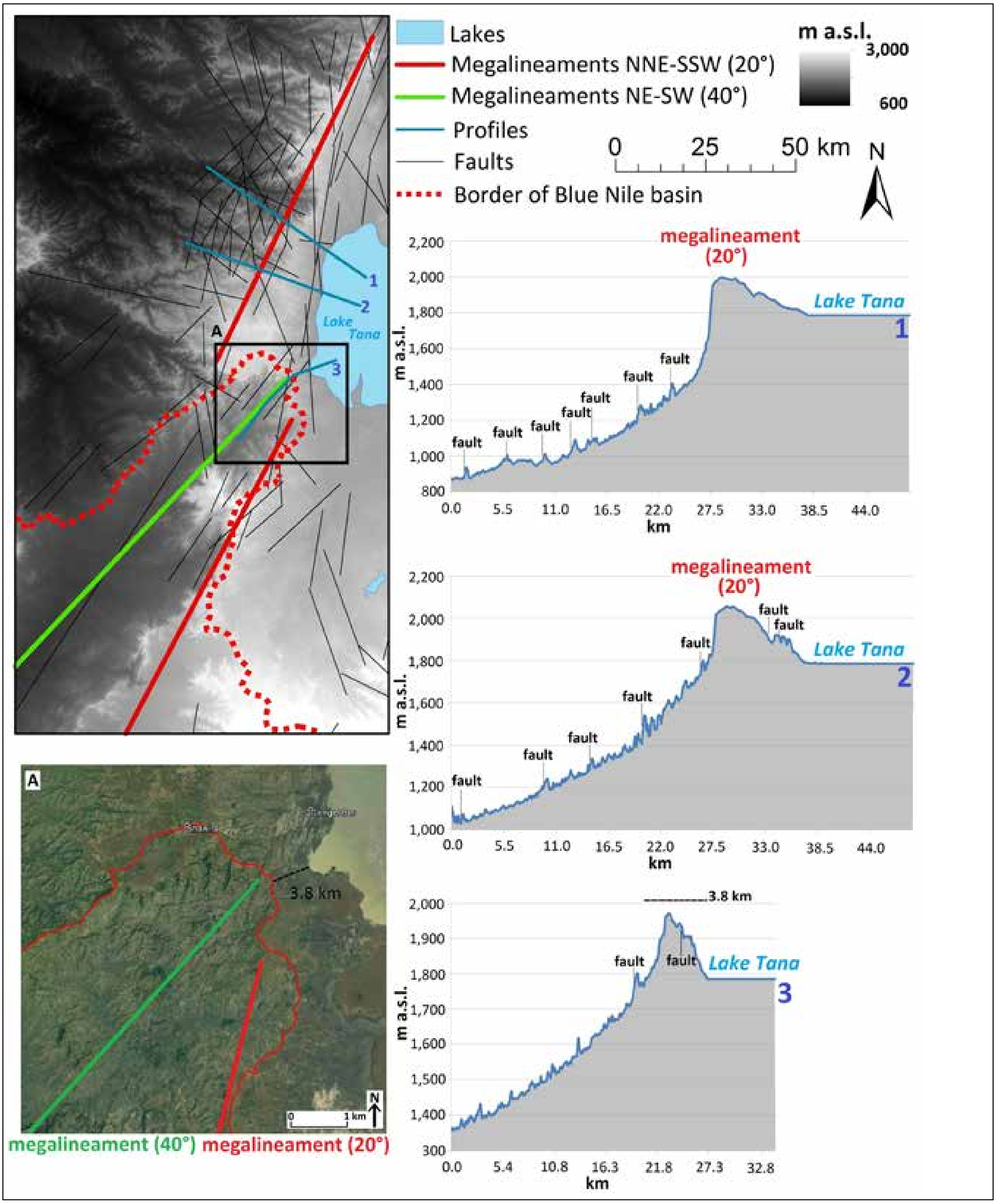

Fig. 12 Longitudinal profiles in the Lake Tana sub-region: Longitudinal profiles 1 and 2 show the situation for rivers stretching WNW-ESE; Longitudinal profile 3 shows the situation for rivers stretching SW-NE.

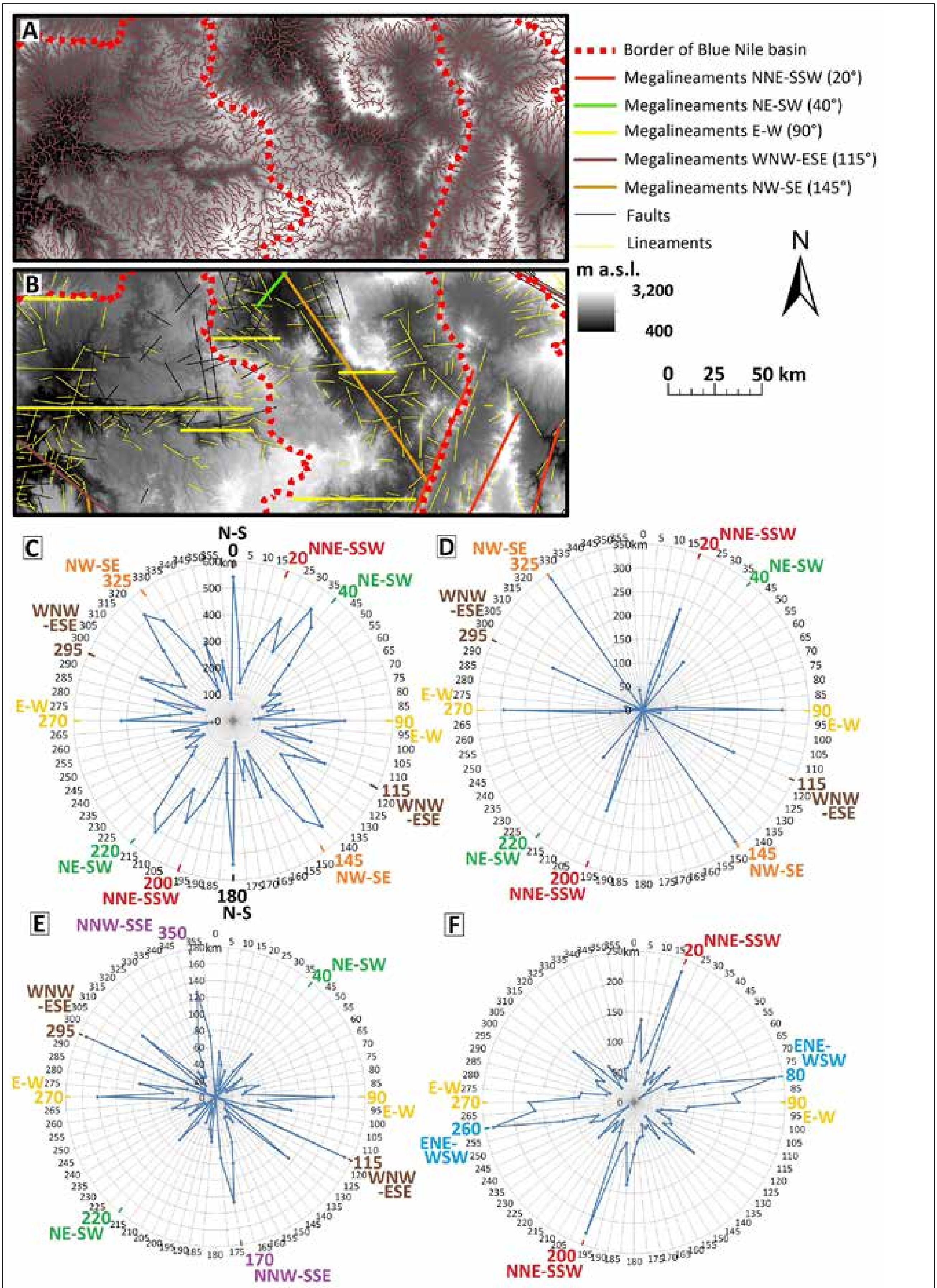

Fig. 13 Sub-region 5 (Didessa River and Baro River): A) map of the valley network and valley border; B) map of megalineaments, faults and lineaments; C) rose diagram of valley network azimuths; D) rose diagram of mega-lineamentsazimuths; E) rose diagram of fault azimuths (E); rose diagram of lineament azimuths (F).

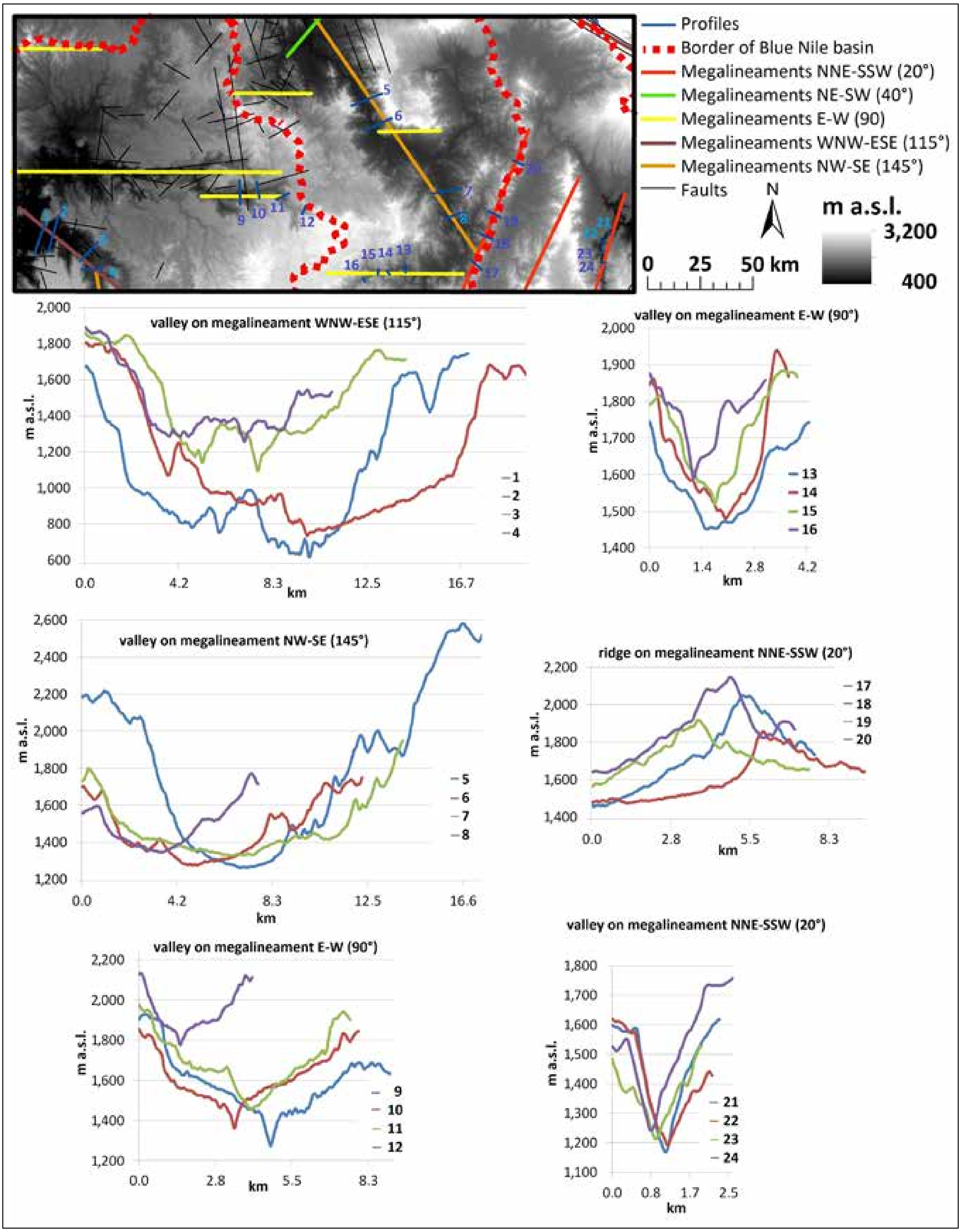

Fig. 14 Profiles in the Didessa River and Baro River sub-region. Note: Profiles 1–4 are cross-sectional profiles directed across the valley marked by a brown line; and profiles 5–8 are cross-sectional profiles directed across the valley marked by an orange line. Profiles 9–16 are cross-sectional profiles directed across the valley marked by a yellow line. Profiles 17–24 are cross-sectional profiles directed across the valley and ridge marked by a red line.

4.2.4 Lake Tana

The valley network seems to be only partly influenced by tectonics, namely in a NNE-SSW direction (azimuth of 20°). The valleys trending WNW-ESE (azimuth of 115°) do not show any connection to the mapped faults and are significant for the area to the west of the Lake Tana Basin. Nevertheless, both directions are marked because they represent a serious potential for river piracy (see the paragraph below). No significant recent erosional activity can be shown in a NNE-SSW direction (red lines), like in other sub-regions closer to the MER. It seems that the influence of the MER is not well pronounced in this part of the Ethiopian Highlands. Nevertheless, NNE-SSW faults exist in this area, and they influence both the lineaments and the megalineaments.

The whole of the Lake Tana Basin, which is tectonically predisposed (e.g. Chorowicz et al. 1998), now drains into the Blue Nile to the SE. Rivers on the western margin of the basin are approaching the lake due to headward erosion from two different directions (Figure 12). Longitudinal profiles 1 and 2 in Fig. 12 show the situation for rivers stretching WNW-ESE (brown lines on the rose diagram in Figure 11). The headwaters are approximately 10 km away from the lake; however, the headward erosion has to remove a 200 m high crest (escarpment) forming the lakeshore in the west (Sembroni et al. 2016). Another situation in the SW of the lake shows, river piracy taking place in a SW to NE direction (along the green lines) and the remnant, which must be crossed, is only 3.8 km long. Headward erosion along this green line is more probable because it is predisposed by the fault. The landscape on the lakeshore is also flatter and is created by a 150-m high remnant of lava flow from Mt. Gish (2,890 m a.s.l.).

Currently, the area of the Blue Nile basin covers 175,400 km^2 (between Lake Tana and the Roseires water reservoir). However, when Lake Tana will change the drainage to the SW, the cut off part of the Blue Nile Basin (98,600 km^2) will not be supplied by water from Lake Tana, but only by precipitation, which is not evenly distributed during the year. The annual rainfall in the Ethiopian Highlands is 1,800 mm and 80% of this precipitation falls from July to September (Klimadiagramme weltweit 2016).

4.2.5 Didessa River and Baro River

Many of the valleys (Figure 13) follow either a NW-SE or NE-SW direction where the valley bottoms are

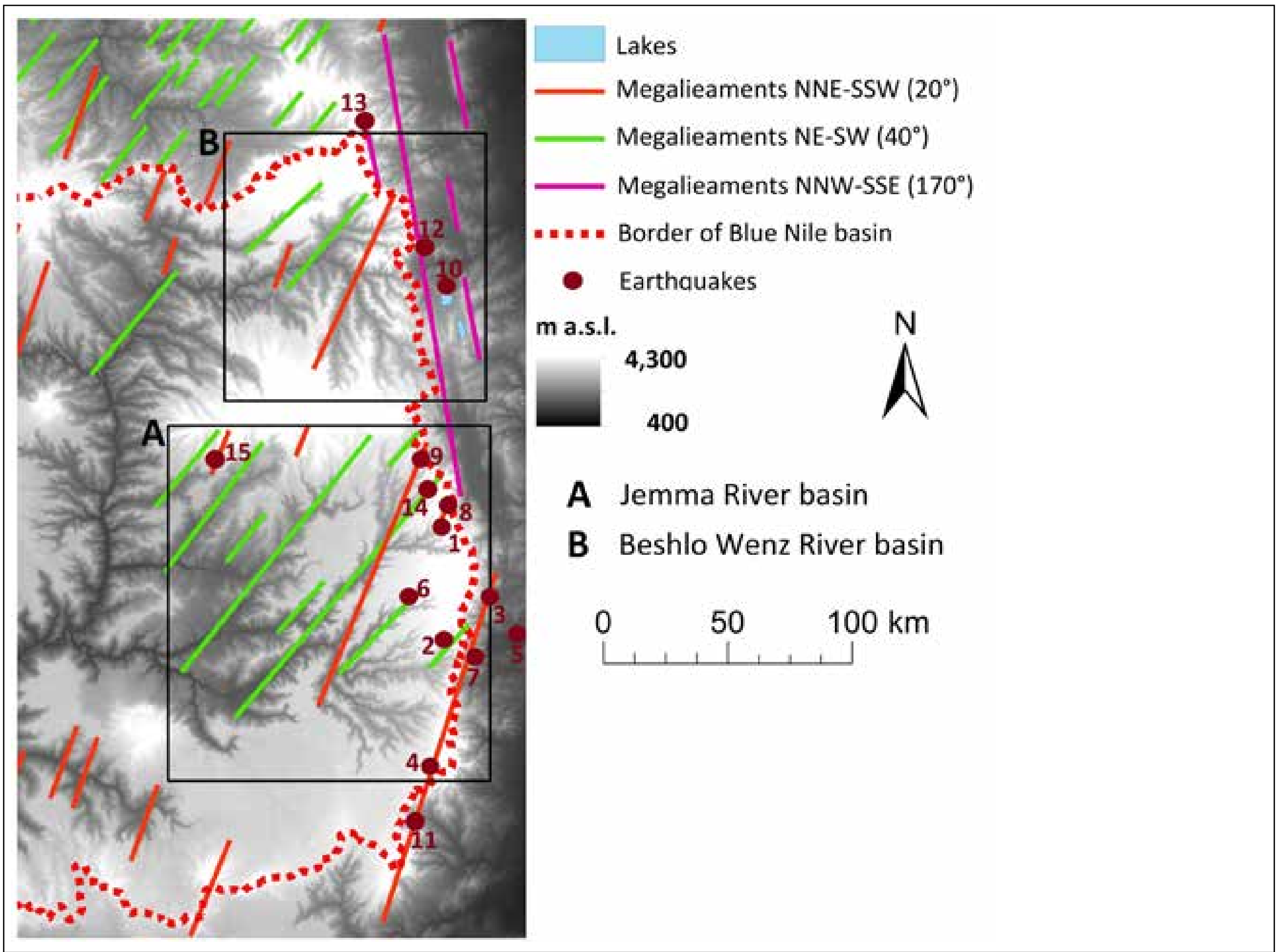

Fig. 15 Earthquake epicentres in the Jemma River Basin (A) and in the Beshlo Wenz Basin (B) (Earthquake Track 2017).

wider, which suggests an older river network. If we check the cross-sectional profiles (1–8) that stretch across brown and orange lines (azimuth 115° and 145°, respectively), the shapes resemble wide and open valleys (Figure 14). In addition, they are rather deep, which suggests a long period of evolution. A slightly different situation can be observed on the profiles crossing the yellow lines, whereby young erosion is already visible on the valley bottoms. An opposite situation to these valley profiles can be found in the valleys depicted by the red lines (profiles 21–24 in Fig. 14), which are again very deep, young and narrow (similar to Figure 7). The area of the Didessa and Baro rivers seems to be older in the sense of the evolution of the valley network with young erosion in the direction of the red lines (azimuth 20°).

4.3 Analysis of earthquake epicentres

A total of 12 earthquakes have occurred in and around the Jemma River Basin since 1961 (Figure 15A; Earthquake Track 2017). Most of the epicentres relate to "red lines" (eight events), only three to "green lines". In the Beshlo Wenz Basin (Figure 15B) the epicentres are located on faults with an azimuth of 170° (NNW-SSE orientation) depicted by a violet colour. This is due to a logical influence of the MER in this catchment. The average depth of the hypocentres is 20 km, which indicates rather shallow earthquakes and the average magnitude is 5.5.

4.4 Features of Guder River and Muger River network reorganisation

Features of drainage network reorganization along lineaments of various types can be detected in various parts of the Blue Nile Basin. The drainage networks of the Guder and Muger catchments (Figure 1*A*, Figure 16) are organised dominantly along diagonal green and orange lines; however, the most active recent river incision is along axial lines (yellow, violet, black and red, see Figure 16). This activity is evident mainly in lower parts of both watersheds as well as in the adjacent part of the Blue Nile valley. On the other hand, valleys along diagonal lines are wider, more denuded and are partly abandoned. A set of highly elevated depressions (2,600–3,000 m a.s.l.) elongated in a NNW-SSE direction (orange lines) can be identified on the southern watershed to the NW of Addis Ababa. Their very flat and wide bottoms (too large for the

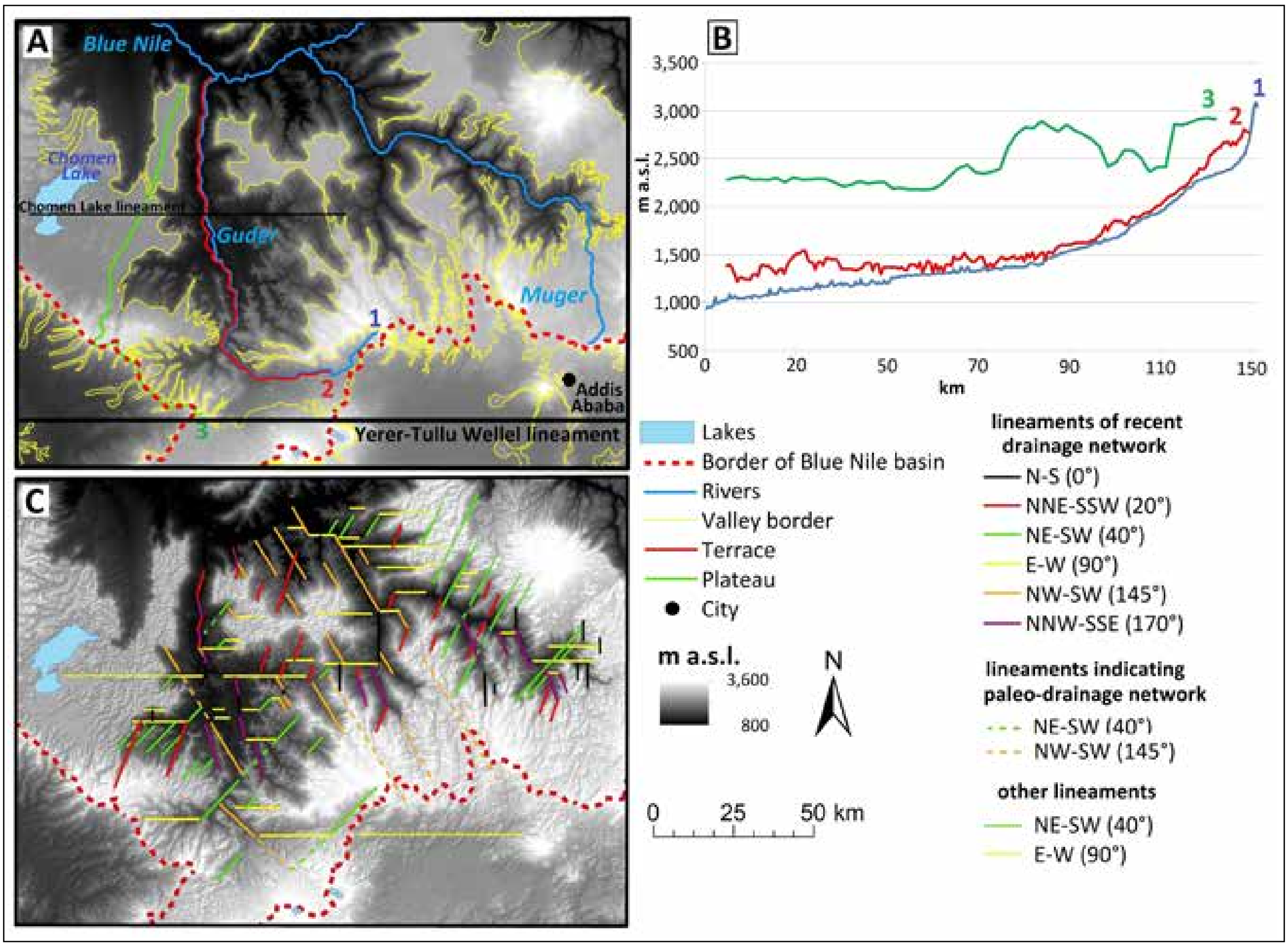

Fig. 16 The Sub-region A (Guder River and Muger River basin sub-region): A) Guder River and Muger River networks organised dominantly along diagonal green and orange lines; B) Profiles of Guga river (1), terrace (2) and plateau (3); C) lineaments of recent drainage network, lineaments indicating paleo-drainage network and other lineaments.

recent small streams) with differing descent (partly to the S and partly to the N) suggest that they belong to an ancient drainage network that might have drained the area towards the SE.

More detailed features of river network reorganisation can be seen in the longitudinal profile of the Guder River (Figure 16B). The river recently rose on the border scarp of the E-W oriented Yerer-Tullu Wellel lineament, below one of the above-mentioned paleo-valleys at an elevation of approximately 3,000 m a.s.l., where only several hundred horizontal meters are missing in order to capture the stream of the paleo-valley by headward erosion. Leaving the scarp, the river flows along the N boundary of the lineament without any notable incision. At an elevation of ~1,800 m a.s.l. it turns to the NW, where it flows across a denuded border scarp in a 0.5 km deep stepped canyon with a small decrease in gradient. The profile only becomes steeper when it crosses the W-E lineament to the east of Lake Chomen and then before it flows into the Blue Nile at an elevation of approximately ~950 m. Along with the typical character of the upper part of the profile, several other morphological features point to reverse flow of the river in the past. The general trend of the plateau inclination (from N to S) has been reversed by a strong uplift of the border of the Yerer-Tullu Wellel lineament (Figure 16); therefore, natural drainage to the S could have existed in the past. The strongly eroded, wide and branched middle part of the watershed contrasts with the narrow lower part closed between well preserved remnants of the volcanic plateau in the north. On the contrary, the uplifted plateau is markedly denuded in the south.

Probably the most interesting feature is the character of a wide terrace (2 in Figure 16B), which is directly incised by the river. The elevation of the terrace rises from the mouth (~1,350 m a.s.l.) only approximately 20 km (~1,450 m a.s.l.) to the narrowest part of the valley bounded by remnants of the plateau. Subsequently (from 20 km to 70 km), the terrace is reversed to recent river as for W-E elevated belt at Lake Chomen lineament. Nevertheless, a small paleo-valley on the southern border sloping to the S is also an indication of reverse flow here. Further to the S, the river is not incised and probably flows down to the terrace (~1,300 m a.s.l.) to the point where a general rise in the elevation of the plateau begins (Figure 16). The former water shed divide, approximately 20 km from the recent mouth, and at least a 50 km long reversed flow of the river can be inferred. Considering all other indices, drainage towards the SSE, mainly along diagonal lines, can be assumed before the uplift of the border of the Yerer-Tullu Wellel lineament. Therefore, reorganization of the drainage network in relation to the current situation took place later by headward erosion, mainly along the axial lines.

5. Discussion

Landscape analysis from SRTM DEM data is of great interest to geoscientists (e.g. Martino et al. 2009; Haider et al. 2015). SRTM DEM data are often used for three types of morphometric analysis: discrete elevation function *f* (x, y, z), hill shading as well as flow direction and flow accumulation. All three were also very useful for performing a morphostructural analysis of the upper Blue Nile Basin and its surroundings.

Hill shading offers many possibilities for interpreting detailed surface structures that are often completely imperceptible using simple DEM visualization (e.g. Kennelly 2008). Hill shading allowed us to perform a much more accurate delineation of lineaments and borders of erosion than mapping from certain satellite images only (Kusák et al. 2016). However, the overlapping of layers (the digital elevation model; hillshade images of various elevation angles of the illumination source and the solar azimuths; satellite images with a pixel size of 15 m placed on the virtual globe of Google Earth) proved to be the most useful when checking the manually mapped landscape shapes (following Wladis 1999; Ehlen 2004; Gani and Abdelsalam 2006). Morphometric analysis based on a calculation of the potential drainage of each pixel of an SRTM DEM shows the shape, density and hierarchy of valley networks, which reflect the influence of structure and erosion processes on the formation of the landscape (*sensu* Stoddart 1997; Kusák et al. 2016), similarly to an evaluation of various topographic profiles. A set of all of these analyses was effectively used in the central part of the area by Gani and Abdelsalam (2006) but our work brings a different type of information.

The character of faults, their direction and space differentiation as well as time of formation and reactivation, are important for the reconstruction of the morphotectonic development. However, direct investigation of faults is limited to outcrops, boreholes and geophysical profiles, which is time demanding. Therefore, they are not mapped equally in detail on geological maps and the distinction of their origin and activation is limited. On the other hand, lineaments reflecting the geological fractures can only be uniformly mapped with a certain probability on the basis of a DEM. Mega-lineaments are most probably connected with important faults, but because of their limited number they are not as suitable for statistical analysis as regular lineaments. The stages of development of structurally determined landforms can also be estimated from topographic profiles and the presence of morphodynamic phenomena (erosion, landslides, earthquakes) (Gani and Neupne 2018). Finally, the combined information from all of these sources can be compared to the results of various geological and geochronological analyses to obtain a comprehensive picture of the morphotectonic development.

Two basic groups of structural features can be distinguished based on our analysis of the upper part of the Blue Nile basin:

a) Axial (including black, yellow, violet, blue and red lines) around N-S and E-W directions.
b) Diagonal (including green, orange and brown lines) around NE-SW and NW-SE directions.

The difference between the red and green lines (representatives of the basic groups mostly contained in the faults and lineaments) is associated with a transition from a 130°E-oriented extension to a 105°E-oriented extension of the MER sometime between 6.6 to 3 Ma ago, which caused a change in the normal fault orientations from predominantly N35°E to N10°E (Wolfenden et al. 2004). The diagonal features also reflect yet older development of the Blue Nile Basin, i.e. its formation as a NW-trending rift (Gani et al. 2009; MacGregor 2015). Axial features mainly respond to Quaternary development, which is influenced by E-W and NNE-SSW-oriented extensions related to an oblique opening of the MER and development of E-trending transverse faults as well as E-W-oriented extensions in western Afar (Gani et al. 2009). Landslides linked with axial directions confirm Quaternary activity too (Kycl et al. 2017).

The young Quaternary character of the majority of the axial lines was also confirmed by our analyses. They are characterized by prevailing deep and narrow valleys, fault scarps, earthquake activity and a tendency for recent river piracy. On the contrary, the diagonal lines are generally linked with wider and more open valleys with a higher tendency for meandering or braided channels, a dominant abundance of higher order streams and significant "well-developed" relief features. Reorganisation of the drainage networks of the Guder and Muger rivers from the diagonal to axial directions supports this dichotomy.

The development of lineaments and river networks is generally accelerated by the uplift of the area. The initial Paleogene surface of the volcanic plateau of the Ethiopian Highlands probably formed at elevation significantly lower than now, and then had been decreasing until the Late Miocene (MacGregor 2015). Indeed, Pike et al. (2003) suggest that erosion initiated in the Blue Nile canyon as early as 25–29 Ma ago, but Gani et al. (2009) evaluated only a very gentle incision rate of the Blue Nile up to the Late Miocene, whereas rapid erosion would have started only 6 Ma ago. Generally, the neotectonic uplift of the territory can be connected with the opening of the MER (11 – 6 Ma ago according to various authors – c.f. Gani et al. 2009; Ismail and Abdelsalam 2012) but uplift of the Lake Tana sub-region probably began earlier, i.e. before the end of the mid-Tertiary flood volcanism when the asthenospheric mantle intruded the lithosphere as a result of plume activity, inducing thermal uplift (Chorowicz et al. 1998). Further an initial radial drainage network could have formed including outflow to the S (see section 4.4), which was later blocked by Late Miocene and Pliocene tectono-thermal uplift of the MER shoulders and relative subsidence of the Tana basin (Chorowicz et al. 1998; Sembroni et al. 2016). We can speculate that in this initial stage the upper part of the Blue Nile could have also drained along the ancient NW-SE (orange) or WNW-ESE (brown) structures to the area of the recent Afar depression. An indication of this may be the mature erosional pattern of the E and SE part of the Blue Nile Basin in comparison with the more homogeneous (younger) erosional pattern of the W and SW parts (Gani and Abdelsalam, 2006) and also indications of diagonally-elongated palaeovalleys in the recent watershed around the city of Dessie.

The spatial distribution of the particular types of the studied linear features can be linked with the time-space character of the background tectonic events. Sub-region 1 lies closely westward of the MER, and sub-region 2 directly on the border of the Afar depression. Sub-regions 3 and 4 are far from the rift boundary and the eastern part of sub-region 5 is close to the Rift (Figure 1). The investigated (coloured) directions represent only 44% of the possible drainage courses. Nevertheless, 62% (sub-region 4) to 87% (sub-region 2) of the drainage agrees with the analysed directions. This not only confirms the significant influence of tectonics on the drainage network, but also an increase in this influence around the boundary of the MER and in particular the Afar depression. Axial directions dominate in sub-regions near or on the rift boundary (sub-regions 1, 2 and 5), whereas diagonal directions are comparably significant only in the last two sub-regions.

The brown direction is mainly about the SW part of sub-regions 3 and 5 with the most mature relief and exposed pre-Tertiary bedrock, which suggests its ancient character. The orange direction has a similar character, but its important representation in sub-region 2 and the eastern part of sub-region 5 suggest neotectonic reactivation (Abebe et al. 1998). The green direction is generally the most represented in all of the linear features, mainly in sub-regions 1, 3 and 4. Transformation of tectonic failures into the drainage network indicates the oldest neotectonic event of regional significance, i.e. opening of the MER.

The red lines are second in the order of areal importance (after the green lines) and are connected with the stress field of the MER, but are less pronounced in the inner part of the Ethiopian Highlands. The black and violet lines have only a limited areal influence on the adjacent parts of the MER and Affar (their higher importance was revealed only in sub-region 2), while they are the youngest and have been active for the short period of time. The yellow lines are located in the south of the upper part of the Blue Nile Basin and they are primarily a result of activation of the Yerer-Tullu Wellel lineament in the Late Miocene. Nevertheless, recessing the lower parts of their cross profiles means that the Quaternary tectonic activity increases in an E-W direction.

The Ethiopian Highlands are one of the most tectonically influenced areas in the world (Beyene and Abdelsalam 2005; Gani et al. 2007, 2009), so river erosion and river piracy are very active there. The valley networks are young and water divides change position easily. Thanks to the steeper slopes and a more humid climate inside the MER, the water courses (tributaries of the Awash River) have more erosive power and the Jemma and Beshlo Wenz river basins lose parts of their catchments as the water divide is moving slightly towards the west (Kusák et al., 2016). However, the most significant effect of river piracy would probably be seen in the Lake Tana Basin.

Currently, the Lake Tana Basin is in the focus of many scientific disciplines in terms of: 1) geology and geochronology (Prave et al. 2016); 2) reproducing historic sediment concentrations (Kabaa et al. 2014); 3) chemical analysis of water and sediment (Ahrens et al. 2016); 4) assessing the implications of water harvesting intensification on upstream–downstream ecosystem services (Dile et al. 2016); etc.. All of the above stated research activities and many other authors write about the Lake Tana Basin as a single unit, but do not include the processes taking place in the surroundings. River piracy approaches the Lake Tana Basin from SW to NE (along the green lines) and the remnant, which has to be surpassed, is only 3.8 km long (Figure 12). The potential change of drainage of Lake Tana to the SW (instead of to NE) will greatly influence the hydrological conditions over a rather large area.

6. Conclusion

The results of simple morphostructural analysis from a DEM can be interpreted in high concordance with the present results of stratigraphic, structural, petrologic, tectonic and geochronologic investigations of the Ethiopian Highlands. The azimuths of lineaments, faults and river networks combined with erosional features, cross-sectional and longitudinal valley profiles reveal a change in the dominant stress fields and stages of topography development.

The NW-SE (orange lines) and WNW-ESE (brown lines) linear features correspond with older stress fields organizing the formation of ancient pre-Neogene rifts and transposed onto the volcanic plateau during the first phases of its rise. They probably determined a palaeo-drainage system that could partially flow to the south and east across the recent rift shoulders and are well pronounced in sub-regions with uncovered pre-volcanic bedrock. These directions are characterized by extremely braided valleys with locally brained channels, small remnants of volcanic plateau and features of well-developed relief'. They are typical for the oldest part of Blue Nile and Didesa valley.

The beginning of the neotectonic stage (during which the modern topography and drainage network formed) can be connected to the opening of the MER during the Late Miocene. The most wide-spread NE-SW (green) linear features parallel to MER represent the long term morphotectonic influence of an NW-SE extension in the whole territory. Reactivation of some NW-SE (orange) linear features mainly near the MER boundary match with activity of the offset faults. Developed (wider, more open) valleys and more 'mature relief' are typical for the leading green direction.

Pliocene counter clockwise rotation of the MER stress field led (mainly in the MER surrounding) to the generation of NNE-SSW (red) linear features. The stress field transition was linked also with intensive uplift of the whole territory, especially the rift shoulders that definitely blocked any existing outflow toward S or E. Continuing relative subsidence of the Lake Tana basin, headward erosion and river piracy could contribute to spreading of the Blue Nile basin too. Morphological features such as incised narrow valleys, indices of river piracy, valley asymmetry and the generally 'young age relief' characteristic for the red direction support this interpretation.

N-S (black and violet) linear features represent the youngest Quaternary development of the territory connected with E-W extension in MER and western Afar. They are morphologically least developed and represented by huge fault slopes with tectonic blocks and active youngest upper parts of valleys. Orthogonal features with E-W direction are dominant along older (Late Miocene) Yerer-Tullu Wellel volcanotectonic lineament, however they show evident features of recent reactivation in cross profiles and tendency to river piracy. Accelerating Quaternary uplift created recent very high relief mainly along the boundary with MER and Afar depression and so conditions for intensive headward erosion from the outside the Blue Nile basin that could lead to its reduction by river piracy in the near future.

Acknowledgements

The study was supported by the GACR project "Mass wasting and erosion as an indicator of morphotectonic activity in the Ethiopian Highlands based on remote sensing approaches" (P209/12/J068) and by the Grant Agency of Charles University in Prague, Czech Republic "Morphostructural analysis of Ethiopian Highlands based on remote sensing" (1436314). Jozef Minár was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0054. The SRTM DEM was provided by the U.S. Geological Survey. This work was carried out thanks to support of the long-term conceptual development research organization RVO: 67985891. We wish to thank Craig M. Hampson and Frederick Rooks for language editing.

References

Abdullah, A., Akhir, J.M., Abdullah, I. (2010): Automatic Mapping of Lineaments Using Shaded Relief Images Derived from Digital Elevation Model (DEMs) in the Maran – Sungi Lembing Area. Malaysia. EJGE 15, 949–957.

Abebe, T., Mazzarini, F., Innocenti, F., Manetti, P. (1998): The Yerer-Tullu Wellel volcanotectonic lineament; a transtensional structure in central Ethiopia and the associated magmatic aktivity. Journal of African Earth Sciences 26(1), 135–150, https://doi.org/10.1016/S0899-5362(97)001413.

Ahrens, L., Gashaw, H., Sjoholm, M., Gebrehiwot S. G., Getahun, A., Derbe, E., Bishop, K., Åkerblom, S. (2016): Poly- and perfluoroalkylated substances (PFASs) in water, sediment and fish muscle tissue from Lake Tana, Ethiopia and implications for human exposure. Chemosphere 165, 352–357, https://doi.org/10.1016/j.chemosphere.2016.09.007.

Ayalew, L., Yamagishi, H. (2003): Slope failures in the Blue Nile basin, as seen from landscape evolution perspective. Geomorphology 57(1–2), 95–116, https://doi.org/10.1016/S0169-555X(03)00085-0.

Ayman, A. A., Abdelkareem, M., Asran, A. M., Mahran, T. M. (2017): Geomorphic and lithologic characteristics of Wadi Feiran basin, southern Sinai, Egypt, using remote sensing and field investigations. Journal of Earth System Science 126(85), 1–25, https://doi.org/10.1007/s12040-017-0861-8.

Belisario, F., Del Monte, M., Fredi, P., Funiciello, R., Palmieri, E. L., Salvini, F. (1999): Azimuthal analysis of stream orientations to define regional tectonic lines. Zeitschrift für Geomorphologie Supplementary Issues 118, 41–63.

Beyene, A., Abdelsalam, M. G. (2005): Tectonics of the Afar Depression: A review and synthesis. Journal of African Earth Sciences 41(1–2), 41–59, https://doi.org/10.1016/j.jafrearsci.2005.03.003.

Chorowitz, J. (2005): The East African rift system. Journal of African Earth Sciences 43(1–3), 379–410, https://doi.org/10.1016/j.jafrearsci.2005.07.019.

Chorowicz, J., Collet, B., Bonavia, F. F., Mohr, P., Parrot, J. F., Korme, T. (1998): The Tana basin, Ethiopia: Intra-plateau uplift, rifting and subsidence. Tectonophysics 295(3–4), 351–367, https://doi.org/10.1016/S0040-1951(98)00128-0.

CiotolI, G., Seta, M. D., Del Monte, M., Fredi., P., Lombardi, S., Palmieri, E. L., Pugliese, F. (2003): Morphological and geochemical evidence of neotectonics in the volcanic area of Monti Vulsini (Latium, Italy). Quaternary International 101–102, 103–113, https://doi.org/10.1016/S1040-6182(02)00093-9.

Dile, Y., T., Karlberg, L., Daggupati, P., Srinivasan, R., Wiberg, D., Rockströmd, R. (2016): Assessing the implications of water harvesting intensification on upstream-downstream ecosystem services: A case study in the Lake Tana basin. Science of the Total Environment 542(part A), 22–35, https://doi.org/10.1016/j.scitotenv.2015.10.065.

Ehlen, J. (2004). Lineation. In Goudie, A. S. (ed.) et al.: Encyclopaedia of geomorphology. Routledge, London, 623–624.

Earthquake Track [online]. [downloaded 10/01/2017]. www.earthquaketrack.com.

ESRI Inc. ArcMap™. Ver. 10.1. Copyright 2015. [online]. [downloaded 01/01/2015]. www.esri.com.

Gani, N., D., Abdelsalam, M. G. (2006): Remote sensing analysis of the Gorge of the Nile, Ethiopia with emphasis on Dejen-Gohatsion region. Journal of African Earth Science 44, 145–150, https://doi.org/10.1016/j.jafrearsci.2005.10.007.

Gani, N. D., Gani, M. R., Abdelsalam, M. G. (2007): Blue Nile incision on the Ethiopian Plateau: Pulsed plateau growth, Pliocene uplift, and hominin evolution. GSA Today 17, 4–11, https://doi.org/10.1130/GSAT01709A.1.

Gani, N. D., Abdelsalam, M. G., Gera, S., Gani, M. R. (2009): Stratigraphic and structural evolution of the Blue Nile Basin, Northweastern Ethiopian Plateau. Geological Journal 44(1), 30–56, https://doi.org/10.1002/gj.1127.

Gani, N. D., Neupane, P. C. (2018): Understanding transient landscape of the Ethiopian Plateau in relation to mantle dynamics. Geological Journal 53(1), 371–385, https://doi.org/10.1002/gj.2903.

Google Earth, Google Inc. Available from URL. [online]. [downloaded 01/01/2015]. http://www.google.cz/intl/cs/earth

Haider, V. L., Kropáček, J., Dunkl, I., Wagner, B., and Von Eynatten, H. (2015): Identification of peneplains by multi-parameter assessment of digital elevation models. Earth Surface Processes and Landforms 40(11), 1477–1492, https://doi.org/10.1002/esp.3729.

Ismail, E. H., Abdelsalam, M. G. (2012): Morpho-tectonic analysis of the Tekeze River and the Blue Nile drainage systems on the Northwestern Plateau, Ethiopia. Journal of African Earth Sciences 69, 34–47, https://doi.org/10.1016/j.jafrearsci.2012.04.005.

Jenson S. K., Domingue, J. O. (1988): Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis. Photogrammetric Engineering and Remote Sensing 54(11), 1593–1600, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.138.6487andrep=rep1andtype=pdf.

Kaba, E., Philpot, W., Steenhuis, T. (2014): Evaluating suitability of MODIS-Terra images for reproducing historic sediment concentrations in water bodies: Lake Tana, Ethiopia. International Journal of Applied Earth Observation and Geoinformation 26, 286–297, https://doi.org/10.1016/j.jag.2013.08.001.

Kazmin, V. (1975): Geological Map of Ethiopia, Geological Survey of Ethiopia, Addis Ababa, Ethiopia.

Kennelly, P. (2008): Terrain maps displaying hill-shading with curvature. Geomorphology 102(3–4), 567–577, https://doi.org/10.1016/j.geomorph.2008.05.046.

Kieffer, B., Arndt, N., Lapierre, H., Bastien, F., Bosch, D., Pecher, A., Yirgu, G., Ayalew, D., Weis, D., Jerram, D. A., Keller, F., Meugniot, C. (2004): Flood and shield basalts from Ethiopia: Magmas from the African Superswell. Journal of Petrology 45(4), 793–834, https://doi.org/10.1093/petrology/egg112.

Klimadiagramme weltweit. [online]. [downloaded 10/10/2016]. www.klimadiagramme.de.

Koronovskya, N. V., Bryantsevaa G. V., Goncharova M. A., Naimarka A. A, Kopaev A. V. (2014): Lineaments, Planetary Jointing, and the Regmatic System: Main Points of the Phenomena and Terminology. Geotectonics 48(2), 151–162, https://doi.org/10.1134/S0016852114020058.

Kusák, M., Kropáček, J., SchillacI, C., Vilímek, V. (2016): Aspects of the evolution of the Jemma River basin in the Ethiopian Highlands determined by means of a relief analysis based on Digital Elevation Model (DEM), Ethiopia. Geografia Fisica e Dinamica Quaternaria 39, 37–50, https://doi.org/10.4461/GFDQ.2016.39.4.

Kusák, M., Krbcová, K. (2017): Analysis of the relationship of automatically and manually extracted lineaments from DEM and geologically mapped tectonic faults around the Main Ethiopian Rift and the Ethiopian Highlands, Ethiopia. AUC Geographica 52(1), 5–17, https://doi.org/10.14712/23361980.2017.1.

Kycl, P., Rapprich, V., Verner, K., Novotný, J., Hroch, T., Mišurec, J., Eshetu, H., Haile, E. T., Alemayehu, L., Goslar, T. (2017): Tectonic control of complex slope failures in the Ameka River Valley (Lower Gibe Area, central Ethiopia): Implications for landslide formation. Geomorphology 288, 175–187, https://doi.org/10.1016/j.geomorph.2017.03.020.

Macgregor, D. (2015): History of the development of the East African Rift System: A series of interpreted maps through time. Journal of African Earth Sciences 101, 232–252, https://doi.org/10.1016/j.jafrearsci.2014.09.016.

Mäerker M., Hochschild V., Maca V., Vilímek V. (2016): Stochastic assessment of landslides and debris flows in the Jemma basin, Blue Nile, Central Ethiopia. Geografia Fisica e Dinamica Quaternaria 39(1), 51–58, https://doi.org/10.4461/GFDQ.2016.39.5.

Mangesha, T., Chernet, T., Haro, W. (1996): Geological Map of Ethiopia (1 : 250,000). Geological Survey of Ethiopia: Addis Ababa, Ethiopia.

Martino, C., Nico, G., Schiattarella, M. (2009): Quantitative analysis of InSAR digital elevation models for identification of areas with different tectonic activities in southern Italy. Earth Surface Processes and Landforms 34(1), 3–15, https://doi.org/10.1002/esp.1681.

Minár, J., Sládek, J. (2008): Morphological network as an indicator of a morphotectonic field in the central Western Carpathians (Slovakia). Zeitschrift für Geomorphologie 53(2), 23–29, https://doi.org/10.1127/0372-8854/2009/0053S3-0023.

Pik, R., Deniel, C., Coulon, C., Yirgu, G., Hoffmann, C., Ayalew, D. (1998): The northwestern Ethiopian Plateau flood basalts: classification and spatial distribution of magma types. Journal of Volcanology and Geothermal Research 81(1–2), 91–111, https://doi.org/10.1016/S0377-0273(97)00073-5.

Pik, R., Marty, B., Carignan, J., Lavé, J. (2003): Stability of the Upper Nile drainage network (Ethiopia) deduces from (U/Th)/He thermochronometry: implications for uplift and erosion of the Afar plume dome. Earth and Planetary Science Letters 215(1–2), 73–88, https://doi.org/10.1016/S0012-821X(03)00457-6.

Pravea, A. R., Batesa, C. R., Donaldsona, C. H., Tolandb, H., Condonc, D. J., Markd, D., Rauba, T. D. (2016): Geology and geochronology of the Tana Basin, Ethiopia: LIP volcanism, super eruptions and Eocene-Oligocene environmental change. Earth and Planetary Science Letters 443, 1–8, https://doi.org/10.1016/j.epsl.2016.03.009.

Sembroni, A., Faccenna, C., Becker, T. W., Molin, P., Abebe, B. (2016): Long-term, deep-mantle support of the Ethiopia-Yemen Plateau. Tectonics 35(2), 469–488, https://doi.org/10.1002/2015TC004000.

Smith, M. J., Clark, C. D. (2005): Methods for the visualisation of digital elevation models for landform mapping. Earth Surface Processes and Landforms 30(7), 885–900, https://doi.org/10.1002/esp.1210.

Stoddart, D. R. (ed.) (1997): Process and form in geomorphology. Routledge, London.

Šilhavý, J., Minár, J., Mentlík, P., Sládek, J. (2016): A new artefacts resistant method for automatic lineament extraction using Multi-Hillshade Hierarchic Clustering (MHHC). Computers and Geosciences 92, 9–20, https://doi.org/10.1016/j.cageo.2016.03.015.

Ukstins, I. A., Renne, P. R. Wolfenden, E., BakeR, J., Ayalew, D., Menzies, M. (2002): Matching conjugate volcanic rifted margins: 40Ar39Ar chrono-stratigraphy of pre and syn-rift bimodal flood volcanism in Ethiopia and Yemen. Earth and Planetary Science Letters 198(3–4), 289–306, https://doi.org/10.1016/S0012-821X(02)005253.

U.S. Geological Survey [online]. [downloaded 10/01/2017]. www.usgs.gov.

Wladis, D. (1999): Automatic Lineament Detection Using Digital Elevation Models with Second Derivative Filters. Photogrammetric Engineering and Remote Sensing 65(4), 453–458, https://pdfs.semanticscholar.org/3c9e/4f7863002102ead6d390954ff4d1225e1f6d.pdf.

Wolela, A. (2010): Diagenetic evolution of the Ansian-Pliensbachian Adigrat Sandstone, Blue Nile Basin, Ethiopia, Journal of African Earth Science, 6 (1), 29–42, https://doi.org/10.1016/j.jafrearsci.2009.05.005.

Wolfenden, E., Ebiger, C., Yirgu, G., Deino, A., Ayalew, D. (2004): Evolution of the northern Main Ethiopian Rift: birth of a triple junction. Earth and Planetary Science Letters 224(1–2), 213–228, https://doi.org/10.1016/j.epsl.2004.04.022.

Is the Second Demographic Transition a useful framework for understanding the spatial patterns of fertility change in Serbia at the beginning of the 21st century?

Vladimir Nikitović[1], Daniela Arsenović[2], Aleksandar Sekulić[3], Branislav Bajat[3,*]

[1] Institute of Social Sciences, Belgrade, Serbia
[2] University of Novi Sad, Faculty of Sciences, Serbia
[3] University of Belgrade, Faculty of Civil Engineering, Serbia
[*] Corresponding author: bajat@grf.bg.ac.rs

ABSTRACT

Gaps in comprehension of demographic change in the region of ex-Yugoslavia after 1990, caused by a lack of reliable data series, frequent change of borders, and distinctive historical and cultural tradition in comparison to other post-communist societies, motivated us to contribute to the understanding of the spatial diffusion of recent profound fertility changes in South-Eastern Europe. We analysed changes in the spatial pattern and distribution of typical fertility indicators of the second demographic transition at the sub-national level in Serbia in order to find out whether these demographic shifts could be interpreted to be similar to those in Central and Eastern Europe. We found that differences in economic, historical, and cultural development between sub-regions of the country strongly affect spatial patterns of fertility change. Also, this paper suggests that the sub-regions forerunners of the first demographic transition could be considered as the cores of diffusion for the second demographic transition.

KEYWORDS

South-East Europe; second demographic transition; sub-national fertility patterns; spatial autocorrelation; Serbia

Received: 7 March 2019
Accepted: 1 August 2019
Published online: 20 September 2019

Nikitović, V., Arsenović, D., Sekulić, A., Bajat, B. (2019): Is the Second Demographic Transition a useful framework for understanding the spatial patterns of fertility change in Serbia at the beginning of the 21st century?. AUC Geographica 54(2), 152–167
https://doi.org/10.14712/23361980.2019.14

1. Introduction

In the European context, the second demographic transition (SDT) was studied through various socio-demographic processes including: fertility decline, mortality impact on fertility change, marital postponement, non-marital fertility, partner relationships, changing household forms etc. Besides reporting general remarks on the very concept of the second demographic transition as regards to its ideational and cultural drivers and universality issues (Coleman 2004; Esping-Andersen, Billari 2015), certain doubts are often expressed with respect to the applicability of this theoretical concept outside the region where it was originally recognized as a process (Lesthaeghe, Neidert 2006). As one of the SDT proponents recently stressed, there is growing evidence that phenomena such as sub-replacement fertility, rising shares of cohabitations, and non-marital fertility are spreading worldwide (Lesthaeghe 2014). Still, some authors associate these factors with the rise of poverty formulated through the 'pattern of disadvantage' (cohabitation as a characteristic of lower social stratum); therefore, expressing doubts that SDT mechanisms of cultural shifts influence demographic changes in Eastern Europe (Perelli-Harris, Gerber 2011). Furthermore, divergence in trends of SDT fertility indicators between and within European countries is evident (Billari, Liefbroer 2010). For that reason, Lesthaeghe (2014) has recently suggested that SDT should be rather conceived as a narrative for understanding the profound cultural change that will sooner or later induce a variety of demographic changes thus implying that cross- and within-country differences in levels of SDT indicators are inevitable.

In this regard, exploring whether a fertility change typical for SDT has been spreading from its source towards the East seems to be a relevant research topic. The interpretation of abrupt and fast demographic changes that followed the fall of the Iron curtain in European post-socialist countries (the CEE term often comprises all former socialist countries including the states successors of Yugoslavia) using the framework of SDT has been a frequent research topic in this century (Philipov, Kohler 2001; Sobotka 2003, 2008, 2011; Sobotka, Zeman, Kantorova 2003; Rašević 2004; Muresan 2007; Hoem et al. 2009; Kuhar, Reiter 2010; Kurek 2011a; Perelli-Harris, Gerber 2011; Botev 2012; Walford, Kurek 2016). Yet, literature explaining demographic changes in CEE countries after 1990 usually do not fully cover the region of former Yugoslavia, which is mainly due to the lack of reliable data series and frequent changes of borders ever since Yugoslavia's dissolution. In addition, the region of former Yugoslavia also had a distinctive history or cultural tradition in comparison to the other former socialist countries (Sobotka 2003).

Moreover, divergent trends in the onset and tempo of the fertility transition had been noted across sub-regions of the former Yugoslavia during the socialist period, which was distinctive from the relative homogeneity of Eastern Europe (Nikitović 2011, 2016; Josipovič 2016; Lerch 2018). The strong sub-regional differences in demographic change, which were well documented in Yugoslav demographic literature (Josipovič 2016), highlighted early adopters of fertility transition in the region (Slovenia, Croatia, and Serbia but excluding Kosovo). Declines in fertility below replacement level started in Croatia and Serbia (excluding Kosovo) approximately at the same time as it started in the vanguard countries of SDT (Nikitović 2016). This peculiarity was acknowledged to a limited extent in newer studies (Frejka, Gietel-Basten 2016; Nikitović, Bajat, Blagojević 2016; Lerch 2018), which partly can be attributed to data issues generated by the frequent change of national boundaries in the region.

When analysing the post-1990 period, Sardon (2001) noticed that strong sub-regional differences of the former Yugoslavia with respect to other former socialist countries could be hidden if the levels of indicators are examined as a whole, especially when analysing nuptiality levels and non-marital fertility. Additionally, specificities of Yugoslav socialism in relation to the Eastern bloc variant were often marked as a distinctive feature that can induce specific demographic outcomes in the socialist context (Rašević 2004). However, these specificities have led some authors to conclude that the patterns of change typical for SDT could not be relevant for explaining fertility changes in the region of former Yugoslavia (Kuhar 2009); whereas, others argue that SDT has been blocked in this region since 1990 (Bobić, Vukelić 2011).

Despite opinions that the concept of the second demographic transition is limited in Balkan countries, recent macro-level evidence suggests that Serbia[1] may have begun to follow the post-socialist pace of this transition after a period of stagnation in the 1990s (Nikitović, Bajat, Blagojević 2016). We therefore aim to examine if there is enough evidence for determining spatial diffusion of fertility change typical for SDT in Serbia at the beginning of this century. In other words, this study aims to contribute to the understanding of the spatial diffusion of recent profound fertility changes occurring throughout South-East Europe.

Although the proponents of SDT have no doubts when it comes to spatial spreading of this transition from North-West to South-East Europe since the 1990s, a recent comprehensive study on non-marital

1 In this paper, it does not include the disputed territory of Kosovo, which unilaterally proclaimed independence of Serbia in 2008. Kosovo is currently recognized as an independent state by 108 of 193 UN member states and 23 out of 28 European Union member states.

fertility (a typical SDT indicator) across sub-national regions of Europe showed that its spatial diffusion is not so straightforward and suggested that state borders are still more relevant than regional ones with respect to non-marital fertility (Klusener, Perelli-Harris, Sanchez Gassen 2013). However, some findings from the aforementioned study showed that a more detailed spatial analysis would shed more light on understanding recent spatial patterns of fertility change in South-Eastern Europe. For that reason, we focus on the spatial pattern and distribution of selected SDT indicators at the municipality level in Serbia based on spatial autocorrelation analysis in order to find out whether recent demographic shifts can be interpreted within the framework of the second demographic transition. In addition, we aim to discuss these findings in the context of regional diversity in comparison to countries of former communist regimes.

Geographic differentials in diffusion of innovations proved to be significant for sub-national differences in the onset of the first demographic transition in Serbia (Nikitović, Bajat, Blagojević 2016). Therefore, we expect that an analysis using lower spatial levels will help us to understand the spatial pattern of the current fertility transition in Serbia and whether it is similar to the patterns observed in CEE countries. The analysis is focused on common SDT indicators in terms of fertility aspects of the process: total fertility rate (TFR), mean age of women at childbearing (MAC), and percentage of births outside marriage (BOM).

2. Overview of the Fertility Change in Serbia in the Context of the Post-Socialist Societies

Fertility changes that could be interpreted in the framework of SDT lagged by at least 20 years in post-socialist countries when compared with the forerunners of the process in North-Western Europe. During the socialist period, CEE countries were characterized by different demographic development, particularly concerning fertility, in relation to the rest of Europe (Frejka, Gietel-Basten 2016). 'A demographic distinctiveness of European former Communist countries as a relatively homogeneous region had gradually developed between the mid-1960s and the mid-1980s' (Sobotka 2003: 453). In the 1970s and 1980s, institutional and cultural factors that jointly sustained the regime of universal and early reproduction under state socialism protected Eastern Europe from the fertility decline experienced in all other parts of Europe. Then, following the collapse of communist regimes, within a decade the CEE region experienced the lowest fertility rates in Europe (Sobotka 2003, 2011).

A sharp decrease in TFR in Eastern Europe during the 1990s was usually associated with political, social and economic distortions during the period. The explanations of this relationship differed – some researchers highlighted the role of macroeconomic factors while others thought of economic and political transitions as a specific part of the second demographic transition in this region (Philipov, Kohler 2001: 38; Kurek 2011b). Anyway, after 1990, the diversification in terms of social, economic, and even demographic development inside the group of former socialist countries is obvious due to different speed of societal transformation (Botev 2012: 69). Therefore, some authors suggest that the process of second demographic transition 'is not a unitary movement that reached all the countries in Central and Eastern Europe roughly at the same time and had the same features throughout, no more than it was in Western Europe' (Hoem et al. 2009: 250). They investigated marriage formation risk in selected countries in Central and Eastern Europe and found that a particular drop in this indicator started in Hungary and Bulgaria after the early 1980s and in Russia and Romania half-a-decade and full decade later, respectively.

According to Sobotka (2011: 262), a strong decrease in period fertility rates have been accompanied and partly caused by a shift towards a later timing of first births. These changes were also followed with family transformations that resulted in the rise of non-marital births. However, in relation to the period before 1990, when the share of non-marital births did not exceed 10 per cent, the CEE region is currently characterised by diversity of this typical SDT indicator – from Nordic values in Bulgaria (more than 50 per cent) to half the West Europe values in Poland (just above 20 per cent).

2.1 Change in Fertility Rates in Serbia, 1950–2017

The general pattern of changes in the period total fertility rate in Serbia since 1950 was similar to those observed in most European countries: post-war baby boom followed by a decrease to the replacement level, and then to the sub-replacement level (Figure 1). The main difference in Serbia is that the baby boom came to its end already in the late 1950s. The decline in fertility was faster and ran deeper than in most European countries in spite of the same general determinants of fertility decline, such as the adoption of new norms and values and growing levels of female labour force participation (Kupiszewski, Kupiszewska, Nikitović 2012). As a result, TFR was already 15 per cent below the replacement level by 1971. Several factors stated in literature could help one to comprehend such a distinctive pattern. Those often include early liberalization (from 1952 to 1969) of women's right to abortion (Rašević and Sedlecky 2009: 385), which was typical for former socialist countries (Sardon 2001; Sobotka, Zeman, Kantorova 2003; Frejka 2008), and structural factors such as fast secularization and industrialization during the period of Yugoslavia implying abrupt migration from rural to urban areas

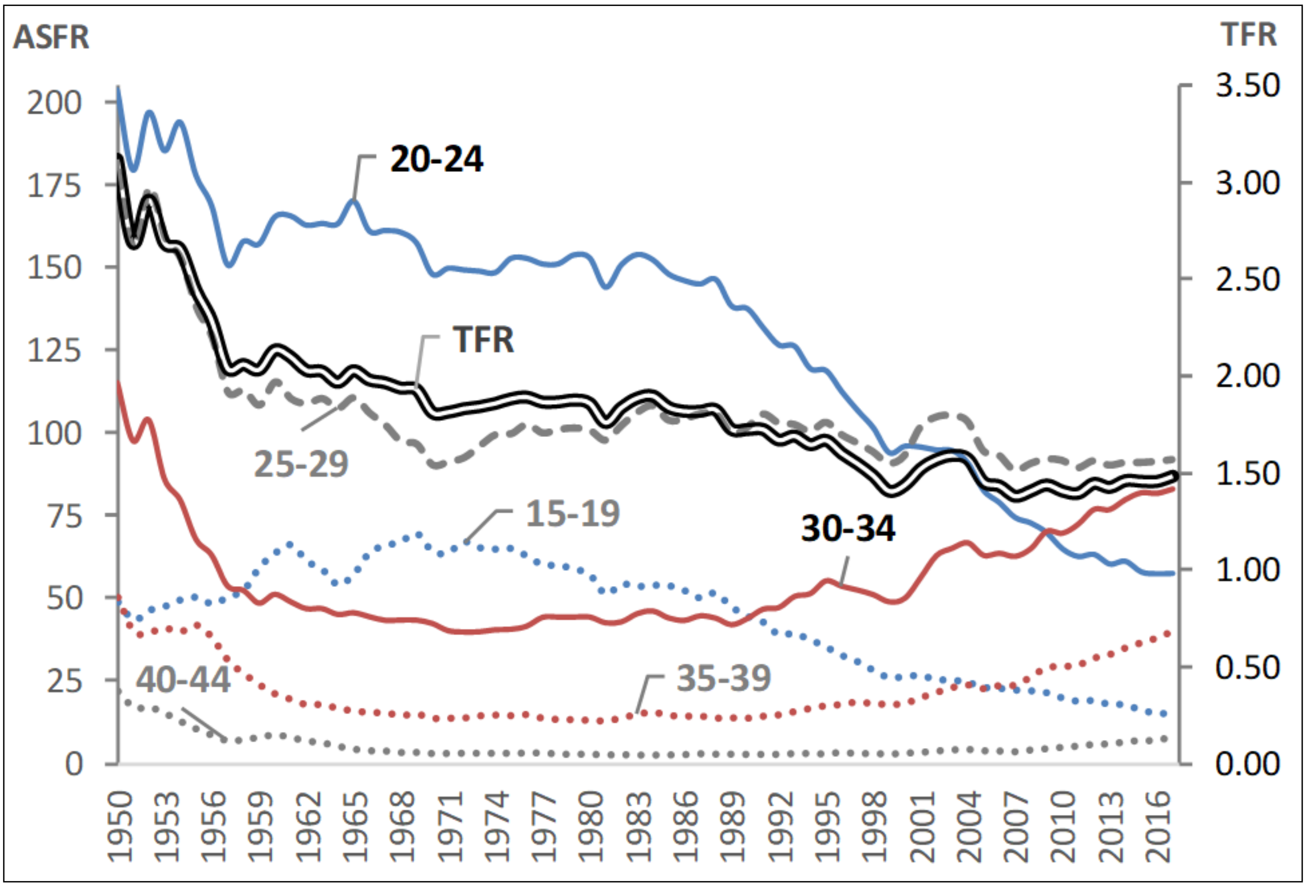

Fig. 1 Total fertility rate (double-line) and 5-year age-specific rates (solid, dashed and dotted lines) in Serbia excluding Kosovo, 1950–2017.
Notes: Time series after the census year of 2011 does not include the effect of international migration. TFR – Total fertility rate (the mean number of children that would be born alive to a woman during her lifetime if she were to pass through her childbearing years conforming to the fertility rates by age of a given year, and surviving); ASFR – age-specific fertility rates (the number of live births occurring during a given year per 1,000 women of reproductive age classified in five-year age groups).
Source: Statistical Office of the Republic of Serbia (Demographic Yearbook 2017; documentation tables).

associated with housing issues in urban areas. Yugoslav society, unlike socialist states in CEE at the time, was quite open to the impact of Western values in the sense of satisfying individual needs to a much greater extent, which arguably contributed to an increase in the economic and psychological cost of children. Furthermore, no efficient measures in the domain of population policy during the period were introduced (Rašević 2004:16).

The subsequent sharp fall in the TFR in Serbia began in the late 1980s, similar to Slovenia and Croatia. These three republics of former Yugoslavia were recognised as the 'early starters' sub-region in terms of fertility transition (Nikitović 2017). Although a steeper fall could have been expected due to wars and institutional crisis during the 1990s, the total fertility rate in Serbia has not experienced the lowest levels recently reported in Europe. Moreover, the war ambience could be a reason that contributed to prolonged consistency of traditional patriarchal values associated with family and childbearing (Sardon 2001; Rašević 2004; Petrović 2011; Lerch 2018). The total fertility rate in Serbia has oscillated between 1.4 and 1.48 since 2005. Most former socialist states that belonged to the lowest fertility group have recently experienced a rise in TFR (Latvia, Russia, and Slovenia have even exceeded 1.5), which should be expected considering the pace of 'postponement transition' (Goldstein, Sobotka, Jasilioniene 2009).

Figure 1 demonstrates a transition in the age pattern of fertility in Serbia, which occurred over the last 60 years. Very high baby-boom rates at ages older than 20 rapidly declined during the late 1950s and early 1960s and kept principally stable during the 1970s and 1980s. Only the fertility rates of adolescents experienced an increase and relatively high levels during all the period of socialist Yugoslavia. This is usually explained by the moral primacy of patriarchy in spite of the new social institutions introduced by the socialist regime (Lerch 2018). However, since the late 1980s, the fertility rates of females younger than 25 began to decline sharply, while those of females older than 30 started to increase, actually at a slower pace. This was particularly prominent for the age groups 20–24 and 30–34, as they swapped their positions in terms of contribution to the total fertility – the latter group currently has a higher rate for the first time in the last 60 years. Moreover, the difference between these two groups regarding their contribution to total fertility rate has been increasing since 2009. The highest fertility rates in the twenty-first century refer to women aged 25–29 (despite the recent drop) and is closely followed by women aged 30–34, which forms the current age pattern that is quite similar to the one observed in Poland (Walford, Kurek 2015). Although the indication of 'the postponement of childbearing for opportunistic reasons and on account of structural impediments' could have been noticed during the period of former Yugoslavia (Rašević 2004: 9), the clear onset in postponing childbirths in Serbia seems to be associated with the downfall of the socialist system, thus resembling the trends observed in other post-socialist societies. Consequently, the average age of childbearing increased from 25.9 in 1991 to 29.8 years in 2017. The latter value is lower by about half a year than in Slovenia, Estonia and Croatia, and by almost a year from the EU average, while it is similar to that in Latvia, Hungary and Poland (Eurostat 2019).

2.2 The Share of Non-Marital Births in Serbia, 1950–2017

The share of non-marital births is one of the typical indicators reflecting changes in demographic behaviour associated with the second demographic transition. This indicator raised in Serbia from 8.0 per cent in 1950 to 26.3 per cent in 2017. However, the position of Serbia regarding the share of non-marital births has shifted downwards in the European context during the last 60 years. Since the late 1950s and during the 1960s the share was stable around 12%, placing Serbia just behind Iceland and Austria in 1960 (12.2 per cent), and behind Iceland, Sweden, Estonia, and Austria in 1970 (11.8 per cent) (Penev, Stanković 2010). Given the universality of marriage and childbearing, the rise of the share at the time was due to a significant drop in the number of marital births caused by the termination of the short-lived baby-boom period (Figure 2), much shorter with respect to most European countries (Kupiszewski, Kupiszewska, Nikitović 2012).

The share of non-marital births was almost the same in 1968 and 1988, while in between it dropped. This was induced primarily by the emergence of the baby-boom echo generations, which accordingly affected the number of marital births and lowered the share of non-marital births. The significant rise of the indicator started in the late 1980s as the end of the socialist system was approaching, a trend that was similar to what was observed in most of former socialist countries. However, the share of non-marital births in Serbia in 1990 (13.1 per cent) was just above those in the Netherlands and Belgium and ahead of most former socialist societies. In 2017,

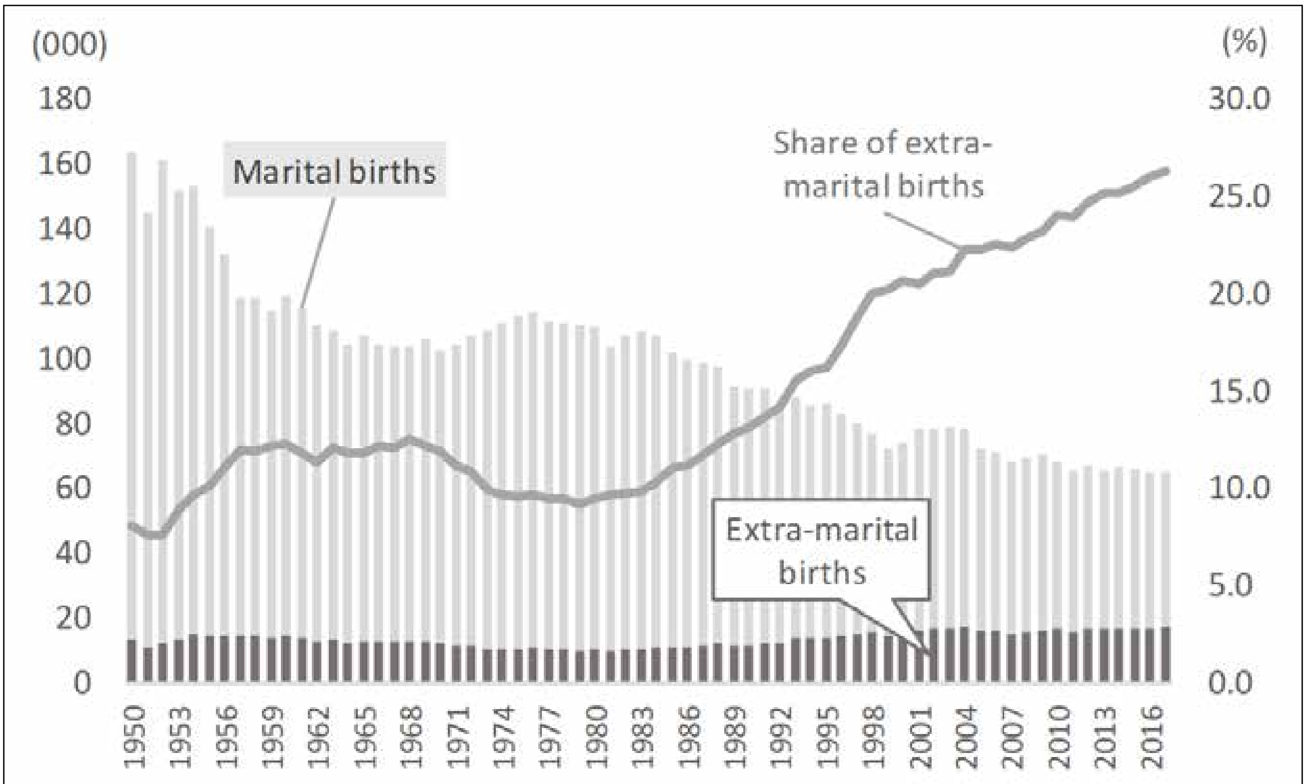

Fig. 2 Number of livebirths by legal marital status (primary axis) and the percentage of extra-marital births (secondary axis) in Serbia excluding Kosovo, 1950–2017.
Note: Time series after the census year of 2011 does not include the effect of international migration. Share of extra-marital births – the proportion of live births outside marriage.
Source: Statistical Office of the Republic of Serbia (Demographic Yearbook 2017; documentation tables).

Tab. 1 Distribution of municipalities (LAU-2) according to the categories of relevant input data in the census years of 2002 and 2011.

Total population size	Frequency of LAU-2		Number of women aged 15–49	Frequency of LAU-2		Number of livebirths	Frequency of LAU-2	
	2002	2011		2002	2011		2002	2011
<10,000	6	11	<2,000	5	13	<100	11	36
10,000–19,999	50	56	2,000–4,999	60	64	100–199	49	44
20,000–49,999	60	51	5,000–9,999	40	37	200–499	50	45
50,000–100,000	27	23	10,000–20,000	30	26	500–1000	30	16
>100,000	18	20	>20,000	26	21	>1000	21	20
Average / LAU	46,571	44,639	*Average* / LAU	11,238	10,033	*Average* / LAU	488	418

Note: Data on the number of livebirths refer to the three-year average over the census years.
Source: Statistical Office of the Republic of Serbia (Census of Population, Households and Dwellings in the Republic of Serbia, 2002 and 2011; documentation tables).

it was positioned much lower, leaving behind, apart from Switzerland, Azerbaijan, Cyprus, Greece, and Turkey, only countries from Eastern Europe (Poland, Belarus, Ukraine, Moldova) and the successor states of the former Yugoslavia except Slovenia (Eurostat 2019). The share of non-marital births in Serbia in the twenty-first century resulted mainly from pre-marital cohabitations that are more typical for people in disadvantaged economic positions, such as lone mothers including those in extended multiple generation families (Penev, Stanković 2010; Stanković 2014). It suggests that the recent rise of the indicator does not reflect spreading of cohabitation as an alternative to marriage (Petrović 2011; Bobić, Vukelić 2011). This is in line with findings from most post-socialist societies, where the diffusion of cohabitation has been rather slow (Muresan 2007, Sobotka 2008).

The following spatial analysis of the SDT indicators (TFR, MAC and BOM) will show that a variety of other factors need to be considered in Serbia in order to gain an understanding of their substantially diverse distribution at sub-regional levels.

3. Data and Methods

3.1 Data sets and territorial disposition

The spatial analyses were conducted at the local administrative unit (LAU) 2 level, which refers to municipalities in Serbia (Figure 3). The values of the analysed indicators were matched to GIS shapefiles of the municipalities' borders. These analyses are exclusively referred to census years as the official population estimates for the years out of intercensal periods do not include international migration. In this way, we wanted to get the values of the indicators as realistic as possible, which is particularly important in case of typical emigration municipalities.

Data on live births and population classified across five-age groups and sex were provided by the Statistical Office of the Republic of Serbia. Legal marital status was used as a criterion for identification of live births outside marriage. Each indicator was calculated for the last two census years – 2002 and 2011, as a three-year average over the periods 2001–2003 and 2010–2012, respectively, in order to smooth out annual fluctuations in the number of livebirths. The calculations were made following the administrative division in 2002 as no data for 2002 were available for the new six municipalities that had been established by the time of the 2011 census. Consequently, the results of the analyses in the paper were distributed across the total of 161 municipalities in Serbia excluding Kosovo. Out of that number, 42 municipalities have city status (16 of them are actually included within the City of Belgrade) based on the administrative criterion, which comprise 63.2 per cent of the total population in the country as to the 2011 census. Table 1 shows the basic descriptive statistics of the relevant input data across municipalities in the census years of 2002 and 2011. The interactive map depicting the values of calculated SDT indicators on the municipality level (LAU-2) is available in HTML format at the Web page http://osgl.grf.bg.ac.rs/static/materials/demography/tfr_mac_bom.html

3.2 Global and Local Spatial Autocorrelation Indices

The global Moran's I statistics (O'Sullivan, Unwin 2003) and local indicators of spatial association (LISA) (Anselin 1995) were calculated in order to outline spatial autocorrelation globally and to identify spatial clusters with respect to percentage difference in the SDT fertility indicators for 2002 and 2011.

As global Moran's *I* index recapitulates autocorrelation values over the entire study area, inherent spatial autocorrelation is necessary to calculate indices like LISA in order to assess significant local spatial clustering around an individual location.

In the first step, Incremental Spatial Autocorrelation (ISA) (ESRI 2012) was conducted to determine

Fig. 3 The map of Serbia with local administrative units (LAU-2).

the distance associated with peak clustering for all SDT indicators. The value obtained from ISA was then used as a distance threshold or radius for determining proximity weights (O'Sullivan, Unwin 2003) when calculating global and local spatial autocorrelation indices.

4. Spatial Analysis of Fertility Change in Serbia, 2002–2011

The analysis of the three fertility indicators connected with SDT at the level of local administrative units refers to the main objective of the study. Table 2 summarizes the changes of these indicators across municipalities in Serbia from 2002 to 2011.

The drop in TFR was significantly higher in rural than in urban areas in this period. As expected, the gap in MAC has even extended between urban and rural areas. Although the higher rise in BOM could be expected in rural than in urban areas in accordance with the recent changes in the European context (Walford, Kurek 2015), a BOM in rural municipalities is 30 per cent higher than in the cities of Serbia according to the 2011 Census. This peculiarity, which is analysed in more details across municipalities, might point to the specific forms of SDT that are atypical for countries where this process was originally described.

The next introductory step of the spatial analysis involves inspecting whether there is a correlation between each of the two pairs of indicators distributed across municipalities during the two census years, 2002 and 2011. The only significant correlation was found indicating that the mean age of mothers at childbirth is negatively associated with the share of births outside marriage, with Pearson coefficient of correlation raising between 2002 ($r = -0.35$, $p < .001$) and 2011 ($r = -0.44$, $p < .001$). This correlation pattern includes two types of association, but with opposite directions – one refers to municipalities where MAC is lower while BOM is higher and the other to municipalities where MAC is higher while BOM is lower. The first one could be expected in less developed and mainly rural areas, while the latter one could be more typical for highly urbanized and more developed sub-regions of the country. In both cases this correlation suggests that cohabitation is still not common alternative to marriage as it was noted in recent studies on fertility and union formation in Serbia. The similar findings on this SDT indicator may be determined among countries of ex-Yugoslavia excluding Slovenia as well as in Poland and Slovakia (Bobić 2014). The following spatial analysis across municipalities in Serbia provides a deeper insight in sub-national differences in fertility.

The two methods used aimed at checking whether there is a spatial spread of SDT fertility indicators throughout Serbia. One implies a descriptive analysis of the spatial differences of each indicator across the country and their dynamics between the two last census years, 2002 and 2011, while the other relies on spatial autocorrelation indices calculated upon relative changes (%) over the intercensal period 2002–2011.

4.1 Total Fertility Rate

According to the 2002 census, TFR was below the replacement level in almost all of the 161 municipalities in Serbia, as only 8 of them reached 2.1 or above (Figure 4(a)). Five of these eight municipalities were predominantly populated by Muslims, the religious group who were the last in the country to enter the final stage of demographic transition. The remaining three municipalities belong to the group with the highest share of persons living abroad for more than a year. Unlike 2011, the TFR for 2002 included the births by this group of emigrants indicating that the TFR was actually lower than 2.1 in these municipalities (Rašević, Penev 2010).

Due to a further decline of TFR, only three municipalities with predominantly Muslim populations were above 2.1 in 2011, while ten municipalities dropped to just around 1 (0.99–1.09) (Figure 4(b)). Only 13 per cent of municipalities (21 of 161) did not decline in TFR between the census 2002 and 2011. The highest increase (mainly lower than 10 per cent) took place in the northern, most urbanized, part of the country including municipalities in the area of the two largest

Tab. 2 Quantitative analysis of the second demographic transition indicators in urban and rural local administrative units in Serbia, 2002–2011.

	Total fertility rate		Mean age of women at childbirth		Births outside marriage (%)	
	Urban	Rural	Urban	Rural	Urban	Rural
year 2002	1.54	1.65	27.37	26.18	20.46	23.69
year 2011	1.44	1.35	28.90	27.16	23.27	30.29
% difference	−6.51*	−17.79**	5.59**	3.73**	13.75*	27.87**

** Significant at $p = 0.001$; * significant at $p = 0.05$ using Mann–Whitney test.

Note: A municipality is classified into urban or rural depending on its administrative status according to the current administrative division of local spatial units (cities and municipalities).

Source: Statistical Office of the Republic of Serbia (documentation tables).

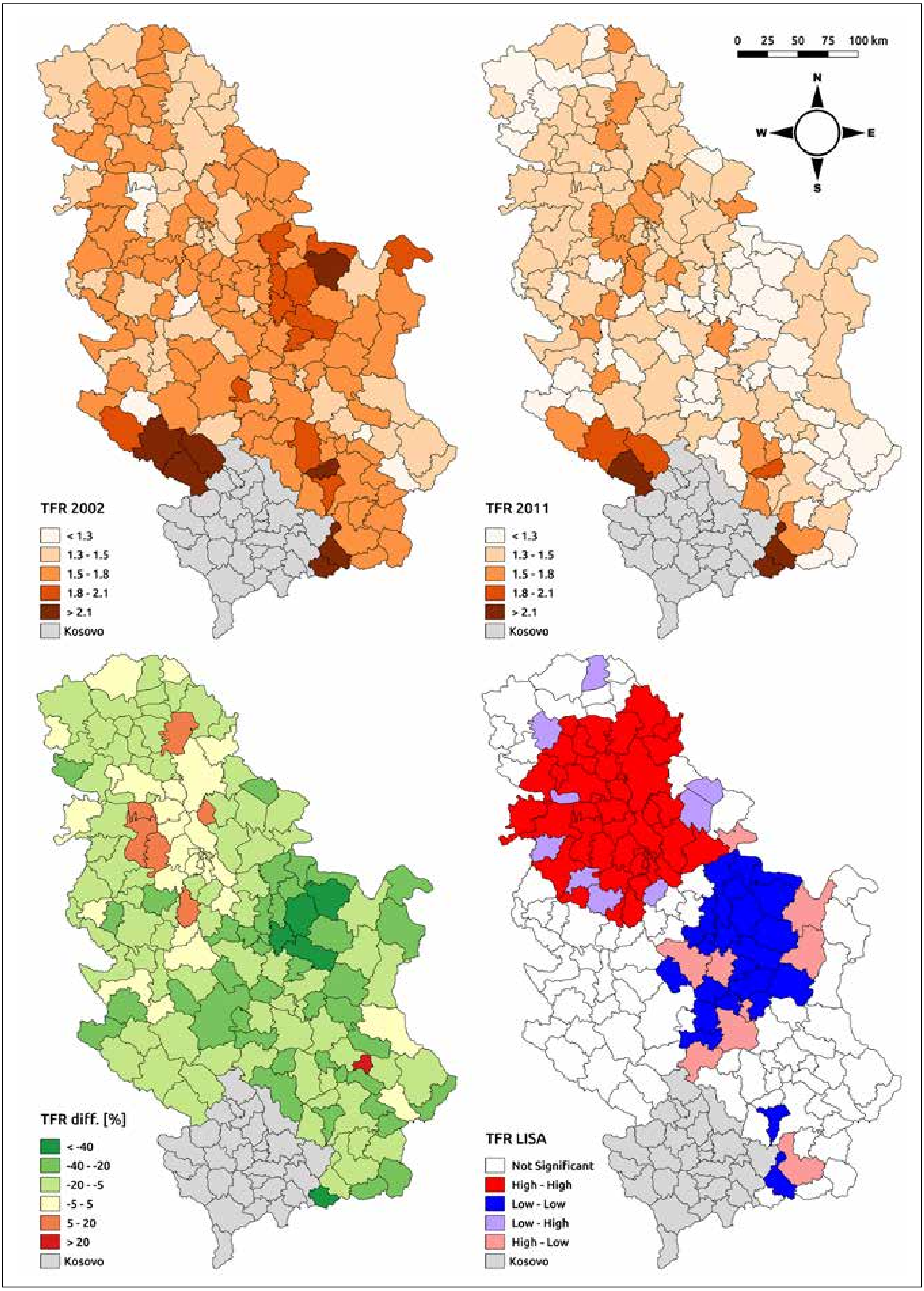

Fig. 4 Total Fertility Rate in Serbia at LAU-2 level: (a) 2002, (b) 2011, (c) percentage difference for the period 2002–2011, (d) LISA cluster map of percentage difference for the period 2002–2011.

cities – Belgrade (the capital) and Novi Sad (the capital of the province of Vojvodina). This result indicates the effect of postponing births to later ages (more typical in urban areas), rather than the rise in the cohort fertility (Rašević 2015).

A quarter of all municipalities had experienced a decrease in TFR between 20 and 40 per cent from 2002 to 2011 (Figure 4(c)). Most of these municipalities are in eastern and southern Serbia, the least developed regions in the country according to the 2014 official scale of development listed on the website of the National Agency for Regional Development. The observed decline during the period was even larger than 40 per cent in several municipalities in the northeast of this area, which is distinguished by the highest share of population living abroad (Penev, Predojević-Despić 2012). Given a difference in methodology when the rates for 2002 and 2011 are compared (the latter one did not include births by emigrants), this strong reduction could not only be attributed to the actual fertility fall (Rašević, Penev 2010).

Global Moran's $I = 0.342$ ($Z = 11.824$, $p < 0.001$) for the TFR (threshold radius h = 70 km) in Serbia indicates moderately clustered spatial patterns with high statistical significance. The LISA cluster map designates locations with significant local Moran indicators assorted by different types of spatial correlation: 1) High-High association, a municipality with high TFR value has neighbouring municipalities with high TFR values (positive spatial correlation); 2) Low-Low association, a municipality with low TFR value has neighbouring municipalities with low TFR values (positive spatial correlation); 3) Low-High association, a municipality with low TFR value has neighbouring municipalities with high TFR values (negative spatial correlation); 4) High-Low association, a municipality with high TFR value has neighbouring municipalities with low TFR values (negative spatial correlation).

The local Moran's *I* statistics singled out a big compact High-High cluster in TFR denoting a metropolitan area that includes the two largest cities (Belgrade and Novi Sad) located in northern Serbia (Figure 4(d)). The cluster, primarily, reflects a quite small increase in TFR between 2002 and 2011 in the area. It might be explained by both the development of housing associated with a recent concentration of industry particularly in suburbs of the area induced by the proximity to important highways, and the ongoing process of south-to-north internal migration (Nikitović, Predojević-Despić, Marinković 2015). A big Low-Low cluster in East Serbia indicates a decrease in TFR, which is partly caused by the changes in methodology between the 2002 and 2011 census, but also reflects the indirect effects of continuous international emigration from this area (Penev, Predojević-Despić 2012; Fassmann, Musil, Gruber 2014).

4.2 Mean Age of Women at Childbirth

The highest values of MAC in 2011 are recorded in the largest urban centres of Serbia, almost exclusively in those located in north and west parts of the country, except for Niš – the largest city in southern Serbia, while the lowest values refer to the east (Figure 5(b)). The number of municipalities with MAC higher than 28 years increased five times (from 12 to 60) between 2002 and 2011. The minimum values of MAC (observed in the southeast – the least developed area of the country) remained almost the same after a decade (24.4 years in 2002 compared with 24.7 in 2011), while maximum MAC (recorded in the historic centre of the Belgrade City) increased from 30.4 to 32.2 years (Figures 5(a–c)). Only 9 of 161 municipalities did not report a rise of MAC between 2002 and 2011, clearly indicating the transition in the peak fertility towards older ages throughout the country. The largest increase in MAC (between 6 and 10 percent) during the 2000s occurred mainly in municipalities with smaller urban centres (most of them located in Central Serbia) and those where relatively lower values of MAC had previously prevailed.

Global Moran's $I = 0.144$ ($Z = 5.338$, $p < 0.001$) for the MAC (threshold radius h = 140 km) suggests less pronounced clustered spatial patterns with high statistical significance. As regards to local Moran's I statistics for the mean age at childbearing, there is almost a continuous High-High cluster connecting the area of the Belgrade City and Central Serbia (Figure 5(d)). The distribution of all four types of clusters could be explained by two processes. The first one refers to areas where MAC was previously very low indicating general spatial diffusion of demographic innovation. The other process is associated with recent trends in urbanisation (mostly pronounced along important road corridors in Serbia) indicating a higher mean age at childbearing in urban than in rural areas (Rašević 2015).

4.3 Live Births Outside Marriage

Eastern Serbia differs quite distinctively from the rest of the country in terms of the percentage of births outside marriage. The BOM values higher than 30 per cent were recorded in almost all municipalities in eastern Serbia, where BOM is even higher than 40 per cent in 11 municipalities for the year 2002. On the other hand, BOM is lower than 20 per cent in 71 of 161 municipalities in the country (Figure 6(a)). In 2011, the observed spatial polarization became even more pronounced (Figure 6(b)). The increase of BOM values in eastern Serbia is noticeable when compared to stagnating rate in southwestern Serbia (9 municipalities with BOM less than 10 per cent). Generally, the increase of BOM values and BOM range throughout the country are obvious in the period of

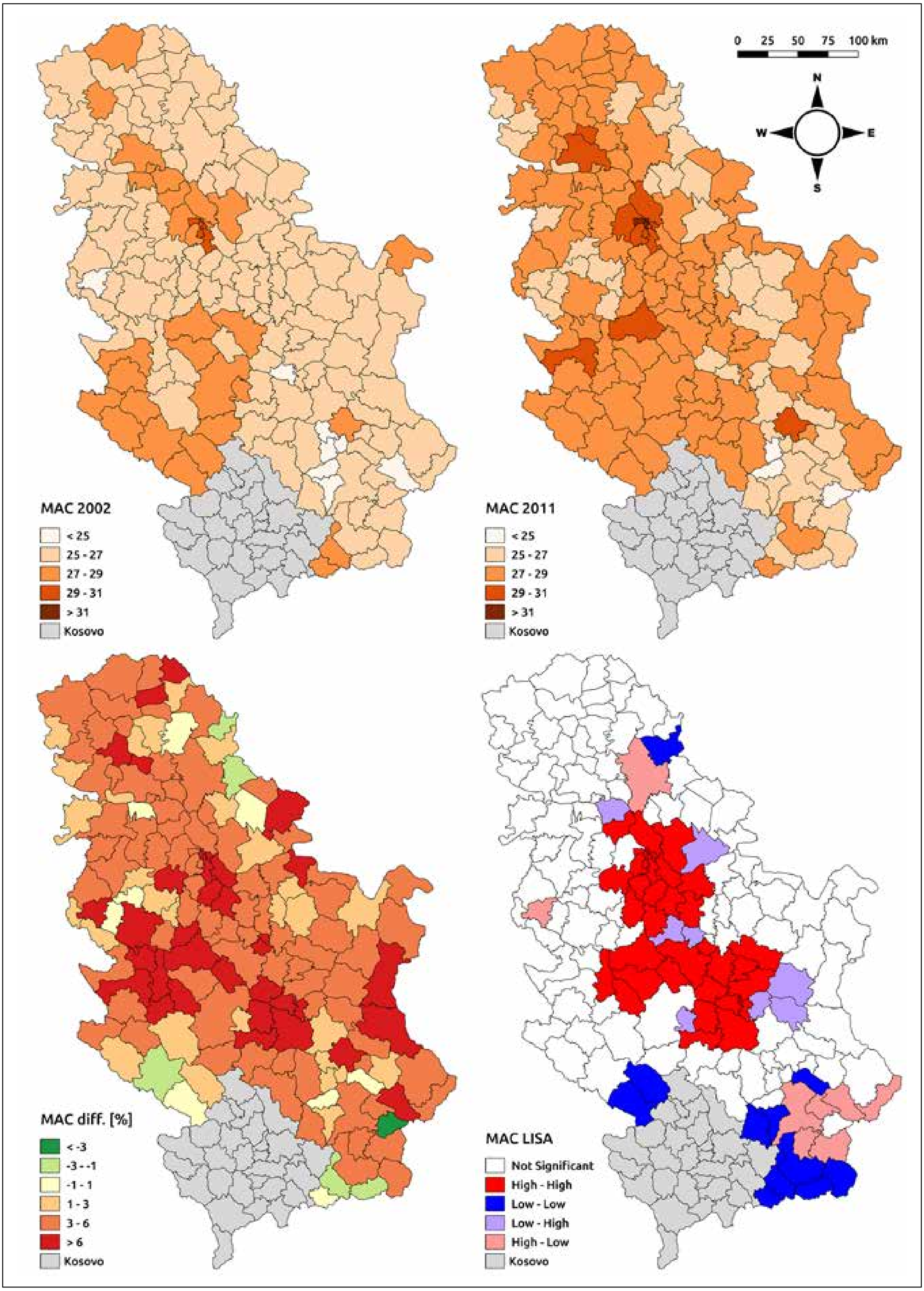

Fig. 5 Mean age of women at childbirth in Serbia at LAU-2 level: (a) 2002, (b) 2011, (c) percentage difference for the period 2002–2011, (d) LISA cluster map of percentage difference for the period 2002–2011.

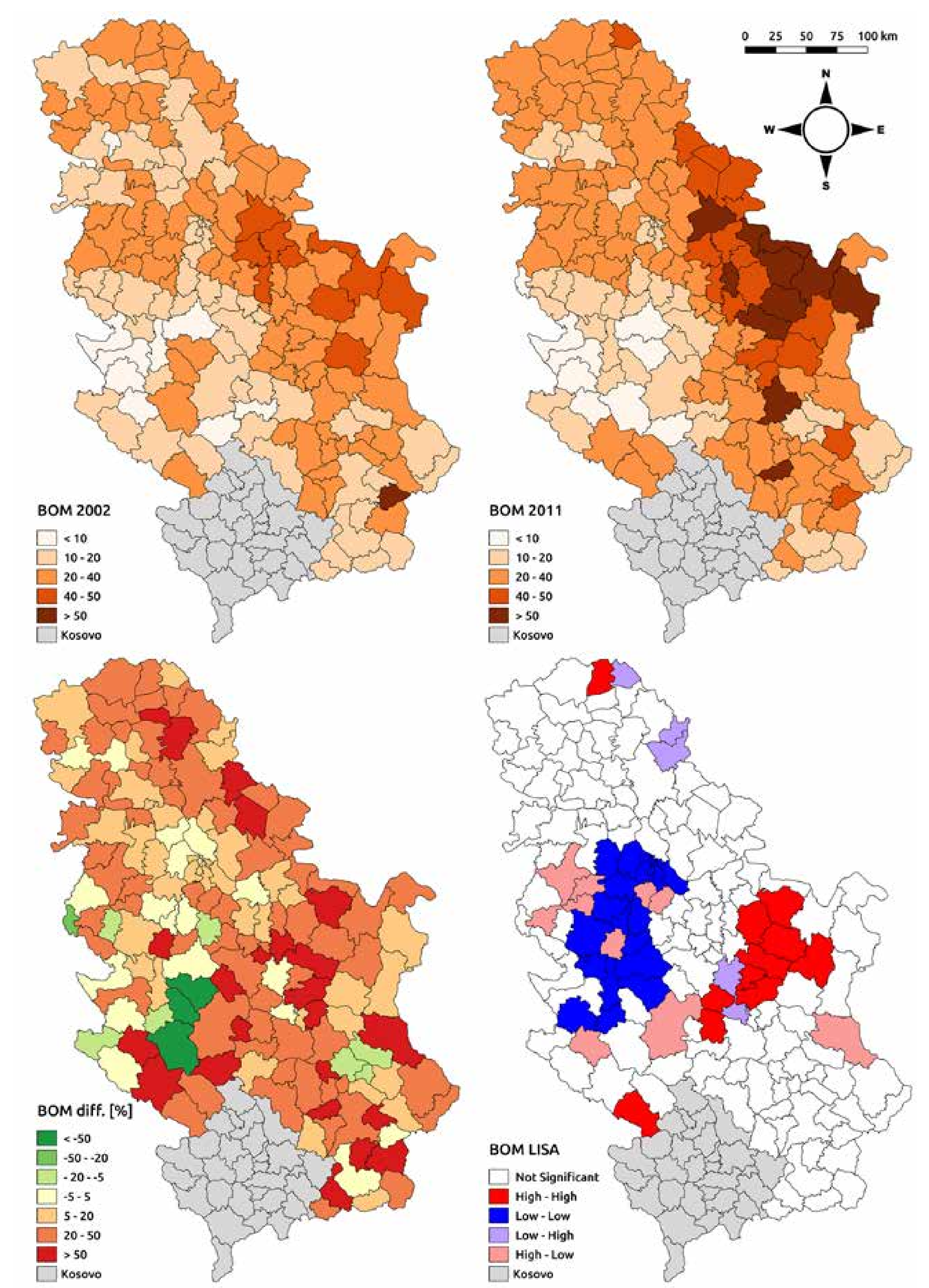

Fig. 6 Share of Live Births Outside Marriage in Serbia at LAU-2 level: (a) 2002, (b) 2011, (c) percentage difference for the period 2002–2011 (%), (d) LISA cluster map of percentage difference for the period 2002–2011.

2002–2011. The number of municipalities with BOM below 20 per cent dropped to 50, while the number of those with a share of births outside marriage higher than 40 per cent reached 29, of which 13 had a share higher than 50 per cent. Only 20 municipalities experienced a decline of BOM between 2002 and 2011, while the strongest drops were localized in the southwestern area characterized by the lowest values of the indicator. A total of 23 municipalities experienced a rise of BOM higher than 50 per cent (Figure 6(c)).

Global Moran's I = 0.078 (Z = 2.818, p < 0.006) for the BOM (threshold radius h = 52 km) indicates a certain degree of clustered spatial patterns with statistical significance. It should be noted that the share of BOM of around or below 20 percent was recorded in the largest cities of the country regardless of their geographic location, while only five urban centres (all of them in eastern Serbia) experienced BOM higher than 30 per cent. The local Moran's I statistics for BOM denoted two clusters that represent two distinctively different areas (east and west in relation to the main road corridor) with regards to the percentage difference in the share of births outside marriage between 2002 and 2011 (Figure 6(d)). One distinctive area refers to the largest urban centres in the more developed west sub-region, including central urban zones of the Belgrade City, characterized by the lowest share of BOM and its change over the period. The other distinctive area refers to municipalities in the less developed east sub-region with the highest share and highest increase of BOM. On the first glance, this observation could be considered as a paradox. However, it was noted that 'cohabitations and extra marital births in Serbia are more common among actors at the lower end of the stratification ladder … [as a result of] … rationally developed strategies in overcoming structural risks, although, without ideational changes typical to the theory of SDT' (Petrović 2011: 78). In addition, East Serbia could be considered as a specific sub-region given the distinctiveness of its demographic change since the mid-nineteenth century (Knežević 2013).

5. Discussion and Conclusions

The objective of this study was to determine whether recent spatial patterns of fertility change in Serbia, which we assumed to be associated with the second demographic transition, agree with those observed in post-socialist societies of Europe. Whereas most of the literature exploring whether and how SDT taking place in former Yugoslavia refers to the national level and to aspects of family change, we wanted to see if sub-national geographic differentials in diffusion of demographic innovations measured by key fertility indicators associated with SDT could shed more lights on the issue. We deployed our analyses from the national level, however special focus was maintained at the level of municipalities.

Concerning the national level, the total fertility rate and mean age at childbearing in Serbia follow paths similar to those in CEE countries. Even though the last sharp decrease of TFR began a little earlier than in CEE countries, it did not reach the lowest values as reported in CEE. This can be explained through a not so sharp transition from state socialism to new societal conditions and stubborn subsistence of traditional values associated with family formation during the civil wars in the 1990s (Petrović 2011; Bobić, Vukelić 2011; Lerch 2018). The latter reason could also have an impact on the much lower rate of non-marital fertility in Serbia than in most of CEE countries, which is almost the same as in Poland, and higher only than those in former Soviet republics and Croatia (Eurostat 2019).

The three general findings concerning changes of the analysed fertility indicators with respect to the type of municipality in Serbia were pointed in the first decade of this century: a) the decline in fertility has diffused from urban to rural areas, b) the gap in mean age at childbearing between urban and rural areas has widened, and c) the share of non-marital births became significantly higher in rural than in urban areas. The diffusion of values and ideas associated with low fertility norms from urban centres to peripheral areas of the country was already well noted during the first demographic transition (Watkins 1990). Also, the pronounced postponing of births in cities is in accordance with the recent changes in post-socialist societies (Walford, Kurek 2015). However, such a discrepancy between rural and urban areas in terms of non-marital fertility might be unexpected if we neglect that non-marital births in Serbia are mostly reported among women from lower social strata, which are often located in rural areas (Bobić 2014).

More specific results of the autocorrelation spatial analysis, as the pivotal analytic tool, denoted that the metropolitan area, which includes the two largest cities located in the northern Serbia, is the only one with a mild but significant increase of TFR in the country at the beginning of this century. At the same time, spatial diffusion of the rise in mean age at childbearing was registered throughout the country but was mostly pronounced along important road corridors connecting bigger towns. This indicates that the process is particularly associated with recent trends in urbanisation. Our results generally coincide with the findings from Czechia indicating that 'this stage in the transformation of reproductive behaviour began earlier in the largest city districts and districts with economic centres.' As such, the spatial patterns of fertility change demonstrate different speed of adaptation to the changes in reproductive conditions (Šprocha, Šidlo 2016: 228). In this respect, East Serbia is clustered in terms of a decrease in TFR, which could be to some extent associated with indirect effects of continuous, long-lasting international emigration from that area. The diffusion of the increase in the proportion of non-marital births from its core in East Serbia was evident with respect to spatial statistics on the percentage difference of this indicator between 2002 and 2011. The two distinctive opposite clusters (East and Southwest Serbia) suggest that not only the tempo of this diffusion was rather slow, but the persistence of traditional values associated with family formation is very strong in the Southwest, which is recognized as the sub-region that was the last to enter fertility transition (Nikitović, Bajat, Blagojević 2016).

Specificities of East Serbia that relate to the analysed indicators of SDT, particularly as to the share of non-marital fertility, belong to the outcomes that might be unexpected to some extent. Furthermore, this sub-region stands as an exception with respect to the Hajnal line that reflects the nuptiality regime division between West and East Europe, which seems to be still relevant for differentials in non-marital fertility between sub-regions in Europe (Walford, Kurek 2015). East Serbia was the innovating sub-region as for the first fertility transition in the country (the late nineteenth century). Also, it could be an innovating area regarding the second transition, similarly to what Lesthaeghe and Neels (2002) concluded exploring the spatial diffusion of SDT in Belgium and France.

The drivers of the first demographic transition in East Serbia were atypical in a way similar to what could be the current specificity. In this sub-region, recognized by the great influence of the cultural model of the Vlach population, the fertility transition already started in the second half of the nineteenth century with the acceptance of birth control through the 'one-child system'. The most commonly identified cause of the transition was agrarian overpopulation due to the extensive farming of predominantly rural populations, which increased after the end of an immigration that was induced by the enlargement of the territory of Serbia. On the other hand, there were also indications that the trigger was endemic diseases such as syphilis (Knežević 2013).

Therefore, the specificity of East Serbia might be better perceived if it is considered from a longer historical and socio-cultural perspective, which highlights the importance of specific ethnic differences for understanding current fertility change in the sub-region. Furthermore, the influence of national borders on the diffusion of fertility changes associated with SDT could be questioned in this part of Europe. A recent study (Klusener, Perelli-Harris, Sanchez Gassen 2013) pointed out that most prominent national dividing lines with respect to non-marital fertility after 1990 shifted from West and Central to Southeast Europe. The finding was probably driven by study limitations regarding different levels of regional decomposition particularly between former Yugoslavia and its neighbours. The spatial analysis in this paper showed that national boundaries in Eastern Europe are not so dividing with respect to SDT indicators. For example, the rise in non-marital fertility rates in Bulgaria resembles the one in East Serbia.

On a national scale, Serbia and Poland are very similar in terms of the level and family settings of non-marital fertility. However, only a closer look at the sub-national patterns of the diffusion of this indicator reveals that 'diversified level of extramarital births in the spatial arrangement reflects cultural and religious differences of the Poland's population' (Kurek 2011a: 393). Likewise, our results suggest that the diffusion of this SDT indicator in Serbia is clearly guided by differences in historical and cultural heritage between sub-regions of the country. Therefore, the spatial pattern of the share of non-marital births, as an indicator of ideational change, shows that typical SDT drivers may not be the decisive determinant for recent fertility change in Serbia. The fact that BOM is not only the result of modern cohabitations but is more attributed to a rise in unstable partnerships that end up with single mothers of poor socio-economic status supports this conclusion. Furthermore, the slow rise of BOM since the beginning of this century along with the highest and increasing level of postponement of births in urban and most developed sub-regions of the country questioning ideational roots of the change assumed by SDT. The societal anomie, slow transition to market economy and stable political conditions typical for the post-socialist countries, may also influence individual decisions to defer childbearing in case of the higher educated or to adapt their strategies associated with family formation in case of persons in disadvantaged socio-economic position (Perelli-Harris, Gerber 2011).

However, as Van de Kaa (2004: 8) pointed in his paper revisiting the concept of SDT, 'it is not of any great significance if countries and regions with different cultural endowments do not rapidly converge to a standard pattern'. Also, it is possible that societies like the Serbian society just follow a specific path of progression of SDT, as suggested by Sobotka (2008). In other words, the pattern of advantage and the pattern of disadvantage could be just two opposite sides of the same coin. This seems to be reasonable because both modern and traditional values influencing decisions of individuals concerning family formation were noticed at the same time in Serbia (Bobić, Vukelić 2011). It could be also interpreted as a very slow transition to what some authors refer to as the gender egalitarian family norms (Esping-Andersen, Billari 2015). The similar trend was noted in Romania indicating that the SDT process might be in progress as most of the 'threshold levels' of the onset of SDT were surpassed with exception of cohabitation, ultimate celibacy, and modern contraception (Muresan 2007: 65). Moreover, our study shows that the transition between traditional and modern concerning the demographic regime might have its geographical dimension in Serbia, too.

Finally, as Thornton and Philipov (2009) noticed, the ideational change behind SDT is important in explaining changes in marriage, cohabitation and childbearing in CEE; particularly for the belief that adopting modern family systems will help to produce modern political and economic accomplishments as they are interdependent. If one assumes that Serbia is on the course of SDT, it seems that further diffusion of the ideational changes throughout the region could be closely related to the process of EU enlargement to the East. For that reason, it will be beneficial to include the whole region of the former Yugoslavia in further research on sub-national patterns of fertility change. Currently, the issues on availability and quality of relevant demographic data are a serious obstacle to accomplish this.

References

Anselin, L. (1995): Local indicators of spatial association-LISA. Geographical Analysis 272, 93–115, http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x.

Frejka, T., Gietel-Basten, S. (2016): Fertility and family policies in Central and Eastern Europe after 1990. Comparative Population Studies 41(1), 3–56, http://dx.doi.org/10.12765/CPoS-2016-03en.

Billari, C. F. (2005): Europe and its fertility: from low to lowest low. National Institute Economic Review 194(1), 56–73, https://doi.org/10.1177/0027950105061496.

Billari, C. F., Kohler, H. P. (2004): Patterns of low and lowest-low fertility in Europe. Population Studies 58(2), 161–176, http://dx.doi.org/10.1080/0032472042000213695.

Bobić, M. (2014): Specific (blocked) post-transition union transformation in Serbia. Sociologija i prostor 52(2), 207–225, https://doi.org/10.5673/sip.52.2.5.

Bobić, M., Vukelić, J. (2011): Second demographic transition de-blocked? Sociologija 53(2), 149–176, https://doi.org/10.2298/SOC1102149B.

Bongaarts, J. (2002): The end of the fertility transition in the developed world. Population and Development Review 28(3), 419-433, http://dx.doi.org/10.1111/j.1728-4457.2002.00419.x.

Botev, N. (2012): Population ageing in Central and Eastern Europe and its demographic and social context. European Journal of Ageing 9(1), 69–79, http://dx.doi.org/10.1007/s10433-012-0217-9.

ESRI (2012): Spatial Statistics Tools, ArcMap 10.1. Redlands, CA, Environmental Systems Research Institute.

Fassmann, H., Musil, E., Gruber, K. (2014): Dynamic historical analysis of longer term migratory, labour market and human capital processes in the SEEMIG region. SEEMIG Working papers No. 3. Budapest, Hungarian Demographic Research Institute. http://www.seemig.eu/downloads/outputs/SEEMIGWorkingPapers3.pdf.

Goldstein, R. J., Sobotka, T., Jasilioniene, A. (2009): The end of lowest-low fertility? Population and Development Review 35(4), 663–699, http://dx.doi.org/10.1111/j.1728-4457.2009.00304.x.

Hoem, J. M., Kostova, D., Jasilioniene, A., Muresan, C. (2009): Traces of the second demographic transition in four selected countries in Central and Eastern Europe: Union formation as a demographic manifestation. European Journal of Population 25(3), 239–255, https://doi.org/10.1007/s10680-009-9177-y.

Josipovič, D. (2016): The post-Yugoslav space on a demographic crossway: 25 years after the collapse of Yugoslavia. Stanovništvo 54(1), 15–40, https://doi.org/10.2298/STNV160415006J.

Klusener, S., Perelli-Harris, B., Sanchez Gassen, N. (2013): Spatial aspects of the rise of nonmarital fertility across Europe since 1960: The role of states and regions in shaping patterns of change. European Journal of Population 29(2), 137–165, http://dx.doi.org/10.1007/s10680-012-9278-x.

Knežević, A. (2013): Istorijskodemografske i etnodemografske osnove razvitka stanovništva Istočne Srbije. [PhD diss.]. Beograd: Univerzitet u Beogradu.

Kuhar, M., Reiter, H. (2010): Transformation and demographic change in the ex-Yugoslav countries – materialist, idealist and institutionalist perspectives on reproductive trends. Annales for Istrian and Mediterranean Studies, Series Historia et Sociologia 20(1), 13–26.

Kupiszewski, M., Kupiszewska, D., Nikitović, V. (2012): Impact of Demographic and Migration Flows on Serbia. Belgrade, International Organization for Migration – Mission to Serbia.

Kurek, S. (2011a): Population changes in Poland: A second demographic transition. Procedia Social and Behavioral Sciences 19, 389–396, http://dx.doi.org/10.1016/j.sbspro.2011.05.146.

Kurek, S. (2011b): Double transitions? Regional patterns of population ageing in Poland. Geografiska Annaler: Series B, Human Geography 93(2), 163–184, https://doi.org/10.1111/j.1468-0467.2011.00367.x.

Lerch, M. (2018): Fertility and union formation during crisis and societal consolidation in the Western Balkans. Population Studies 72(2), 217–234, https://doi.org/10.1080/00324728.2017.1412492.

Lesthaeghe, R. (2014): The second demographic transition: A concise overview of its development. Proceedings of the National Academy of Sciences 111(51), 18112–18115, http://dx.doi.org/10.1073/pnas.1420441111.

Lesthaeghe, R., Neels, K. (2002): From the first to the second demographic transition: An interpretation of the spatial continuity of demographic innovation in France, Belgium and Switzerland. European Journal of Population 18(4), 325–360, https://doi.org/10.1023/A:1021125800070.

Lesthaeghe, R., Neidert, L. (2006): The second demographic transition in the United States: Exception or textbook example? Population and Development Review 32(4), 669–698, http://dx.doi.org/10.1111/j.1728-4457.2006.00146.x.

Muresan, C. (2007): Family dynamics in pre and post-transition Romania: a life table description. MPIDR Working Paper WP 2007-18. Rostock, Max Planck Institute for Demographic Research. https://www.demogr.mpg.de/papers/working/wp-2007-018.pdf.

Nikitović, V. (2011): Functional data analysis in forecasting Serbian fertility. Stanovništvo 49(2), 73–89, https://doi.org/10.2298/STNV1102073N.

Nikitović, V. (2016): Long-term effects of low fertility in the region of former Yugoslavia. Stanovništvo 54(2), 27–58, https://doi.org/10.2298/STNV161115009N.
Nikitović, V. (2017): The impact of migration on demographic processes in the region of former Yugoslavia. In: Bobić, M., Janković, S. Towards Understanding of Contemporary Migration: Causes, Consequences, Policies, Reflections. Belgrade: Faculty of Philosophy, University of Belgrade and Serbian Sociological Society, 61-81.
Nikitović, V., Bajat, B., Blagojević, D. (2016): Spatial patterns of recent demographic trends in Serbia 1961–2010. Geografie 121(4), 521–543.
Nikitović, V., Predojević-Despić, J., Marinković, I. (2015): Migracije stanovništva. In: Nikitović, V. Populacija Srbije početkom 21. veka. Beograd, Republički zavod za statistiku, 97–128.
O'sullivan, D., Unwin, D. J. (2003): Geographical Information Analysis. New Jersey, John Wiley and Sons.
Penev, G., Predojević-Despić, J. (2012): Spatial aspects of emigration out of Serbia: Three 'hot' emigration zones. Stanovništvo 50(2), 35–64, https://doi.org/10.2298/STNV1202035P
Penev, G., Stanković, B. (2010): Characteristics of extramarital births in Serbia in the second half of the 20th and at the beginning of the 21st century. Stanovništvo 48(1), 1–23, https://doi.org/10.2298/STNV1002001P.
Perelli-Harris, B., Gerber, T. P. (2011): Nonmarital childbearing in Russia: Second demographic transition or pattern of disadvantage? Demography 48(1), 317–342, https://doi.org/10.1007/s13524-010-0001-4.
Petrović, M. (2011): Changes of marital behaviour and family patterns in post-socialist countries: Delayed, incomplete or specific second demographic transition? Stanovništvo 49(1), 53–78, https://doi.org/10.2298/STNV1101053P.
Philipov, D., Kohler, H.-P. (2001): Tempo effects in the fertility decline in Eastern Europe: Evidence from Bulgaria, the Czech Republic, Hungary, Poland, and Russia. European Journal of Population 17(1), 37–60, https://doi.org/10.1023/A:1010740213453.
Rašević, M. (2004): Fertility trends in Serbia during the 1990s. Stanovništvo 42(1–4), 7–27, https://doi.org/10.2298/STNV0404007R.
Rašević, M. (2015): Fertilitet ženskog stanovništva. In: Nikitović, V. Populacija Srbije početkom 21. veka. Beograd, Republički zavod za statistiku, 74-95.
Rašević, M., Penev, G. (2010): Opštine republike Srbije. Osnovni demografski, ekonomski i socijalni pokazatelji relevantni za populacionu politiku. Beograd, UNFPA and Društvo demografa Srbije.
Rašević, M, Sedlecky, K. (2009): The abortion issue in Serbia. European Journal of Contraception and Reproductive Health Care 14(6), 385–390, http://dx.doi.org/10.3109/13625180903215422.
Sardon, J-P. (2001): Demographic change in the Balkans since the end of the 1980s. Population: An English Selection 13(2), 49–70.
Sobotka, T. (2003): Re-emerging diversity: rapid fertility changes in Central and Eastern Europe after the collapse of the communist regimes. Population 58(4/5), 451–486, http://dx.doi.org/10.1007/s10433-012-0217-9.
Sobotka, T. (2008): The diverse faces of the second demographic transition in Europe. Demographic Research 19, 171–224, http://dx.doi.org/10.4054/DemRes.2008.19.8.
Sobotka, T. (2011): Fertility in Central and Eastern Europe after 1989: collapse and gradual recovery. Historical Social Research 36(2), 246–296, http://dx.doi.org/10.12759/hsr.36.2011.2.246-296.
Sobotka, T., Zeman, K., Kantorova, 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.
Stanković, B. (2014): Lone mothers and their network support: Sociodemographic research of nonmarital parenthood in Serbia. Stanovništvo 52(1), 55–76, https://doi.org/10.2298/STNV1401055S.
Šrocha, B., Šidlo, L. (2016): Spatial differentiation and fertility postponement transition in Czechia. AUC Geographica 51(2), 217–233, https://doi.org/10.14712/23361980.2016.18.
Van de Kaa, D. J. (1987): Europe's second demographic transition. Population Bulletin 42(1), 1–59.
Van de Kaa, D. J. (2004): Is the second demographic transition a useful research concept: Questions and answers. Vienna Yearbook of Population Research 2, 4–10, http://dx.doi.org/10.1553/populationyearbook2004s4.
Walford, N., Kurek, S. (2016): Outworking of the second demographic transition: National trends and regional patterns of fertility change in Poland, and England and Wales, 2002-2012. Population, Space and Place 22(6), 508–525, http://dx.doi.org/10.1002/psp.1936.
Watkins, S. C. (1990): From local to national communities: The transformation of demographic regimes in Western Europe, 1870–1960. Population and Development Review 16(2), 241–272, https://doi.org/10.2307/1971590.

Data sources

Fertility indicators. Luxembourg, Eurostat Data Base. (cited on May 13, 2019). http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_find&lang=en.
Demographic Yearbook 2017. Belgrade, Statistical Office of the Republic of Serbia. http://publikacije.stat.gov.rs/G2018/PdfE/G201814015.pdf.
Population by Age and Sex, Book2, 2002 Census of Population, Households and Dwellings in the Republic of Serbia. Belgrade, Statistical Office of the Republic of Serbia (In Serbian). http://publikacije.stat.gov.rs/G2002/Pdf/G20024002.pdf.
Population by Age and Sex, Book2, 2011 Census of Population, Households and Dwellings in the Republic of Serbia. Belgrade, Statistical Office of the Republic of Serbia. http://publikacije.stat.gov.rs/G2012/Pdf/G20124002.pdf.

Effect of rural transport infrastructure on the intensification of purchased input use for major food crop production: the case of smallholder farmers in Horro Guduru Wollega Zone, Western Ethiopia

Sileshi Tamene*, Tebarek Lika Megento

Department of Geography and Environmental Studies, Addis Ababa University, Ethiopia
* Corresponding author: sileshitamene@gmail.com

ABSTRACT

This paper examines the effect of rural transport on smallholder farmers' purchased input use. A random sample of 500 respondents was selected and relevant data was collected. Descriptive, correlation, and regression statistics were used to analyze the data. The multiple linear regression analysis revealed that farmers' purchased input use was found to be significantly and negatively related to distance to major market, distance to all weather road, distance to farm plot, transport cost, and size of land holding. In contrast, farmers' purchased input use was found to be significantly and positively related to family size, off farm income, membership in a cooperative, being in Horro district, having animal cart, and access to good road. Further, the results of hierarchical multiple regression showed that approximately 82% of the total variation in purchased input use can be explained by the linear combination of all independent variables. Furthermore, the result showed that rural transport infrastructure-related variables, as a set, contributed 13.3% to the prediction of farmers' purchased input use over and above the remaining predictors. The results suggest that improving the rural road infrastructure and access to rural transportation services is vital in encouraging farmers' purchased input use.

KEYWORDS

hierarchical regression; Horro Guduru; intensification; purchased input use; rural transport

Received: 13 January 2019
Accepted: 10 September 2019
Published online: 19 December 2019

Tamene, S., Megento, T. L. (2019): Effect of rural transport infrastructure on the intensification of purchased input use for major food crop production: the case of smallholder farmers in Horro Guduru Wollega Zone, Western Ethiopia. AUC Geographica 54(2), 168–181
https://doi.org/10.14712/23361980.2019.15

1. Introduction

The achievement of major food crop intensification remains the greatest challenge facing smallholder farmers in developing countries. The poor state of the rural transportation network and inefficient logistics continue to hinder agricultural intensifications in Africa (Delaney et al. 2017). The key rural transport infrastructure challenges are inadequate and poor conditions of the rural road network and limited availability of vehicles, which has led to an increase in the cost of transportation further affecting agricultural input prices (Salami et al. 2010). The very poor condition of transport infrastructure, the effect of geographic isolation, high transport costs, and time lost to roadblocks, can completely undermine the returns to investments in crop intensification practices (Delaney et al. 2017).

Like in many other sub-Saharan African countries, small-scale agriculture is the most important sector for achieving sustainable household food security in Ethiopia. It accounts for 80% of the working population, 86% of the total foreign exchange earnings, and 48% of gross domestic product (GDP) (Worku 2011). An increasing use of modern agricultural inputs such as chemical fertilizers and improved seeds remains one of the best hopes for greater agricultural production and productivity of rural Ethiopia, where more than 80% of the population lives (Minten et al. 2013). However, lack of enhanced supply and promotion of improved seeds, organic and inorganic fertilizers, and low level of an irrigation system are major obstacles to sustain the agricultural production in the country (Elias et al. 2006).

As a landlocked country with largely non-navigable rivers, road transport plays a significant role in the performance of the Ethiopian economy. In Ethiopia immature rural transport and other key physical infrastructure have led to high transport costs for agricultural products to the market as well as of farm inputs, reducing farmers' competitiveness (Fufa and Hassan 2006; Lulit 2012). This could be a disincentive to the use of productivity-enhancing agricultural inputs and therefore could discourage smallholder farmer's major food crop intensification.

Most of the literature on small-scale agriculture market participation concerns only the output side of marketing production (Arethun and Bhatta 2012; Bekele et al. 2010; Gebremedhin and Hoekstra 2008). However, the sustainable marketing of smallholders also requires integration into the input markets (Pingali and Rosegrant 1995). To bridge the gap in the literature on the marketing participation of rural household on the input side, we analyze the rural transport determinants of purchased modern agricultural inputs use. In Ethiopia, there is relatively large literature dealing with the role of technological innovation and diffusion in increasing agricultural productivity and intensification (Elias et al. 2013; Katungi et al. 2011; Weir and Knight 2004). Past agricultural research in the country also focused on the impact of improved agricultural technologies on smallholder farm income and its implication for poverty reduction strategies (Hailu 2014; Katungi et al. 2011; Salami et al. 2010).

Although these past studies provided useful information on the trends, patterns, and determinants of agricultural input adoption, rigorous assessments of rural transport constraints on the use of purchased (variable) inputs for major food crop production, as a measure of household commercialization from the input side have rarely been studied in Ethiopia. A better understanding of rural transport constraints that hinders smallholder farmers' participation in agricultural input markets as a buyer is therefore important for designing promising pro-poor agricultural and transport policies that could stimulate the use of modern agricultural inputs and increase small-scale agricultural production. This study aims to fill this knowledge gap and provide quantitative information to empirically address the relationship between smallholder farmers' food crop intensification and rural transport infrastructure. The findings of this research are supposed to be used by different stakeholders involved in rural transport policy, agricultural land use planning, and sustainable food crop production strategies.

The overall objective of this study was to investigate rural transport constraints of major food crop intensification strategies by small-scale farmers in Ethiopia. The specific objectives of this study were to: (1) examine relationship between proximity to all-weather roads and total values of purchased input use among smallholder farmers; (2) determine the extent to which a combination of rural transport infrastructure, institutional factors, resource endowment, and physical factors predict smallholder farmers' purchased agricultural input use; (3) identify the extent to which rural transport infrastructure (distance to major market, distance to all-weather road, distance to farm plot, transport cost, mode of transport and road conditions) predicts smallholder farmers' purchased agricultural input use, controlling for the effects of demographic, institutional, resource endowment, and physical factors.

2. Theoretical underpinnings of the study

This study aims to look at the nature and extent of rural transport infrastructure and its effect on modern agricultural input use among smallholder farmers of Horro Guduru Wollega Zone, Western Ethiopia. There are several theories that attempted to explain how rural transport infrastructure investment can bring about economic growth and development (Banerjee et al. 2012; Didenko et al. 2017; Jelilov and Kachallah 2017; Margarian 2011; Roland-holst 2009).

Of these multiple theories, “the theory of induced technical and institutional change” was used as relevant theoretical perspective to design the research questions that this study is based on. In addition, this theoretical perspective was used for organizing and interpreting the findings of this study. The theory of induced technical and institutional change provides the structure to define how this particular research will philosophically, epistemologically, methodologically and analytically approached.

Economic historians are increasingly drawing on the theory of induced technical change in attempting to interpret differential patterns of productivity growth among countries and over time (Ruttan 2008; Ruttan and Hayami 1984). Agricultural economists like Hayami and Ruttan (1993, p. 6) argue by saying that in agriculture, changes in the ‘relative resource endowments’, especially land and labor, induce a derived demand for technological innovations to facilitate the replacement of relatively less scarce and cheap factors for more scarce and expensive ones. For example, when labor is in short supply, there is a tendency for capital in the form of labor-saving machinery to be substituted for human labor. Whereas, in a land-scarce economy, yield-increasing and land enhancing inputs such as fertilizers and improved seeds are substituted for land which in turn depends on the agricultural input market conditions (Hayami and Ruttan 1985, 1993; Ruttan 2008). Moreover, roughly 22 years ago, Ruttan (1996, p. 54) in one of his seminal papers addressed that induced technical change acts to make the ‘scare factor more abundant’.

Induced technical and institutional change theory offers a theoretical understanding appropriate to examine the complex and dynamic relations between rural transport access and smallholder farmers’ purchased input use. When we examine the appropriateness of this theory for understanding the nature of rural transport and purchased input use in the study area, two related realities emerge: first, the effort of promoting the rural agricultural economy and bringing maximum benefits to smallholder farmers needs technical change and innovations to transform the most common and tedious traditional rural transport mechanisms-human porterage (head, shoulder, and back-loading) to improved rural transport means-pack animals and animal drawn carts. Such rural transport improvements brought about by rural transport innovations can enhance the use of purchased input use among smallholder farmers. Second, this theory also acknowledges that improved production technologies (fertilizer and improved seed) as well as improved farm management practices (credit and extension institutions) can play in replacing relatively scarce resources like land and labor.

3. Data and Methods

3.1 Selection and description of study site

The study was conducted in Horro Guduru Wollega Zone, western Ethiopia. This Zone lies between Latitude 9°10′ N and 9°50′ N and Longitude 36°00′ E and 36°50′ E (Figure 1). It has a total land area of 8,097 km^2 (CSA 2011; Tamene and Megento 2017). Shambu is the capital town of the zone and found 314 km West of Addis Ababa. According to the report of CSA (2011), this zone had a total population of 641,575 of which 50.09% are male and 49.91% are female. According to the same source, about 89% of the population lives in the rural areas driving their livelihoods from agriculture.

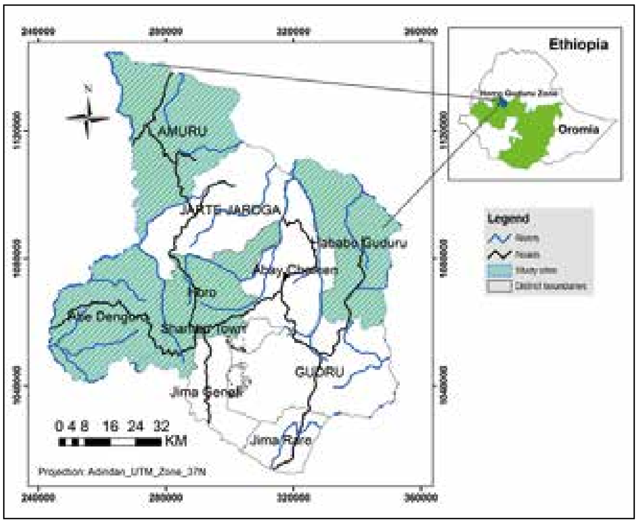

Fig. 1 Location map of Horro Guduru Wollega Zone.
Source: Adapted Finance and Economic Development Bureau of Oromia, 2016.

The average annual temperature in the study area is 22.1 °C, with an average minimum of 13 °C and an average maximum of 30 °C (Beyene et al. 2015). The average altitude of Horro Guduru Wollega Zone ranges from 860 to 2657 meters above sea level (Beyene et al. 2015). Mixed crop-livestock agriculture is the mainstay in the study area with notable food crops including wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), teff (*Eragrostis tef*), maize (*Zea mays*), pulses (*Vicia faba*, *Pisum sativum*) and cash crops like sesame (*Sesamum orientale*), niger (*Guizotia abyssinica*), and linseed (*Linum usitatissimum*) (CSA 2014).

Even though few independent variables used in our previously published article (Tamene and Megento 2017) are also used in current manuscript, a greater number of independent variables are newly introduced to the current manuscript to emphasize on the originality of this paper. Furthermore, due the following reasons each manuscript has a distinct focus and purpose. First, each manuscript addressed different research questions. Second, each manuscript studies the data from completely different angles. Third, each manuscript used different relevant literature. Fourth, they both differ in their analytic methods, interpretations, and conclusions. Lastly, the dependent variables in both manuscripts are conceptually different and empirically not related. Therefore, the manuscripts should be considered independently.

3.2 Study Design, Sampling, Data Collection, and Analysis

The study utilized a smallholder household based cross-sectional quantitative survey design using a structured questioner with face to face interview. Descriptive and analytical cross-sectional micro-level data have been used to estimate the effect of rural transport infrastructure on smallholder farmers' purchased input use. Horro Guduru Wollega Zone was identified as one of the potential cereal crop producing corridors of Ethiopia. On the contrary, the existing rural road transport infrastructure in the zone is not satisfactory to support smallholder agricultural production system including the purchases of necessary inputs (Tamene and Megento 2017). Therefore, Horro Guduru Wollega Zone was purposively selected.

A multistage simple random sampling technique was sequentially employed to select four districts from Horro Guduru Wollega Zone, four rural kebeles (RKs) from each district, and rural farm households from each rural kebele-Kebele is the smallest administrative unit in Ethiopia. The first stage involves a random selection of four districts from the nine districts of Horro Guduru Wollega Zone. As a result, four districts (Hababo Guduru, Horro, Amuru and Abe Dondoro) were selected (Figure 1). The second stage involves the random selection of four RKs from each of the four districts making a total of 16 RKs.

The third and final stage was the random selection of farm households with farmland size of 0.25 ha and above from each RK. The list of farm households in each RK was compiled with the assistance of the extension agents and RK managers. According to Gray et al. (2007) suggestion, the researcher used a 95% confidence level to determine the sample size for this specific study. It is usual that RKs may vary considerably in the number of smallholder farmers they contain and hence to avoid bias, probability proportional to size (PPS) was employed. Thus, 500 smallholder farmers from the four districts were sampled for the study.

The household survey was conducted from February to June 2016, which followed shortly after the main season (Meher) harvest. Interviews were conducted in places convenient to farmers either at home or in the field.

Statistical analysis such as descriptive statistics, correlation, and multiple regression with a hierarchical model specification was performed using the Statistical Package for Social Sciences (IBM SPSS) software program version 20. To determine whether rural transport infrastructure added significantly to the prediction of smallholder farmers' purchased input use, over and above the variance predicted by demographic, household resource endowment, institutional infrastructure, and location specific variables, the independent variables were entered in to two separate blocks of the regression analysis (hierarchical method). In the first block, all independent variables except rural transport infrastructure related variables (variables of interest) were included into the regression analysis. At the second step, the new independent variables (rural transport infrastructure related variables) were added, and all of the independent variables (those entered at first block) remain in the independent set to see the unique contribution of rural transport infrastructure related variables in the intensification of purchased agricultural input use.

3.3 Description of Variables

Our variables of interest (both dependent and independent) used in this study and their levels of measurement are shown in table 1. These variables are supposed to capture the influence of the potential independent variables on the purchased input use as a dependent variable.

3.3.1 Rural transport infrastructure

It is assumed that a well-functioning rural transportation infrastructure is significant determinants of the form and pace of food crop intensification of smallholder farmers. Smallholder level purchased input use for major food crop production (Y) is therefore modeled as a function of smallholder farmers' access to rural transport infrastructures and services. Access to rural transport infrastructures and services, captured as an average distance to major market, average distance to farm plot, average distance to all weather-road, are expected to be negatively correlated to the total values of purchased input use (Bekele et al. 2010). Availability of good quality rural access road is considered crucial to improving access to agricultural input markets, resulting in greater use of productivity-enhancing modern agricultural inputs (Jayne et al. 2003). Ownership of Intermediate Means of Transportation for local-level transport services in rural areas is also expected to promote increased use of purchased modern inputs. Hence, household level purchased input use is modeled as a function of distance to major market, distance to farm plot, distance to all weather-road, transport cost, ownership of Intermediate Means of Transportation and road quality.

3.3.2 Demographic factors

To capture the effects of demographic factors on smallholder farmers' purchased input use, we used the age of household head and family size. It was assumed that older households tend to be more

Tab. 1 Summary of variables used and their measurements.

Variable name and type	Description	Measurement	Expected sign
Purchased input use	A dependent variable indicating the total amount of birr spent for purchasing agricultural inputs for major food crops	Ethiopian Birr*	
Age of household head (continuous)	Age of household head	Number of years	–
Family size (continuous)	Total family size of the household head	Number	+
Size of land holding (continuous)	Area of farm land owned by the household	Hectare	+
Livestock ownership (continuous)	Total livestock ownership of the household	Total livestock unit (TLU)	−/+
Off farm income (continuous)	Income earned from non-agricultural activities	Ethiopian Birr*	+
Membership in a cooperative (Dummy)	Being a member of an agricultural cooperatives	1 = Yes; 0 = No	+
Extension visit monthly (Dummy)	Frequency of extension visit	1 = Monthly visit; 0 = Twice in a year	+
Level of annual rainfall (continuous)	Amount of annual precipitation	Millimeter (mm)	−/+
Distance to major market (continuous)	Distance travelled by the household to reach the nearest major market	Kilometer	–
Distance to farm plot (continuous)	Average farm plot distance from the homestead	Kilometer	–
Distance to all weather road (continuous)	Distance travelled by the household to reach the nearest all weather road	Kilometer	–
Transport cost (continuous)	Transport cost incurred to move 100 kg of agricultural input over 1 km	Birr per 100 kg per km	–
Animal cart (Dummy)	Transport mode used	1 = Animal cart; 0 = Headloading	+
Good road (Dummy)	Road quality	1 = Good road; 0 = Bad road	+
High land (Dummy)	Agro-ecology type	1 = Highland; 0 = Otherwise	−/+
Horro (Dummy)	District type	1 = Horro district; 0 = Otherwise	–

Source: Own construction.
* During data collection 1 USD equals 23.73 Ethiopian birr.

market-oriented and have higher agricultural input market participation. The number of members in the household is also assumed to be very important in the use of purchased input use since the use of such productivity-enhancing inputs are said to be more labor-intensive than conventional subsistence farming.

3.3.3 Household resource endowment

Off-farm income, livestock ownership and size of landholding increase the welfare of farmers because they help farmers in getting the required input needed for agricultural production (Abdullah et al. 2017). Purchased agricultural inputs are mainly financed through cash from off-farm activities and livestock sales (Christiaensen 2017). Therefore, household resource endowment (size of land holding, livestock ownership, and off farm income) are hypothesized to increasingly recognized as a resource that can significantly influence the quantity of purchased input use among smallholder farmers.

3.3.4 Institutional infrastructure variables

Apart from household resource endowment and demographic factors several other factors such as the institutional frameworks affect the demand for and consumption of purchased farm inputs such chemical fertilizers (DAP and urea), improved seeds (maize and wheat) and pesticides. Therefore, the third important set factors affecting purchased input use among smallholder farmers in rural areas include the institutional infrastructure variables such as membership in a cooperative (Hellin et al. 2009; Markelova et al. 2009) and contact with agricultural extension agents (Belay 2015).

3.3.5 Physical and location specific factors

Several environmental variables were hypothesized to encourage/discourage farmers to invest in purchased agricultural input use for major food crop production. These include the amount of annual rainfall received and agro-ecology. The total values of purchased input use and physical factors (amount of annual rainfall received and agro-ecology where a household belongs) are related. It is expected that, on average, smallholder farmers in midland agro-ecology and with sufficient rainfall tend to use more purchased inputs. To capture the differences in purchased input use among study districts, district dummy was considered.

3.3.6 Rural transport infrastructure

It is assumed that a well-functioning rural transportation infrastructure is significant determinants of the form and pace of food crop intensification of smallholder farmers. Smallholder level purchased input use for major food crop production is therefore modeled as a function of smallholder farmers' access to rural transport infrastructures and services. Access to rural transport infrastructures and services, captured as an average distance to major market, average distance to farm plot, average distance to all weather-road, are expected to be negatively correlated to the total values of purchased input use (Bekele et al. 2010). Availability of good quality rural access road is also considered crucial to improving access to agricultural input markets, resulting in greater use of productivity-enhancing modern agricultural inputs (Jayne et al. 2003). Ownership of Intermediate Means of Transportation for local-level transport services in rural areas is also expected to promote increased use of purchased modern inputs.

The above described independent variables were entered in to into a hierarchical linear regression analysis through sequential block-enter approach so as to predict smallholder farmers' purchased input use. The choice of how to include independent variables was determined by researchers based the overall purpose of the analysis. In the first block of hierarchical linear regression analysis, demographic variables, household resource endowment variables, institutional infrastructure variables, and physical and location specific variables were entered as the first block of independent variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{10} X_{10} + \varepsilon \quad (1)$$

Since we are also interested in examining the effect of rural transport infrastructure variables on smallholder farmers' purchased input use after demographic variables, household resource endowment variables, institutional infrastructure variables, and physical and location specific variables have been controlled for, we entered rural transport infrastructure variables (variables of interest) in the subsequent blocks of independent variables in the hierarchical linear regression analysis (block 2). Therefore, to see if the rural transport infrastructure variables predict smallholder farmers' purchased input use above and beyond the effect of the controls and to test if successive model fit better than previous one the following model was developed.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{16} X_{16} + \varepsilon \quad (2)$$

Here in the case of both models:
Y is the financial cost or monetary value of variable inputs (Ethiopian birr),
β_0 is the constant term or the intercept,
β_{1-16} are the regression coefficients associated with respective independent variables
X_1 is age of household head (years),
X_2 is family size (number),
X_3 is size of land holding (hectare),
X_4 is livestock ownership (TLU),
X_5 is off farm income (Ethiopian birr),
X_6 is membership in a cooperative (dummy),
X_7 is extension visit monthly (dummy),
X_8 is level of annual rainfall,
X_9 is high land (dummy),
X_{10} is Horro (district dummy),
X_{11} is distance to major market (km),
X_{12} is distance to all weather road (km),
X_{13} is distance to farm plot (km),
X_{14} is transport cost (Ethiopian birr),
X_{15} is animal cart (Dummy),
X_{16} is good road (Dummy),
ε is the random error component reflecting the difference between the observed and fitted linear relationship.

After fitting a hierarchical linear regression model and computing the parameter estimates some predictor variables are omitted since they are much less important or most likely affect the explanatory power of the model if included.

4. Results

4.1 Preliminary analyses

There are a number of assumptions that must be met for multiple linear regression model to be reliable (Osborne and Waters 2002). Preliminary analysis to ensure the non-violation of the assumptions of normality (Kim 2015; Miot 2017; Yap and Sim 2011), linearity (Osborne and Waters 2002), homoscedasticity and multicollinearity (Daoud 2017; Friday and Emenonye 2012; Imdadullah et al. 2016) were completed prior to the analysis. The preliminary analysis revealed that these assumptions were not seriously violated.

4.2 Descriptive statistics

Descriptive statistics were used to summarize data. Table 2 shows a descriptive analysis of continuous variables by presenting numerical facts about the quantitative dataset. The total number of observations (n) was 500 smallholder farmers in four districts of Horro Guduru Wollega Zone, Western Ethiopia. The mean score for household size was 6.48, with a standard deviation of approximately 3.31 points,

Tab. 2 Descriptive statistics for continuous variables (n = 500).

Variable	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic
Age of household head	17	64	34.29	13.02	.56	−.77
Family size	2	12	6.48	3.31	.19	−1.29
Size of land holding	.4	6.0	2.41	1.18	.45	−.48
Livestock ownership (TLU)	.55	11.51	3.26	1.64	.95	1.59
Off farm income	0	9100	2416.30	2730	.89	−.68
Distance to major market	5	35	20.18	7.87	−.15	−1.16
Distance to farm plot	.4	13.0	6.41	3.38	−.08	−1.07
Distance to all weather road	0	27	12.97	6.76	−.16	−.79
Transport cost	.50	3.20	2.18	.78	−.53	−.74
Purchased input use	717.50	37644.43	10096.92	9357	1.44	1.09

Source: Compiled from field data, 2016.

figures which are above the national average of 4.6 persons (CSA 2017).

The average distance to the nearest major market is about 20.18 km, indicating poor market access of smallholders in the study area, while the average distance to all-weather roads is 12.97 km. The values for skewness and the kurtosis indices are very small and fall within the acceptable range which indicates that the variables most likely do not include influential cases or outliers (see Table 2).

The average purchased input (chemical fertilizer, improved seed, and herbicides) value used for major food crop production (maize, wheat, and teff) is ETB 10096.92. The result also indicates that on average smallholder farmers get about ETB 2416.30 income from off-farm employment. A household on average operates about 2.41 ha (SD = 1.18), a result which is two times greater than the national average of 1.14 ha (CSA 2015; Tamene and Megento 2017). Finally, on average, households incur a 2.18 ETB to transport one kg of farm input for 100 km from input market center to home.

4.3 Correlation analysis

The strongest negative significant Pearson product-moment correlation coefficient for the total values of purchased input use was with the proximity to all-weather roads: $r(498) = -.772$, $p < .001$. Whereas the strongest positive Pearson product-moment correlation coefficient for the total values of purchased input use was with off-farm income: $r(498) = .654$, $p < .001$ (Table 3). The simple coefficient of determination ($r^2 = 0.596$) indicated that distance to all weather road explains 59.6% of the variation in total values of purchased input use by smallholder farmers. Approximately, the other 40% of the total variance between distance to all weather road and purchased input use remains unexplained. This value of simple coefficient of determination shows the binary linear relationship between distance to all weather road and purchased input use in the absence of other independent variables. Thus distance to all weather road seems to explain a significant amount of variation in purchased input use.

Tab. 3 Pearson correlations of purchased input use (dependent variable) and independent variables.

Independent Variables	Correlation coefficient (r)	P value
Age of household head (years)	−.419**	.000
Family size (no.)	.571**	.000
Size of land holding (ha)	−.416**	.000
Livestock ownership (TLU)	.263**	.000
Off farm income (ETB)	.654**	.000
Distance to major market (km)	−.745**	.000
Distance to farm plot (km)	−.467**	.000
Distance to all weather road (km)	−.772**	.000
Transport cost (km)	−.727**	.000
Level of annual rainfall (mm)	.114*	.011

n = 500, ** $P < .001$ (2-tailed), * $P < .05$ (2-tailed)
Source: Compiled from field data, 2016.

4.4 Hierarchical multiple regression

A multiple regression analysis was conducted to examine whether rural transport infrastructure, institutional factors, resource endowment, and physical factors could significantly predict smallholder farmer's purchased input use. The results of multiple regression analysis using all sixteen independent variables are summarized in table 4 of model 2. A strong relationship between purchased input use and the independent variables was observed and the model was a significant predictor of purchased input use ($R^2 = .822$, $F(16, 483) = 138.95$, $p < .001$).

The multiple coefficient of determination (R^2) value of 0.822, indicates that approximately 82.2% of the total variation in purchased input use can be accounted for by the linear combination of explanatory variables. But the remaining 17.8% of the variance has been attributed to other variables not included in the model and disturbance term. The adjusted R^2 value is 0.816, which is very close to the multiple R^2, indicating that we shouldn't worry too much about whether we are using too many variables in the model.

4.4.1 Rural transport infrastructure

The estimated unstandardized regression coefficients displayed in table 4 showed the relative importance of each predictor in the model. Among rural transport infrastructure variables, transport cost was the strongest negative predictor as indicated by its estimated unstandardized regression coefficient, ($\beta = -3707.88$, $p < .001$) followed by good road dummy ($\beta = 1832.57$, $p < .001$) and animal cart dummy ($\beta = 1090.49$, $p < .05$). The unstandardized regression coefficients for the association between distance to major market, distance to all weather road and distance to farm plot on one hand and smallholder

Tab. 4 Hierarchical egression for variables predicting purchased input use ($n = 500$).

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	22750.229	11824.431		1.924	.055
	Age of household head	−98.328	23.189	−.137	−4.240	.000
	Family size	651.186	102.507	.230	6.353	.000
	Size of land holding	−2572.812	288.302	−.325	−8.924	.000
	Livestock ownership (TLU)	577.724	205.493	.101	2.811	.005
	Off farm income	1.240	.113	.362	10.989	.000
	Membership in a cooperative	2801.652	602.726	.145	4.648	.000
	Extension visit monthly_Dummy	571.666	542.847	.027	1.053	.293
	Level of annual rainfall (mm)	−8.864	7.432	−.038	−1.193	.234
	High land_Dummy	2078.012	725.521	.100	2.864	.004
	Horro_Dummy	1046.011	611.352	.044	1.711	.088
2	(Constant)	22594.157	10233.971		2.208	.028
	Age of household head	−7.988	18.552	−.011	−.431	.667
	Family size	189.180	82.571	.067	2.291	.022
	Size of land holding	−1070.479	246.050	−.135	−4.351	.000
	Livestock ownership (TLU)	38.776	168.814	.007	.230	.818
	Off farm income	.812	.092	.237	8.848	.000
	Membership in a cooperative	1457.595	467.319	.076	3.119	.002
	Extension visit monthly_Dummy	722.301	424.473	.035	1.702	.089
	Level of annual rainfall (mm)	1.000	6.519	.004	.153	.878
	High land_Dummy	−269.070	574.761	−.013	−.468	.640
	Horro_Dummy	3379.733	492.646	.142	6.860	.000
	Distance to major market	−160.575	46.674	−.135	−3.440	.001
	Distance to all weather road	−316.427	50.336	−.229	−6.286	.000
	Distance to farm plot	−215.478	61.132	−.078	−3.525	.000
	Transport cost	−3707.879	380.257	−.310	−9.751	.000
	Animal cart_Dummy	1090.493	548.139	.044	1.989	.047
	Good road_Dummy	1832.569	457.855	.096	4.003	.000
a. Dependent Variable: Purchased input use						

Note: $R^2 = .689$ for model 1, $R^2 = .822$ for model 2 and $\Delta R^2 = .133$.
Source: Compiled from field data, 2016.

farmers' purchased input use on the other hand are −160.57, −316.43 and −215.48 respectively; the associated standard errors for these regression coefficients are 46.67, 50.34 and 61.13 respectively.

4.4.2 Household demographics

The results of the regression model showed that family size is an important factor identified to influence purchased input use (β = 189.18, t(485) = 2.29, $p = .02$).

4.4.3 Household resource endowment

Size of land holding significantly predicted purchased input use, β = −1070.48, $t(485) = -4.35$, $p < .001$. Off-farm income is positively associated with total values of purchased input use such that, holding everything else constant, for each additional ETB off-farm income, the total values of purchased input use is predicted to increase by .81 ETB, and this association is statistically significant (p < .001).

4.4.4 Institutional infrastructure variables

As can be seen in table 4, membership in a cooperative had a significant positive regression coefficient, indicating smallholder farmers who are members of cooperative associations were expected to invest ETB 1457.59 more than the nonmembers, after controlling for the other variables in the model, and this result is statistically significant ($p = .002$).

4.4.5 Location specific factors/district dummy

The value associated with being a farmer living in Horro district is ETB 3379.73 (adjusting for the other variables in the model), and the coefficient on this dummy variable is both positive and statistically significant (p < .001).

4.4.6 The unique contribution of rural transport infrastructure related variables to purchased input use

The percent of the variability in the purchased input use that can be accounted for by all the predictors together is 82.2%. This is a significant contribution and hence is an excellent model. The change in variance accounted for (ΔR^2) was equal to .689 − .822 = .133, which was significantly different from zero $F(6, 483) = 59.76$, $p < .001$. In this case, the percentage of variability in purchased input use accounted for went up from 68.9% to 82.2%. The F change associated with R^2 change of .133 is statistically significant showing that adding rural transport infrastructure variables to the model increases the model's predictive capacity. This is to mean that rural transport infrastructure explained additional13.3% of the variance in purchased input use, after controlling for the possible effects of potential confounding variables [$\Delta R^2 = 0.133$, $\Delta F\ (6, 483) = 59.764$, $p < 0.001$]. The unstandardized regression coefficients (β), intercept, and the standardized regression coefficients (*Beta*), for the full model are reported in table 5.

5. Discussion

5.1 Family size and purchased input use

The results of hierarchical multiple regression revealed a positive relationship between family size and purchased input use indicating that large families spend more on purchased input use as compared to small families. There are many prior research findings that explain how the number of family members' influences purchased input use decisions of smallholder farmers. For example, Nambiro (2008) found a significant positive effect of family size on the proportion of farm allocated to the cultivation of improved hybrid maize seed. Another study by Kamara (2004) and Perz (2003) found availability of family labor as a precondition for greater use of chemical fertilizers, pesticides, and improved seeds. Therefore, labor-augmenting technologies are important to encourage small families so that they can take part in input market participation. As a result, this finding supports induced technical innovation theory that agricultural equipment designed for use in small farm plots make it feasible for farmers to shift from labor-intensive practices to higher-yielding mechanized practices (Hayami and Ruttan 1993).

5.2 Household-level asset variables and purchased input use

5.2.1 Landholding

This study found a negative and statistically significant relationship between farm size and smallholder farmer's purchased input use. Past studies have found a mixed result. For instance, the largest share of households renting mechanization is more likely related to large farming size (Diao et al. 2016; Ma et al. 2018). In addition, Ma et al. (2018) investigated the determinant role of increased farmland values on the level of supplementary feed used for dairy production. Kiplimo and Ngeno (2016) and Hung et al.(2007) also found a negative relationship between a continuous reduction in farm size (farm fragmentation) and farm household level input use. In contrary to these findings, FAO (2015) found the use of seed and fertilizer technologies to be scale-neutral, which is their use intensity does not depend on farm size. FAO further underlined that, since fertilizer is 'a land augmenting input', smallholders use it intensively, probably to substitute for land (p. 10). Our finding goes in line with FAO's finding supporting the theory of induced technical innovation which advocates the need to substitute the relatively less scarce and cheap factors of production (fertilizer) for more scarce and expensive ones (land) (Hayami and Ruttan 1985;Hayami and Ruttan 1993).

5.2.2 Off-farm income

The study revealed that income from off-farm sources positively influenced the application of complementary inputs such as fertilizer, pesticides, and high yielding seed varieties. This is because off-farm income was an important source for smallholder farmers to increase their market access to agricultural inputs to intensify production. Findings of this study are in line with those of Dahal et al. (2007), who concluded that off-farm income earning opportunities drive smallholders towards agricultural intensification. Kamara (2004) also found that farmers' access to adequate and sustainable off-farm income has a significant effect on their use of modern agricultural inputs. Moreover, in their seminal work, Lim-Applegate et al. (2002) pointed out the significance of off-farm employment as a source of income for Australian farm families.

5.3 Cooperative membership and purchased input use

Membership in agricultural cooperatives is among the variables that determine the propensity of smallholder farmer's participation in the agricultural input market. In this research, it was expected to have a positive influence. Accordingly, membership in farmer cooperatives was found to significantly influence the level of participation in agricultural input marketing. There are several points that help us to maintain this view. Firstly, agricultural cooperatives play a pivotal role in subsidizing fertilizer and seed distribution. Secondly, participating in farmer organizations has the potential to secure better prices for produce. And thirdly, they also play a key role in improving farmer's access to technical advice. The results thus obtained are compatible with previous studies. For instance, in their research findings, Birachi et al. (2011) indicated that membership to cooperative society was the significant driver of agricultural commercialization among food crop farmers in Burundi. Furthermore, Carrer et al. (2018) found significant and positive relationship between participation in pools (cooperatives) and the adoption of forward contracts among citrus growers in the State of Sao Paulo, Brazil.

5.4 Rural transport infrastructure and purchased input use

5.4.1 Distance to major market

Access to agricultural input markets is expected to be negatively correlated with the total values of purchased input use. It is, therefore, hypothesized that reduced distance to major market will positively affect smallholder farmers' purchased input use. As expected, there was a negative and significant association between distance to the nearest major market and total values of purchased input use. The current results confirmed the hypothesis that farmers with reduced physical distance to input markets have a higher probability of using modern agricultural inputs than those who are remote (Hailu and Fana 2017). This is partly because the costs of obtaining agricultural inputs such increase more quickly with increased distance to input supply centers. This finding also corroborates the ideas of Nin-Pratt (2016) and Katungi et al. (2011), who suggested that greater access to urban markets increases the intensity of input use and productivity in the rural sector.

5.4.2 Distance to farm plot

As expected, the regression result showed that plot distance from the homestead has a negative and significant relationship with total values of purchased input use. It may be the case therefore that the more remote the farm plot from farmer's residence, the lesser would be the probability of purchased agricultural input utilization. This result is in line with the findings of a great deal of the previous work in this field. For instance Hailu et al. (2014) found a statistically significant negative relationship between plot distance from the homestead and probability of chemical fertilizer adoption decision. By using a dataset from Ghana, Kotu et al. (2017) also found that plots located adjacent to the homestead are more likely to adopt sustainable agricultural intensification practices than the more distant ones.

5.4.3 Distance to all weather road

The current study found a negative and statistically significant association between distance to all weather road and total values of purchased input use. It is evident that the cost of transport is determined by taking account of road roughness and seasonality. All weather road was reported from earlier studies to be an important variable which explains variations in purchased input use. For example, in Madagascar, Ninnin (1997) reported that dry season fares were less costly than wet season fares. By using data drawn from longitudinal Ethiopian Rural Household Surveys, Wondemu and Weiss (2012) also reported that improving the class of rural roads to a degree that allows all-weather road access sharply increases average household income. They further established that with the equal level of farmland ownership, having paved road access allows a smallholder farmer to generate 82% higher income than would be the case with poor access road. Using cross-sectional data, Beshir (2014) also examined the factors that affect the probability of improved forage seeds adoption in two districts of South Wollo zone, Ethiopia. He found a negative and statistically significant relationship between distance to all weather road and the probability of adoption and intensity of use of improved forage seeds.

5.4.4 Transport cost

The coefficient of transport cost incurred had the expected negative sign and significant effect on the total values of purchased input use (see Table 4). This is because, on the whole, it has been established that there is a strong correlation between transport cost incurred and the ability of smallholder farmers to purchase and use modern agricultural inputs. This finding is consistent with previous studies of Kotu et al. (2017), who reported that because of inefficient input markets characterized by high transaction and transport costs, in Ghana, farmers mostly pay higher than official prices for nitrogen, phosphorus and potassium (NPK) fertilizer. In another study in Australia, Freebairn (2003), reported that Australians who live in remote rural areas will incur additional transport costs to get access to some services offered in large urban areas. Similarly, according to Wondemu and Weiss (2012), high input prices due to lack of infrastructure, such as underdeveloped rural road networks have led to high transport costs for farm inputs, thus holding back farmers' demand for purchased input use.

5.4.5 Mode of transport

The survey showed a marked variation in the purchased input use among smallholder farmers by type of transport mode owned and used. This research found that an improvement in the mode of rural transport use from head-loading to animal cart will result in an additional 1090.493 birr investment in purchased input use by smallholder farmers. The supply of agricultural inputs was expected to increase substantially with the increased probability of modal shift. The overall efficiency of the transport mode used can seriously affect access to farm inputs. As a result, with respect to transport mode owned and used, modern input use was expected to be higher for smallholder farmers who owned animal cart as compared to those who use the various methods of human porterage (head, shoulder, and back-loading). Our finding reveals that ownership of pack animals is the most deriving factor for input market participation where the difference in purchased input use was seen among those who own and not.

Past empirical findings suggested that limited access to an improved mode of transport that helps to move farm inputs from input delivery center to homestead and from homestead to farm remains a major challenge for smallholder farmers (Hine 2004). According to Zewdie (2015), those households who face binding transport constraints may be unsuccessful to afford the maximum desired levels of input use. The traditional mode of transport like human porterage negatively affected the level of input use for agricultural production in Nigeria (Akramov 2009; Orakwue et al. 2015). Yet, the use of improved inputs, such as fertilizer and improved seeds is very low among those smallholders who do not own transport animals as compared to those who own the same. In Ethiopia too, transport mode choice is said to be an increasingly important area in getting access to agricultural inputs for smallholder farmers (Kassa 2014).

5.4.6 Road Condition

Linking smallholder farmers with a good road network was found to be positive and significant in access to and utilization of purchased farm inputs. If all other variables are controlled, a good road condition, as opposed to bad road condition, will result in an additional ETB 1832.569 investment in purchased input use by smallholder farmers. This current study further showed that a good road system from input market to the farms would allow easy and timely access to purchased inputs. This might also partly explain why good-quality roads (paved roads) provide a good stimulus to farm profitability and productivity of the rural economy.

A study by Quan (2009) showed that good physical connections to input markets are a fundamental enabler for smallholder farmer's purchased input use. He further showed that good road access being paramount for farm input commercialization through new technologies that increased the yield of basic food crops. Another recent study in Kenya and Tanzania by Bradbury et al. (2017) explicitly revealed the poor accessibility challenges that smallholder farmers experience in getting agricultural inputs from the market to the farms. In their examination of the transport costs and access constraints for well connected and remote rural farmers of Kenya and Tanzania, they found that smallholders who are linked by a network of unclassified, earth access tracks that are poorly maintained and mostly impassable during the rainy seasons are less likely to use of purchased inputs than those who are connected to good road networks linking farming areas to major markets.

6. Conclusion and policy implications

It is incontestable that rural road connectivity and rural transport services are among the key components for rural development, as it promotes access to economic and social services, stimulating the demand and consumption of purchased agricultural inputs that in turn enhance production and productivity of the farmers. To this end, this study revealed that size of land holding, distance to major market, distance to all-weather road, distance to farm plot, transport cost have a significant negative relationship with purchased input use. Whereas family size, off-farm income, and membership in a cooperative are found to be significant and positively related to smallholder farmers' purchased input use.

Two important policy implications emerged from the study: First, to free smallholder farmers from a vicious cycle of subsistence production, policy reforms in the area of rural infrastructure, access to input markets and to credit facilities must be the central government and local government's rural development top priority. Second, input use intensification need a close policy follow-up so as to enhance production and productivity of farmers. Hence, a policy-mix that can increase smallholder farmers' off-farm income is desirable as income is a critical predictor of improved seed and fertilizer use.

7. Limitations and future research directions

First and foremost, the most obvious limitation of this study is its cross-sectional design. This limitation calls upon future research to adopt a longitudinal (time series) study approach to robustly capture the impact of public infrastructure investment towards determining the exact role of rural transport infrastructure in purchased input use. Second, the environmental costs of modern agricultural input use were overlooked by this study; therefore, future research concerns should consider a more in-depth analysis of the impact of modern agricultural input use on environmental quality. To measure smallholder farmers' purchased input use intensity, the total financial cost (monetary value) of variable inputs was used. Purchased input use in this study is therefore refers to the quantity of money that smallholder farmers spent on major food crop variable inputs (chemical fertilizer, improved seed, and pesticides) in 2015/16 crop production season. Indeed, in order to estimate the extent of purchased input use we used information on quantity of variable inputs used and the prices at which they are purchased. Since the effect of agricultural input use on farmers' productivity is not the concern of this paper, it can be another potential area for future study.

References

Abdullah, Ahamad, R., Ali, S., Chandio, A. A., Ahmad, W., Ilyas, A., Din, I. U. (2017): Determinants of Commercialization and its impact on the Welfare of Smallholder rice Farmers by using Heckman's two-stage approach. Journal of the Saudi Society of Agricultural Sciences 18(2), 224–233, https://doi.org/10.1016/j.jssas.2017.06.001.

Akramov, K. (2009): Decentralization, agricultural services and determinants of input use in Nigeria (IFPRI Discussion Paper No. 00941). Washington, D.C., USA: International Food Policy Research Institute (IFPRI). Retrieved from http://www.ifpri.org/category/publication-type/discussion-papers.

Arethun, T., Bhatta, B. P. (2012): Contribution of rural roads to access to- and participation in markets: Theory and results from Northern Ethiopia. Journal of Transportation Technologies 2, 165–174, https://doi.org/10.4236/jtts.2012.22018.

Banerjee, A., Duflo, E., Qian, N. (2012): On the Road: Access to transportation infrastructure and economic. Paper presented at IGC Conference, London, https://doi.org/10.3386/w17897.

Bekele, A., Kassa, B., Legesse, B., Lemma, T. (2010): Effects of crop commercial orientation on productivity of smallholder farmers in drought-prone areas of the Central Rift Valley of Ethiopia. Ethiopian Journal of Agricultural Sciences 20, 16–34.

Belay, K. (2015). Agricultural extension in Ethiopia: The case of participatory demonstration and training extension system. Journal of Social Development in Africa 18(1), 49–84, https://doi.org/10.4314/jsda.v18i1.23819.

Beshir, H. (2014). Factors affecting the adoption and intensity of use of improved forages in North East Highlands of Ethiopia. American Journal of Experimental Agriculture 4(1), 12–27, https://doi.org/10.9734/AJEA/2014/5481.

Beyene, B., Hundie, D., Gobena, G. (2015): Assessment on dairy production system and its constraints in Horoguduru Wollega Zone, Western Ethiopia. Science, Technology and Arts Research Journal 4(2), 215–221, https://doi.org/10.4314/star.v4i2.28.

Birachi, E. A., Ochieng, J., Wozemba, D., Ruraduma, C., Niyuhire, M. C. (2011): Factors influencing smallholder farmers' bean production and supply to market in Burundi. African Crop Science Journal 19(4), 335–342.

Bradbury, A., Hine, J., Njenga, P., Otto, A., Muhia, G., Willilo, S. (2017): Evaluation of the effect of road condition on the quality of agricultural produce (Report No. RAF2109A). United Kingdom: Cardno Emerging Markets (UK) Ltd.

Carrer, M. J., Silveira, Rodrigo L. F., Filho, H. M. (2019): Factors influencing hedging decision: evidence from Brazilian citrus growers. Australian Journal of Agricultural and Resource Economics 63(1), 1–19, https://doi.org/10.1111/1467-8489.12282.

Christiaensen, L. (2017): Agriculture in Africa – Telling myths from facts: A synthesis. Food Policy, 67, 1–11, https://doi.org/10.1016/j.foodpol.2017.02.002.

CSA. (2011): Federal Democratic Republic of Ethiopia Central statistical Agency Statistical Abstract, Addis Ababa, Ethiopia.

CSA. (2014): Agricultural sample survey: Report on farm management practices for private peasant holdings, 3, Addis Ababa, Ethiopia.

CSA. (2015): Key Findings of the 2014/2015 Agricultural Sample Surveys. The Federal Democratic Republic of Ethiopia Central Statistical Agency, Addis Ababa, Ethiopia.

CSA. (2017): Ethiopia: Demographic and health survey 2016. Central Statistical Agency, Addis Ababa, Ethiopia and ICF International, Maryland, USA.

Dahal, B. M., Sitaula, B. K., Bajracharya, R. M. (2007): Sustainable agricultural intensification for livelihood and food security in Nepal. Asian Journal of Water, Environment and Pollution 5(2), 1–12.

Daoud, J. I. (2017): Multicollinearity and regression analysis. Journal of Physics: Conference Series 949, 1–6, https://doi.org/10.1088/1742-6596/949/1/012009.

Delaney, S., Livingston, G., Schonberger, S. (2017). Right place, right time: Increasing the effectiveness of agricultural development support in sub-Saharan Africa development support in sub-Saharan Africa. South African Journal of International Affairs 18(3), 341–365, https://doi.org/10.1080/10220461.2011.622950.

Diao, X., Silver, J., Takeshima, H. (2016): Agricultural mechanization and agricultural transformation (Background Paper for African Transformation Report 2016: transforming Africa's agriculture). International Food Policy Research Institute.

Didenko, N., Skripnuk, D., Mirolyubova, O., Radion, M. (2017): Analysis of rural areas development of the region using the adl-model. Research for Rural Development 2, https://doi.org/10.22616/rrd.23.2017.061.

Elias, A., Nohmi, M., Yasunobu, K., Ishida, A. (2013): Effect of agricultural extension program on smallholders' farm productivity: Evidence from three peasant associations in the highlands of Ethiopia. African Journal

of Agricultural Research 6(2), 476–487, https://doi.org/10.5539/jas.v5n8p163.
FAO. (2015): The economic lives of smallholder farmers: An analysis based on household data from nine countries, Rome, Italy.
Freebairn, J. (2003): Economic policy for rural and regional Australia. Australian Journal of Agricultural and Resource Economics 47(3), 389–414, https://doi.org/10.1111/1467-8489.00220.
Friday, O. R., Emenonye, C. (2012): The detention and correction of multicollinearity effects in a multiple regression diagnostics. Elixir International Journal 49, 10108–10112.
Fufa, B., Hassan, R. (2006): Determinants of fertilizer use on maize in Eastern Ethiopia: A weighted endogenous sampling analysis of the extent and intensity of adoption. Agrekon 45(1), 38–49, https://doi.org/10.1080/03031853.2006.9523732.
Gebremedhin, B., Hoekstra, D. (2008): Market orientation of smallholders in selected grains in Ethiopia: Implications for enhancing commercial transformation of subsistence agriculture. IPMS (Improving Productivity and Market Success) of Ethiopian Farmers. Project Working Paper No. 11. ILRI.
Gray, P. S., Williamson, J. B., Karp, D. A., Dalphin, J. R. (2007): The Research Imagination: An Introduction to Qualitative and Quantitative Methods. Cambridge: Cambridge University Press, https://doi.org/10.1017/CBO9780511819391.
Hailu, B. K., Abrha, B. K., Weldegiorgis, K. A. (2014): Adoption and impact of agricultural technologies on farm income: evidence from Southern Tigray, Northern Ethiopia. International Journal of Food and Agricultural Economics 2(4), 91–106.
Hailu, C., Fana, C. (2017): Determinants of Input Commercialization as Buyers of Agro-chemicals and improved seed: Evidence from farm households' of Ambo and Toke Kutaye districts, West Shewa Zone, Ethiopia. American Research Journal of Agriculture 3(1), 1–14, https://doi.org/10.21694/2378-9018.17004.
Hayami, Y., Ruttan, V. W. (1985): Agricultural development. Johns Hopkins University Press, Baltimore.
Hayami, Y., Ruttan, V. W. (1993): Induced technical and institutional change evaluation and reassessment: two chapters. Bulletin Number 93-1, University of Minnesota Economic Development Center, Minnesota, USA.
Hellin, J., Lundy, M., Meijer, M. (2009): Farmer organization, collective action and market access in Meso-America. Food Policy, 34(1), 16–22, https://doi.org/10.1016/j.foodpol.2008.10.003.
Hine, J. (2004): Good Policies and Practices on Rural Transport in Africa Planning Infrastructure and Services (SSATP Working Paper No. 100). Washington, D.C., USA: The World Bank.
Hung, P. Van, MacAulay, T. G., Marsh, S. P. (2007): The economics of land fragmentation in the North of Vietnam. Australian Journal of Agricultural and Resource Economics 51, 195–211, https://doi.org/10.1111/j.1467-8489.2007.00378.x.
Imdadullah, M., Aslam, M., Altaf, S. (2016): mctest: An R Package for Detection of Collinearity among Regressors. The R Journal 8(2), 495–505, https://doi.org/10.32614/RJ-2016-062.
Jayne, T. S. Ã., Govereh, J., Wanzala, M., Demeke, M. (2003): Fertilizer market development: a comparative analysis of Ethiopia, Kenya , and Zambia. Food Policy 28, 293–316, https://doi.org/10.1016/j.foodpol.2003.08.004.
Jelilov, G., Kachallah, M. B. (2017): The nexus among road transport and the economic growth in Nigeria. The Journal of Middle East and North Africa Sciences 3(9), 22–29, https://doi.org/10.12816/0040813.
Kamara, A. B. (2004): The impact of market access on input use and agricultural productivity: evidence from Machakos district, Kenya. Agrekon 43(2), 202–218, https://doi.org/10.1080/03031853.2004.9523645.
Kassa, B. (2014): Assessment of factors affecting agricultural production : Evidence from smallholder farmers of Southern Tigray, Northern Ethiopia (master's thesis). Mekelle University, Mekelle, Ethiopia.
Katungi, E., Horna, D., Gebeyehu, S., Sperling, L. (2011): Market access, intensification and productivity of common bean in Ethiopia: A microeconomic analysis. African Journal of Agricultural Research 6(2), 476–487.
Kim, N. (2015): Tests based on skewness and kurtosis for multivariate normality. Communications for Statistical Applications and Methods 22(4), 361–375, https://doi.org/10.5351/CSAM.2015.22.4.361.
Kiplimo, L. B., Ngeno, V. (2016): Understanding the effect of land fragmentation on farm level efficiency: An application of quantile regression-based thick frontier approach to maize production in Kenya. Paper presented at the 5th International Conference of the African Association of Agricultural Economists, September 23–26, 2016, Addis Ababa, Ethiopia.
Kotu, B. H., Alene, A., Manyong, V., Hoeschle-Zeledon, I., Larbi, A. (2017): Adoption and impacts of sustainable intensification practices in Ghana. International Journal of Agricultural Sustainability 15(5), 539–554, https://doi.org/10.1080/14735903.2017.1369619.
Lim-Applegate, H., Rodriguez, G., Olfert, R. (2002): Determinants of non-farm labour participation rates among farmers in Australia. Australian Journal of Agricultural and Resource Economics 46(1), 85–98, https://doi.org/10.1111/1467-8489.00168.
Lulit, A. (2012): Impact of Road on Rural Poverty Evidence Form Fifteen Rural Villages in Ethiopia (master's thesis). Erasmus University Rotterdam, Institute of Social Studies (ISS), The Hague, the Netherlands.
Ma, W., Bicknell, K., Renwick, A. (2019): Feed use intensification and technical efficiency of dairy farms in New Zealand. Australian Journal of Agricultural and Resource Economics 63(1), 20–38, https://doi.org/10.1111/1467-8489.12283.
Margarian, A. (2011): Endogenous Rural Development: Empowerment or Abandonment? Paper presented at the 4th International Summer Conference in Regional Science, Dresden.
Markelova, H., Meinzen-dick, R., Hellin, J., Dohrn, S. (2009): Collective action for smallholder market access. Food Policy 34(1), 1–7, https://doi.org/10.1016/j.foodpol.2008.10.001.
Minten, B., Koru, B., Stifel, D. (2013): The last mile(s) in modern input distribution: Pricing, profitability, and adoption. Agr Econ 44(6), 629–646, https://doi.org/10.1111/agec.12078.
Miot, H. A. (2017): Assessing normality of data in clinical and experimental trials. Vasc Bras 16(2), 88–91, https://doi.org/10.1590/1677-5449.041117.

Nambiro, E. (2008): Trends in land use and agricultural intensification in Kakamega, Western Kenya (Doctoral dissertation, Rheinische Friedrich-Wilhelms University, Nairobi, Kenia). Retrieved from http://hss.ulb.uni-bonn.de/diss.

Negari, K. I. (2017): Compiled Body of Works in Field Epidemiology. Ethiopia Field Epidemiology Training Program(EFETP) Addis Ababa University.

Nin-Pratt, A. (2016): Agricultural Intensification and Fertilizer Use. In: Samuel Benin (Ed.), Agricultural productivity in Africa : trends, patterns, and determinants (pp. 199–246). Washington, DC: International Food Policy Research Institute.

Ninnin, B. (1997): Transport et Developpement A Madagascar. French Co-operation Ministry and Malagasy Public Works Ministry, INRETS.

Olana, B. T. (2006): People and Dam: Environmental and Socio-economic changes induced by reservoir in Fincha Water shades, Western Ethiopia. PhD Dissertation, Wageningen University, Netherland.

Orakwue, C., Umeghalu, I., Ngini, J. (2015): Effects of Road Transport on Agricultural Productivity: A Case Study of Ayamelum Local Government Area of Anambra State, Nigeria. Inter J Appl Sci Engr 3(1), 1–4.

Osborne, J. W., Waters, E. (2002): Four assumptions of multiple regression that researchers should always test. Practical Assessment, Research, and Evaluation 8(2), 1–5.

Perz, S. G. (2003): Social determinants and land use correlates of agricultural technology adoption in a forest frontier: A case study in the Brazilian Amazon. Human Ecology 31(1), 133–165, https://doi.org/10.1023/A:1022838325166.

Pingali, P. L., Rosegrant, M. W. (1995): Agricultural commercialization and diversification: processes and policies. Food Policy 20(3), 171–185, https://doi.org/10.1016/0306-9192(95)00012-4.

Quan, T. T. (2009): Transition from subsistence farming to commercial agriculture in Quang Binh province, Vietnam (doctoral dissertation). Lincoln University, Oxford, USA.

Roland-holst, D. (2009): Infrastructure as a Catalyst for Regional Integration, Growth, and Economic Convergence: Scenario Analysis for Asia. In: F. Zhai (Ed.), From Growth to Convergence: Asia's Next Two Decades (108–149). Palgrave Macmillan UK, https://doi.org/10.1057/9780230250604_4.

Ruttan, V. W. (2008): Induced technical change induced institutional change and mechanism design. Paper presented at the 10th International Workshop on Institutional Economics, Institutions, Technology and their Roles in Economic Growth, University of Hart.

Ruttan, V. W. (1996): Induced Innovation and Path Dependence: A Reassessment with Respect to Agricultural Development and the Environment. Technological Forecasting and Social Change 53, 41–59, https://doi.org/10.1016/0040-1625(96)00055-8.

Ruttan, V. W., Hayami, Y. (1984): Toward a theory of induced institutional innovation. The Journal of Development Studies 20(4), 203–223, https://doi.org/10.1080/00220388408421914.

Salami, A., Kamara, A. B., Brixiova, Z. (2010): Smallholder Agriculture in East Africa: Trends, Constraints and Opportunities. Working Papers Series No. 105, African Development Bank, Tunis, Tunisia.

Tamene, S., Megento, T. L. (2017): The effect of rural road transport infrastructure on smallholder farmers' agricultural productivity in Horro Guduru Wollega zone, Western Ethiopia. AUC Geographica 52(1), 79–89, https://doi.org/10.14712/23361980.2017.7.

Vandercasteelen, J., Tamru, S., Minten, B., Swinnen, J. (2016): Cities and Agricultural Transformation in Africa: Evidence from Ethiopia (Working Paper No. 374/2016). Belgium: LICOS Centre for Institutions and Economic Performance, https://doi.org/10.2139/ssrn.2744504.

Weir, S., Knight, J. (2004): Externality effects of education: Dynamics of the adoption and diffusion of an innovation in rural Ethiopia. University of Chicago, https://doi.org/10.1086/423254.

Wondemu, K. A., Weiss, J. (2012): Rural Roads and Development: Evidence from Ethiopia. EJTIR 12(4), 417–439.

Worku, I. (2011): Road sector development and economic growth in Ethiopia. Ethiopia Support Strategy Program II, International Food Policy Research Institute, Addis Ababa, Ethiopia, 101–146.

Yap, B. W., Sim, C. H. (2011): Comparisons of various types of normality tests. Journal of Statistical Computation and Simulation 81(12), 2141–2155, https://doi.org/10.1080/00949655.2010.520163.

Zewdie, T. D. (2015): Access to Credit and the Impact of Credit constraints on Agricultural Productivity in Ethiopia: Evidence from Selected Zones of Rural Amhara (master's thesis). Addis Ababa University, Ethiopia.Salami, A., Kamara, A. B., Brixiova, Z. (2010): Smallholder Agriculture in East Africa: Trends, Constraints and Opportunities. Working Papers Series No. 105, African Development Bank, Tunis, Tunisia.

Tamene, S., Megento, T. L. (2017): The effect of rural road transport infrastructure on smallholder farmers' agricultural productivity in Horro Guduru Wollega zone, Western Ethiopia. AUC Geographica 52(1), 79–89, https://doi.org/10.14712/23361980.2017.7.

Vandercasteelen, J., Tamru, S., Minten, B., Swinnen, J. (2016): Cities and Agricultural Transformation in Africa: Evidence from Ethiopia (Working Paper No. 374/2016). Belgium: LICOS Centre for Institutions and Economic Performance, https://doi.org/10.2139/ssrn.2744504.

Weir, S., Knight, J. (2004): Externality effects of education: Dynamics of the adoption and diffusion of an innovation in rural Ethiopia. University of Chicago, https://doi.org/10.1086/423254.

Wondemu, K. A., Weiss, J. (2012): Rural Roads and Development: Evidence from Ethiopia. EJTIR 12(4), 417–439.

Worku, I. (2011): Road sector development and economic growth in Ethiopia. Ethiopia Support Strategy Program II, International Food Policy Research Institute, Addis Ababa, Ethiopia, 101–146.

Yap, B. W., Sim, C. H. (2011): Comparisons of various types of normality tests. Journal of Statistical Computation and Simulation 81(12), 2141–2155, https://doi.org/10.1080/00949655.2010.520163.

Zewdie, T. D. (2015): Access to Credit and the Impact of Credit constraints on Agricultural Productivity in Ethiopia: Evidence from Selected Zones of Rural Amhara (master's thesis). Addis Ababa University, Ethiopia.

The fertility revolution in Zimbabwe with special regards to proximate determinants of fertility

Collet Muza*

Charles University, Faculty of Science, Department of Demography and Geodemography, Prague, Czechia
* Corresponding author: colletmuza@gmail.com

ABSTRACT

The role of proximate determinants in moderating fertility decline is well documented in developing countries. In Zimbabwe, however, there is a limited understanding of the role of proximate determinants on fertility levels and trends. This study aimed to examine the role of proximate determinants of fertility (namely marriages, postpartum infecundity and contraception) using the Bongaarts proximate determinants model. The impact of these determinants is studied on a sample of married women aged 15–49 years' and corresponding cross-sectional data obtained through the six consecutive Zimbabwe Demographic Health Surveys (ZDHSs) hold in 1988, 1994, 1999, 2005 and 2015. The results reveal that the overall fertility declined from 5.4, 3.8 and 4.0 children per woman observed among 1988, 1999 and 2015 ZDHSs, respectively. This change was caused by the contraceptive inhibitive effect, which correspondingly increased from 3.00 to 4.65 and 6.45 children per woman. The fertility stalling observed in 1999 and after that is caused by postpartum infecundity and marital fertility inhibition which decreased with time. Moreover, contraceptive inhibition effect increased with education, wealth quintiles, and urban residence. In contrast, marital and postpartum infecundity fertility inhibition effects inversely correlate with education, wealth quintiles, and the place of residence. Therefore, to foster further fertility decline to replacement level, policies should promote contraceptive adoption, more extended breastfeeding periods and delay entry into early marriages. Furthermore, women empowerment, especially the promotion of female education to higher education and female employment, could be useful tools to further fertility decline.

KEYWORDS

proximate determinants; fertility stalling; decomposition; Bongaarts model

Received: 18 April 2019
Accepted: 25 September 2019
Published online: 9 December 2019

Muza, C. (2019): The fertility revolution in Zimbabwe with special regards to proximate determinants of fertility.
AUC Geographica 54(2), 182–193
https://doi.org/10.14712/23361980.2019.16

1. Introduction

A significant number of African countries have experienced either sustained high fertility or fertility stalling at midway transition, which has put future prospects of fertility decline into doubt and made demographers question the demographic transition theory. According to Bongaarts (2008), fertility stalling is when a country does not experience significant fertility decline of 10% or more between two successive surveys demographic and health surveys. Consequently, because of sustained high fertility and low mortality, the population growth rate in Africa has remained high at about 2.5% per annum since 1960 (United Nations 2017). The population doubling time associated with such an annual growth rate is 28 years. Thus, the population of Africa grew from 0.3 to 1.2 billion from 1960 to 2015 (ibid.). The challenge is that high population growth has been shown to negatively affect socio-economic and human development (Birdsall and Sinding 2001). This is alarming since Africa is already the world's least developed region (United Nations 2017).

In 1956, it was propounded that socioeconomic background factors operate through 11 proximate determinants to influence the level of fertility in any society (Davis and Blake 1956). Further, Bongaarts (1978) revised and reduced Davies and Blake's 11 into 7 proximate determinants of fertility. Furthermore, Bongaarts (1982) demonstrated, that 96% of fertility levels and variation in developing countries is a result of four main proximate determinants; marriage, postpartum infecundity, abortion and contraception. The remaining seven variables explain a small part of the total fertility rate variation (ibid.). The challenge is that Africa lacks historical and current complete vital registration systems for conducting such analysis (Dyson 2013). However, the seminal work of Notestein (1945) has established that the fertility revolution from natural to controlled fertility regimes is a result of the deliberate adoption of birth control methods (contraception). Unlike Europe where women of older reproductive ages used stopping reproductive behaviour and natural methods of fertility limitation, in Africa, efficient modern methods of contraception were used for both child spacing and stopping in younger and older reproductive ages respectively (Caldwell and Caldwell 2002). Thus, in Africa, the fertility revolution, although still incomplete, once started, is more rapid than Europe's historical fertility transitions.

In Zimbabwe, fertility had declined from *TFR* 5.4 in 1984 to 3.8 in 1999 has been both development and crisis-driven fertility decline (Mhloyi 1988; Müller et al. 2013). Development-driven fertility decline is typical of the Becker's theory of fertility decline when for instance, highly urbanized, educated and high incomes women were further ahead in the fertility transition than disadvantaged socioeconomic groups (Mhloyi 1988; Müller et al. 2013). In contrast, crisis-driven fertility decline is when a combination of high costs of living, declining incomes, civil unrest and persistence droughts in Zimbabwe has forced couples, regardless of their levels of modernization, to adopt their fertility downwards (Mhloyi 1988). For instance, the costs related to the education of the children rather than the education of mothers themselves have led women to decrease their demand for children. In other words, the high costs of living reverse the wealth flow from children to parents (Caldwell and Caldwell 2002).

However, Gould (2015) has argued that crisis-driven fertility decline cannot be sustained in the long run. Precisely fertility in Zimbabwe has stalled at midway transition at about 4 children per woman since 2000 (ZIMSTAT and ICF International 2016). The question is whether fertility stalled as a result of changes in background variables or changes in proximate determinants?

Although the effects of the proximate determinants on fertility have been documented in Zimbabwe (Guilkey and Jayne 1997; Letamo and Letamo 2001; Mhloyi 1986; Muhwava and Muvandi 1994; Sibanda 1999), little has been done recently to show the current scenario with regard to proximate determinants. Undoubtedly, the socioeconomic circumstances appear to have changed since significantly changed, which might also change the proximate determinants. Thus, this study seeks to assess the proximate determinants of fertility in Zimbabwe, incorporating recent data. The data were pooled from 6 consecutive (1988, 1994, 1999, 2005 2010, and 2015) Zimbabwe Demographic and Health Surveys (ZDHSs). This study will contribute to our knowledge and understanding of proximate determinants of fertility in Zimbabwe and how they have associated fertility levels and trends.

A substantial body of literature has generally agreed that Africa's early and universal nuptiality conditions are more conducive to high fertility rates than witnessed in Europe's historical conditions (Chesnais 1992; Coale 1973; Dyson 2013). Undeniably, entry into marriage serves as a risk factor to childbearing as most of the childbearing was happening in marriages (Bongaarts 1982; Coale 1973). In Zimbabwe, a recent study by Sayi and Sibanda (2018) based on 2014 Zimbabwe Multiple Indicators Cluster Survey (MICS) indicates that 1 in 4 women aged 15–19 years were currently married, while among 20–49 years about 32% were married before 18 years of age. Further, the gap between marriage and childbearing is very small as couples seek to strengthen their marriage with children (Chitereka and Nduna 2010). Marital dissolutions have been found to be insignificant to fertility levels as remarriages occur early (Mhloyi 1988). However, it is possible that nuptiality patterns changed with time and consequently fertility levels.

Polygyny is generally high and variable in Africa. McDonald (1985) reported that polygyny ranged from 10% to 67% in Lesotho and Senegal, respectively. In

Zimbabwe in a recent survey, polygamous unions have been found to be 8% and 16% in urban and rural areas, respectively (ZIMSTAT and ICF International 2016). Studies on the effects of polygamy on fertility offer contradictory evidence. On the one hand, since women in the polygamous union share time of husband, they have less exposure to sex than women in monogamous relationships and correspondingly have less risky to pregnancy. Moreover, women in polygamous unions have been noted to breastfeed longer than women in monogamous unions (Sayi and Sibanda 2018; Mhloyi 1988). On the other hand, Mhloyi (ibid.) argued that women in polygamous unions' demand for children is higher than women in monogamous unions, as more children offer them security, respect and access to husband's wealth. The question is, has the polygyny patterns changed with inevitable modernization?

Furthermore, although variable and decreasing with modernization, Africa's long postpartum abstinences were strategies ensuring the survival of already born children through prolonged birth spacing rather than deliberate birth control methods (ibid.). Caldwell and Caldwell (2002) have argued that child spacing is embedded in the African culture, and high levels of contraceptive use might be for spacing and not stopping childbearing. Nonetheless, it can be hypothesised that postpartum infecundity has decreased with modernization. In the absence of alternative contraceptive adoption or increments, fertility can increase.

Several studies have employed the Bongaarts framework for the analysis of the fertility differentials, levels and transitions in Africa (Chola and Michelo 2016; Majumder and Ram 2015; Rutaremwa et al. 2015; Mturi and Kembo 2011; Sibanda 1999; Bongaarts 1978). A recent analysis of countries in the early stages of fertility transition namely Namibia (Palamuleni 2017); Zambia (Chola and Michelo 2016), Uganda (Rutaremwa et al. 2015), Malawi (Palamuleni 1996), and found the dual importance of marriage and postpartum infecundability as most important predictors of fertility outcomes. The contribution of contraception was least although it was increasing rapidly.

In African countries that have experienced significant fertility decline, it has been established that the contribution of contraception is the most important and amenable to fertility reduction (Finlay, Mejia-Guevara and Akachi 2018; Majunder and Ram 2015; Mturi and Kembo 2011; Sibanda 1999). Sibanda (1999) looking at the relative contribution of proximate determinants in Zimbabwe and Kenya using two consecutive surveys for each country show that in Zimbabwe contraception is the most important factor in fertility decline in younger and middle ages. In Kenya, with slightly higher fertility than Zimbabwe, it was found that postpartum infecundity was the most important factor followed by marriage. In a later study using 2005/06 DHS data on Zimbabwe Mturi and Kembo (2011) reveal that high contraception prevalence rate was the most important factor of fertility decline even during periods of socioeconomic development and socioeconomic crisis. The question is why fertility in Zimbabwe has stalled at midway transition given such high contraceptive prevalence rate? South Africa (*TFR* 2.3) and Botswana (*TFR* 2.6) with similar contraceptive prevalence levels and they have lower fertility approaching replacement level fertility (United Nations 2017). This is important given that studies have revealed contraception prevalence of 75% is associated with replacement level fertility (Mahjabeen and Khan 2011).

It is also possible that contraceptives such as condoms might be used for HIV prevention but not parity-specific fertility limitation behaviour (Terceira et al. 2003). Nonetheless, studies have revealed the contributory effects of the proximate determinants vary positively with women empowerment, i.e. women's education, wealth quintiles, urban and rural residence. In Zambia, secondary education and urban residence and wealth were established to be positively related to higher relative fertility inhibition effect in marriage and contraception whilst postpartum infecundability was inversely correlated (Chola and Michelo 2016). Similar findings were made in Uganda (Rutaremwa et al. 2015) and in Zimbabwe (Mturi and Kembo 2011; Sibanda 1999).

2. Data and methods

2.1 Sources of data

This paper utilizes pooled data from 6 consecutive ZDHSs conducted in 1988, 1994, 1999, 2005, 2010 and 2015. In each survey year, a nationally representative survey of ever-married women of age 15–49 which were conducted under the collaboration of Zimbabwe National Statistics Agency (ZIMSTAT) and ICF International. The ZDHSs provides nationally representative data on basic health social and demographic indicators. The study utilized this data in order to fit the aggregate fertility model, thereby assessing the contribution of marriage, contraception and postpartum infecundability in Zimbabwe. The fertility estimates were also disaggregated by a number of selected socioeconomic variables, namely residence, education and wealth quintiles.

2.2 Sampling

The sample sizes of the interviewed women age (15–49) were selected based on a master sampling plan which was provided by ZIMSTAT (1988–2015). Sampling was done using a two-stage cluster sampling process. Initially, clusters are selected from a list of clusters obtained from the master sampling plan provided by ZIMSTAT, followed by a section of households from each cluster. The data obtained were stratified by rural and urban areas. The samples are considered adequate to enable analysis and comparisons that would be useful in the identification of socioeconomic and demographic locus that could guide fertility and population policy interventions in Zimbabwe.

2.3 Analytical framework

Bongaarts' proximate determinants model (1978, 1982) was applied for analyzing proximate determinants of fertility from the six successive ZDHSs named above. Bongaarts and Potter (1983) developed a technique to quantify the impact of four proximate determinants on fertility, namely marriage, contraception, abortion, and postpartum infecundability. They assume that the total fecundity rate (TF) of all women is the same, but their real reproductive performance is modified by the above four mentioned proximate determinants. The mechanism of the model is summarized by relating the fertility measures to the proximate determinants. The basic model equations are:

(a) $TFR = C_m \times C_i \times C_c \times C_a \times TF$ (Bongaarts 1982),
(b) $TM = C_i \times C_c \times C_a \times TF$ (Chola and Michelo 2016),
(c) $TN = C_i \times TF$ (ibid.),

where *TFR* is the number of births, a woman would have at the end of her reproductive years if she were to bear children at the prevailing age-specific fertility rates throughout the reproductive period. Total marital fertility rate (*TM*) is the number of births a woman would have at the end of her reproductive year if she were to bear children at the prevailing age-specific marital fertility rates and remain married during the entire reproductive period (Bongaarts 1987). Total natural fertility rate (*TN*) is observed under conditions in which contraception and abortions are eliminated (Bongaarts 1982). Bongaarts noted that whilst *TFR*, *TM*, and *TN* vary in many populations, total fecundity is constant in all populations. Total natural fecundity rate (*TF*) index is estimated as follows:

$TF = TFR / (C_m \times C_i \times C_c \times C_a)$ (Bongaarts 1982),

where, C_m, C_i, C_c, and C_a are indices of marriage, postpartum infecundability, and contraception and induced abortion, respectively. The indices can only take values between 0 and 1. Where there is no fertility inhibiting effect of a given intermediate fertility variable, the corresponding index equals one, and when the fertility inhibition is complete the index equal 0 (Bongaarts 1982). Abortion is illegal in Zimbabwe (ZIMSTAT and ICF International 2016; Sibanda 1999), and there is limited and unreliable information. Therefore, the index of abortion in this study will be 1. The contribution of abortion is regarded as insignificant. The indices can be approximated from the measure of proximate variables.

2.4 Index of marriage (C_m)

The index of marriage (C_m) measures the inhibiting effects of marriage on fertility in the population. The lower the proportions of married, the higher the inhibiting effects of marriage and the inverse is true. However, age-specific marital proportions are used since inhibiting effects of marriage are marital age distribution sensitive. Childbearing is greatest in the central age distribution years. Marriage is defined as formal or consensus marriages. Implicit in this definition is the assumption that only women in marriages are exposed to the risk of childbearing. The ZDHSs done in 1999, 2005, 2010 and 2015 has classified cohabitation as a marital union in Zimbabwe. This was not the case in 1988 and 1994 since no information was collected on this variable. In order to make the ZDHSs marital definition comparable, we take the earlier definition used in 1988 and 1994. The index is estimated as follows:

$$C_m = \frac{\sum m(a)g(a)}{\sum g(a)} \quad \text{(Bongaarts 1982)},$$

where $m(a)$ is age-specific proportions currently married women. This characteristic is computed by dividing the number of married women by the total number of women in the same age group. Also, $g(a)$ is age-specific marital fertility rates. It is computed by dividing the births of a particular age group (from married women) by the number of married women in the same age group.

2.5 Index of contraception (C_c)

The index of contraception measure the inhibiting effects of modern methods of contraception on fertility in a population which also varies with the prevalence and use-effectiveness of contraception used by couples in the reproductive age groups. The higher the level of contraception in a population, the higher the inhibiting effect of contraception and vice versa. The index of contraception is estimated using the following:

$C_c = 1 - 1.08 \times u \times e$ (Bongaarts and Potter 1983),

where u is the average proportion of married women currently using contraception; is average contraceptive effectiveness. The coefficient 1.08 is the sterility correction factor (represents the adjustment for the fact that women do not use contraception if they know they are sterile). The parameter e values proposed for particular contraceptives are as follows: pill 0.90, IUD 0.95, injection 0.99, sterilization 1.00, others 0.70 (Bongaarts and Potter 1983).

2.6 Index of postpartum infecundability (C_i)

The index of postpartum infecundability is a measure of the inhibiting effect of breastfeeding or sexual abstinence on fertility in a population (Bongaarts

1982). The index of postpartum infecundability in the model is estimated using the effect of breastfeeding (lactation amenorrhea) or postpartum abstinence. The index was calculated as follows:

$C_1 = 20/18.5 + i$ (Bongaarts 1982),

where the symbol i represents the average duration in months of infecundability from birth to the first postpartum ovulation (menses), in this research, the index of postpartum infecundability was estimated using the mean duration of breastfeeding.

2.7 Index of abortion (C_a)

The index of abortion measures the inhibiting effect of abortion on fertility in a population. In this research, the index of abortion was set 1.0 due to lack of data. Abortion is illegal in Zimbabwe excerpt for health and legal reasons (ZIMSTAT and ICF International 2016). Moreover, abortion data in the ZDHSs include stillbirths and miscarriages. Therefore, it is difficult to isolate abortion data. The index of abortion is estimated using the following formula:

$$C_a = TFR/TFR + b \times TA = TFR/TFR + 0.4 \times (1 + U) \times TA$$
(Bongaarts and Potter 1983)

where U is contraceptive prevalence use, b is the average number of births averted per induced abortion and $b = 0.4(1 + U)$; b is 0.4 when $U = 0$ and $b = 0.8$ when $U = 1.0$. TA is total abortion (an average number of induced abortion per woman at the end of the reproductive period if induced abortions remain at prevailing levels throughout the reproductive period). Then $C_a = 1.0$ if $TA = 0$. Therefore, the index of abortion in this study is equal to .

Based on the studies of historical populations with the highest recorded fertility, Bongaarts recommends

Tab. 1 Percentage distribution of respondent women, 15–49 years by selected characteristics (weighted), 1988–2015, Zimbabwe.

		1988	1994	1999	2005	2010	2015
Children born	1–3	34.9%	38.1%	44.7%	47.6%	52.3%	50.2%
	4–5	16.7%	15.0%	12.9%	14.0%	14.6%	17.0%
	6+	19.9%	17.5%	12.6%	9.0%	6.7%	6.1%
Marital status	Never married	27.0%	26.9%	27.7%	27.0%	24.0%	25.2%
	Married	62.9%	61.8%	56.3%	56.3%	59.4%	58.7%
	Divorced	7.6%	7.8%	3.5%	4.5%	3.7%	5.0%
	Widowed	2.5%	3.5%	4.2%	7.5%	6.1%	4.4%
Age group	15–19	24.3%	24.0%	24.5%	24.2%	21.2%	22.1%
	20–24	20.0%	20.7%	21.9%	21.9%	20.1%	17.0%
	25–29	16.2%	14.9%	17.5%	16.5%	18.4%	16.6%
	30–34	14.0%	14.2%	11.3%	13.7%	14.1%	16.3%
	35–39	11.0%	10.8%	10.8%	9.4%	11.5%	12.4%
	40–44	7.6%	8.7%	7.9%	7.8%	8.0%	9.7%
	45–49	6.9%	6.6%	6.1%	6.6%	6.8%	5.8%
Residence	Urban	33.5%	32.2%	38.6%	39.3%	38.7%	38.5%
	Rural	66.5%	67.8%	61.4%	60.7%	61.3%	61.5%
Education	No education	13.5%	11.1%	6.7%	4.3%	2.3%	1.35
	Primary	55.9%	47.3%	40.2%	32.6%	28.0%	25.8%
	Secondary	29.7%	40.0%	50.2%	60.15	65.1%	65.6%
	Higher	0.9%	1.6%	2.8%	3.0%	4.6%	7.3%
Wealth quintile	Poorest	–	18.2%	17.1%	17.4%	16.9%	17.1%
	Poorer	–	16.1%	17.3%	16.8%	17.4%	17.0%
	Middle	–	18.9%	18.4%	17.4%	18.3%	17.6%
	Richer	–	21.8%	22.3%	22.5%	22.6%	23.2%
	Richest		25.0%	24.9%	25.9%	24.8%	25.1%
Family Planning (married)	Modern (any)	36.1%	42.2%	50.4%	58.4%	57.3%	65.8%
	Traditional	7.0%	4.3%	2.8%	1.4%	1.1%	1.0%
Mean duration of breastfeeding	Months	18.9	18.8	19.0	18.7	17.5	18.1
Total (n)		4,201	6,128	5,907	8,907	9,171	9,955

– missing values
Source: ZDHSs

using 15.3 as the maximum number of births per woman. This is referred to as the Total fecundity rate (*TF*) (Bongaarts 1978; 1982). The value is the theoretical number that a woman would have if she were to continuously married from age 15–44, did not use contraceptives and did not abort any pregnancies. Multiplying all the indices together by the total fecundity rate of 15.3 produces the predicted *TFR* for the population. The predicted *TFR* will normally differ from the observed *TFR* because of underreporting of births; misreporting of behaviours measured by the indices or omission of proximate factors that help determine fertility levels in the population under study. The Bongaart's model was also applied for calculation of the proximate determinants for selected background characteristics as the place of residence, education and wealth quintiles.

3. Results

The individual characteristics of the respondents for six successive ZDHSs conducted between 1988-2015 are shown in table 1. The results show that the percentages living in rural areas decreased from 67% in 1988 to 61% by 2015. The results reveal that women with no education and primary education were reduced from 14% to 1% and 56% to 26% while women with secondary education simultaneously increased from 30% to 66% between 1988 and 2015 respectively. The percentage of women with higher education is still very low, although it increased from 1% to 7% from 1988 to 2015 respectively. The distribution of the respondents by age groups clearly demonstrate that percentage distribution diminished with an increase in age for the respective periods. The 1988 ZDHS did not collect data on wealth quintiles. Nonetheless, for the later ZDHSs periods with available data, the wealth quintiles stayed relatively the same with respondents in the lowest wealth quintile slightly decreasing from 18% to 17% and a simultaneous increase in the fourth wealth quintile from 22 to 23% in 1988 and 2015 respectively.

Table 1 also shows the distribution of the study by selected key proximate determinants of fertility. The percentage currently married and single declined from 63%, 56%, 59% and 27%, 24%, 25% in 1988, 1999 and 2015 respectively. The currently married percentages were considered as proportion currently married in computing the indices for marriage. Mean duration of breastfeeding from ZDHSs reports was, on average 18.5 months from 1988 to 2015, respectively, and this is what was used in the calculations to represent the average duration of the postpartum infecundability. With regard to contraception, there has been an increase in modern contraception use by married women from 36% in 1988 to 66% by 2015 and a simultaneous decrease in traditional methods from 7% to 1% from 1988 to 2015 respectively. The above-mentioned information was used in the calculation of fecundity. As highlighted in the equation above, the observed fertility levels in a population are a product of the relationship of proximate determinants and the maximum biological level of fertility.

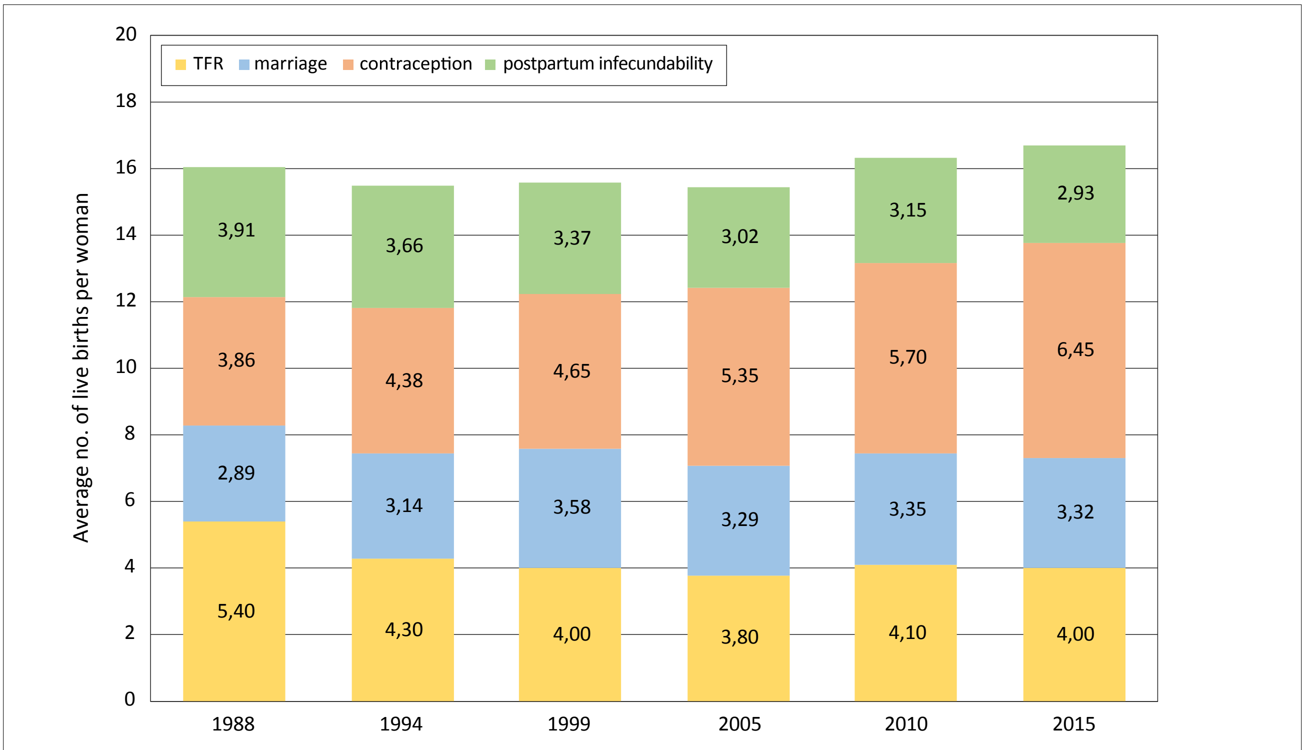

Fig. 1 Effects of proximate determinants on the total fertility rate.

Tab. 2 Estimated proximate determinants and their effect on fertility reduction, 1988–2015, Zimbabwe.

Proximate determinant index	1988	1994	1999	2005	2010	2015
	Proximate determinants indices					
Index of marriage (C_m)	0.72	0.69	0.61	0.62	0.65	0.63
Index of contraception (C_c)	0.65	0.60	0.53	0.46	0.47	0.41
Index of postpartum infecundability (C_i)	0.65	0.65	0.63	0.65	0.66	0.66
Index of abortion (C_a)	1.00	1.00	1.00	1.00	1.00	1.00
	Impact on fertility reduction percentage					
Index of marriage (C_m)	18.9%	20.6%	23.4%	21.5%	21.9%	21.7%
Index of contraception (C_c)	25.2%	28.6%	30.4%	35.0%	37.3%	42.2%
Index of postpartum infecundability (C_i)	25.5%	23.9%	22.0%	19.7%	20.6%	19.2%
Index of abortion (C_a)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
TFR predicted	4.64	4.12	3.14	2.82	3.10	2.60
TFR observed	5.40	4.30	4.00	3.80	4.10	4.00

Source: ZDHSs and own calculations

The fecundity rate has been observed to vary from 13–17 per woman, with an average of 15.3 births per woman.

Figure 1, reveals the effect of each proximate determinant in absolute terms towards fertility reduction from the biological maximum of 15.3 children per woman, for 6 ZDHSs from 1988 to 2015. In 1988, the biggest fertility inhibition (*TFR*) was caused by postpartum infecundity (C_i) 3.91, followed by contraception (C_c) 3.86 and marriage (C_m) 2.89 children per woman. The contraception fertility inhibition effect continuously increased to 6.45 children per woman by 2015, whilst marriage fertility inhibition increased to a peak of 3.58 in 1999 before decreasing to 3.32 children per woman in 2015. The postpartum infecundability fertility inhibition effect has decreased to 2.93 children per woman by 2015. *TFR* has also decreased from 5.40 in 1988 to 4.10 by 2015. Surprisingly, the *TFR* decline reversed from 3.80 in 2005 to 4.10 in 2010.

In terms of percentages, the impact of each proximate determinant is displayed in Table 2. In 1988 contraceptive and postpartum infecundity accounted for 26% and 25% of fertility reduction, with marriage accounting the least of 19% respectively. Similarly as in levels and trend in absolute *TFR* inhibition effect, in 2015 the fertility reduction effects of marriage and contraception significantly increased to 42% and 22% respectively, whilst the effect of postpartum infecundability decreased to 20% from 26% for the same period. The values in Table 1 also show that the effect of marriage and postpartum infecundability stagnated/reduced in 1994 and 1999, respectively, whilst the effect of contraception significantly increased continuously. This suggests that married women who used contraception contributed the most towards fertility reduction.

Values in Table 3 shows the effects of the proximate determinants on fertility by residence, education and wealth. As noted above, the closer the index to zero or one, the bigger or, the lesser the effect on fertility reduction, respectively. Table 3; reveal contraception fertility inhibition effect increased with time and is positively related to the socioeconomic variables. For example, between 1994 and 2015 contraception inhibition effect of higher educated women increased from 60% to 68% respectively, whilst least educated women contraception fertility inhibition effect increased from had 25% to 42% in 1994 and 2015 respectively. The wealth quintile data collection starts from 1994 and shows that the higher the wealth quintile, the stronger the effect of contraception, for example in 1994 poorest and richest quintiles had 30% and 53%, which increased to 55% and 65% respectively by 2015. Urban areas had higher contraception inhibition effect than rural areas which increased from 47% and 30% to 64% and 56% between 1988 and 2015 respectively. Table 3; shows contraception fertility inhibition stagnation or reversals for period 2010 for most socioeconomic variables. Surprisingly, married women in rural areas, of middle and poorest wealth quintiles did not experience contraception fertility inhibition stagnation for periods under analysis. In 2015, married women with no education were the only socioeconomic groups with fertility contraception inhibition effect below 50% with 42%.

The index of postpartum infecundability was observed to be inversely related to the selected socioeconomic characteristics. The postpartum infecundity, although fluctuating, has slightly increased from 1988 to 2015. For example, in 1988 to 2015 poorest and richest women had 37% and 32% postpartum infecundity, which decreased to 35% to 30% respectively. This suggests that the fertility inhibition effect of post-partum infecundity is higher at lower socioeconomic levels than high socioeconomic levels and have been slightly reduced across all socioeconomic groups over time.

The marriage inhibition index effect on fertility is inversely related to education, urban and wealth

Tab. 3 Estimated indices for the four principle determinants by selected background variables, 1988–2015, Zimbabwe.

Index of contraception (Cc)		1988	1994	1999	2005	2010	2015
Residence	Urban	0.53	0.49	0.42	0.37	0.45	0.36
	Rural	0.70	0.64	0.60	0.51	0.49	0.44
Education	No education	0.76	0.75	0.68	0.72	0.62	0.58
	Primary	0.67	0.63	0.59	0.52	0.52	0.46
	Secondary	0.50	0.48	0.45	0.40	0.46	0.39
	Higher	–	0.40	0.38	0.30	0.40	0.32
Wealth Quintile	Poorest	–	0.70	0.62	0.58	0.52	0.45
	Poorer	–	0.69	0.61	0.49	0.51	0.45
	Middle	–	0.64	0.60	0.49	0.48	0.44
	Richer	–	0.53	0.50	0.39	0.45	0.38
	Richest	–	0.47	0.37	0.35	0.42	0.35
Index of postpartum infecundability (Ci)							
Residence	Urban	0.68	0.67	0.64	0.68	0.68	0.69
	Rural	0.63	0.64	0.62	0.63	0.65	0.65
Education	No education	0.61	0.62	0.62	0.62	0.63	0.65
	Primary	0.64	0.64	0.61	0.63	0.66	0.65
	Secondary	0.69	0.66	0.65	0.65	0.66	0.67
	Higher	0.73	0.85	0.73	0.77	0.70	0.72
Wealth Quintile	Poorest	–	0.63	0.61	0.63	0.65	0.65
	Poorer	–	0.63	0.61	0.64	0.65	0.65
	Middle	–	0.64	0.65	0.63	0.65	0.66
	Richer	–	0.65	0.64	0.65	0.68	0.68
	Richest	–	0.68	0.64	0.71	0.70	0.70
Index of marriage (Cm)							
Residence	Urban	0.71	0.69	0.60	0.62	0.65	0.65
	Rural	0.72	0.69	0.62	0.62	0.63	0.62
Education	No education	0.69	0.67	0.62	0.61	0.58	0.69
	Primary	0.70	0.66	0.59	0.60	0.61	0.60
	Secondary	0.71	0.69	0.62	0.62	0.64	0.64
	Higher	0.82	0.77	0.65	0.64	0.73	0.72
Wealth Quintile	Poorest	–	0.69	0.61	0.62	0.62	0.62
	Poorer	–	0.68	0.62	0.62	0.64	0.62
	Middle	–	0.70	0.63	0.61	0.62	0.63
	Richer	–	0.67	0.60	0.60	0.63	0.62
	Richest	–	0.70	0.61	0.64	0.67	0.67

– missing values
Source: ZDHSs and own calculations

quintiles. However, the marriage inhibition effect across all the selected socioeconomic variables slightly decreased with time (1988-2015). Marital fertility inhibition effect trend is fluctuating between rural and urban areas between 1988 and 2005 rural. However, in 2005 and 2015, slightly higher marital fertility inhibition effect was observed in rural areas (37% and 38%, respectively) than in urban areas (35% and 35%, respectively). Higher educated women had lower marital fertility inhibition than less educated women. Further, there is a general fluctuating but increase in marital fertility inhibition across all educational categories with the least decrease experienced by women with no educated. This suggests education increases are related to marital fertility reduction over time.

4. Discussion

The results from the Bongaarts model presented above reveal levels, trends and the impact of proximate determinants of fertility, namely marriage, postpartum infecundability, and contraception for the six consecutive ZDHSs. The results quantify the impact of each proximate determinant in absolute as well as percentage terms to total fecundity and fertility decline in Zimbabwe. The results uncovered that contraception was the most important factor contributing to fertility decline experienced in Zimbabwe by married women from *TFR* of 5.4 to 4.0 in 1988 to 2015 respectively. Precisely, the contraception inhibition effect increased significantly and continuously from 25% to 42%, whilst marital inhibition effect although fluctuating increased from 19% to 22% and postpartum infecundity inhibition effect decreased from 26% to 19% in 1988 to 2015 respectively. A similar trend is observed of fertility (*TFR*) suppression, as presented in Figure 1. This finding concurs with previous studies that found that contraception adoption is the most important factor for fertility decline in Zimbabwe (Mturi and Kembo 2011; Sibanda 1999). Moreover, this finding is consistent with other previous studies in sub-Saharan Africa (Finlay, Mejia-Guevara and Akachi 2018) and Asia (Majumder and Ram 2015). More broadly, the results confirm the hypothesis that contraception adoption is associated with further fertility decline as countries move from non-parity specific fertility to parity-specific fertility regime (Coale 1973).

Further, the decomposition of contraception fertility inhibition by residence, education and wealth quintiles demonstrate a positive relationship. Analysis of change associated with contraception use helps us examine in some limited way whether any increase in the fertility-inhibiting effects of contraception constitute evidence of what Coale (1973) postulated as the necessary conditions for a fertility decline to start: (1) fertility is within the calculus of the conscious choice, (2) lower fertility is seen as advantageous, (3) effective means of birth control are readily available and couples are willing to use them. The idea of using contraception (and socioeconomic variables) to achieve a smaller ideal family size is grounded in van de Walle's concept of numeracy (van de Walle 1992; Sibanda 1999). This suggests that contraception adoption is single proximate determinant responsible for marital fertility decline in Zimbabwe.

In contrast to findings in this study, Palamuleni (2017) and Chola and Michelo (2016) found out that marriage not contraceptive was the most important proximate determinant for fertility inhibition in Namibia and Zambia, respectively. The differences between Zimbabwe and Namibia might be explained by the fact that in Namibia, percentage currently married for women aged 15–49 years was about 25%, whilst in Zimbabwe, it was about 60% for the respective periods. The discrepancy between Zimbabwe and Zambia which have similar marital rates can plausibly be explained by the fact that Zambia is in the early stages of fertility transition (*TFR* 6.2 in 2007) where fertility is more amenable to Malthusian preventative checks (Malthus 1798), while Zimbabwe is in the midway of fertility transition (*TFR* 4.0 in 2005, 2010 2015) where fertility is more susceptible to contraception adoption. This is made possible by increases in the prevalence rate of modern contraceptive use from 36% in 1988 and 66% in 2015 and was responsible for fertility decline (ZIMSTAT and ICF International 2016).

TFR at the national level fell from1988 (5.4) to 2005 (3.8) (Figure 1). Thus the fall in *TFR* is well above 10%, a figure that is cited in the historical fertility transitions as a key threshold for entry into non-reversible decline (Coale 1973). However, the fertility levels between 2005, 2010 and 2015 increased from *TFR* 3.8, 4.1 and to 4.0 respectively. This means fertility decline in Zimbabwe stalled, as there was less than 10% sustained *TFR* decline between successive surveys as defined by Bongaarts (2008). The fertility stalling finding in Zimbabwe is in line with other findings in a substantial number of sub-Saharan African countries (Bongaarts 2006; Garenne 2013; Goujon, Lutz and Samir 2015; Lutz, Goujon and Kabat 2015). However, during the respective stalling periods, the overall contraceptive fertility inhibition effects continuously strengthened from 5.35, 5.70 and 6.45 children per woman, respectively. Viewed from this empirical evidence, this suggests that fertility stalling in Zimbabwe has not been caused by a lack of priority assigned to family planning services suggested by (Bongaarts 2008). However, it is plausible that modern contraceptive methods such as condoms were used for HIV prevention rather than for fertility limitation, given high HIV prevalence rate of 17% among women (15–49 years) in Zimbabwe (ZIMSTAT and ICF International 2016).

Results indicated that marital fertility inhibition effects of marriage were the least important of all the proximate determinants in 1988 and 1994 and thereafter became the second important after contraception. Equally important, marital fertility inhibition effects increased from 2.89 to 3.58 children per woman in 1988 and 1999, before declining to 3.29 and 3.32 children per woman in 2005 and 2015, respectively. It is possible that the erosion of the marital fertility inhibition effect might be partly responsible for fertility stalling during 2005 and 2015. This finding is consistent with other studies that have found the erosion of marital fertility inhibition effect as countries move along the fertility transition (Chola and Michelo 2016; Mturi and Kembo 2011; Muhwava and Muvandi 1994; Sibanda 1999). Ministry of Health and Child Care (2016), research show an increase in early teenage pregnancies in Zimbabwe. This is important, considering the gap between marriages and childbearing is very small in Zimbabwe (Chitereka and Nduna 2010; Sayi and Sibanda 2018). Thus, fertility stalling might be due to sustained cultural and traditional practices which promote early marriage and childbearing.

The results show that the index of postpartum infecundability fertility-reducing effect has since decreased from the biggest (26%) in 1988 to the least fertility inhibition effect by 1999 (22%) and continuously thereafter. This is consistent with previous findings in Zimbabwe (Sibanda 1999; Kembo and Mturi 2011) and other countries in the region: Namibia (Palamuleni 2017), in Zambia (Chola and Michelo 2016), Malawi (Palamuleni 1996). This means short breastfeeding possibly contributes to fertility stalling noted above. In India, it was found that the risk of pregnancy increases substantially after giving birth in the absence of breastfeeding (Singh, Suchindran and Singh 1993; Bongaarts and Potter 1983). Besides that, prolonged breastfeeding practices and postpartum abstinence have been used effectively in Africa, not only for birth spacing and subsequently reducing total fertility but also for increased child survival (Guikey and Jayne 1997; Mhloyi 1988). ZIMSTAT (2015) notes higher infant mortality among employed than unemployed women in Zimbabwe, which is plausibly related to inadequate and short breastfeeding by employed women. Child survival is considered a principal component of fertility decline (Coale 1973; Dyson 2013). This suggests that the promotion of universal breastfeeding among women can contribute to significant fertility decline through postpartum amenorrhea and increased child survival.

The analysis of education and wealth quintiles shows an inverse relationship with marriage and postpartum infecundability. Rural women have generally had higher postpartum infecundity and marriage fertility reduction effect than urban areas. The fertility inhibition effect of the selected socioeconomic variables generally increased, although they were fluctuating. This corresponds with what emerged from other researches in Zambia (Chola and Michelo 2016), Uganda (Rutaremwa et al. 2015) and Ethiopia (Alazbih, Tewabe and Demisse 2017) and sub-Sahara Africa (Finlay, Mejia-Guevara and Akachi 2018). This could be attributed to the effects that marriage and wealth have on fertility.

The effect of contraception was positively correlated with regard to education, wealth and residence. Furthermore, contraceptive fertility reduction effect increased with time across all socioeconomic groups (rich or poor, educated or not, urban or rural). These results are consistent with other findings that show that contraceptive inhibition effect of fertility from biological maximum is greater among high socioeconomic groups than lower socioeconomic groups; in Uganda (Rutaremwa et al. 2015), Namibia (Palamuleni 2017), Ethiopia (Alazbih, Tewabe and Demisse 2017) and Zambia (Chola and Michelo 2016). This could imply that women empowerment (access to education especially tertiary, female employment, poverty alleviation) could be powerful tools for fertility decline.

An interesting finding is a decrease in the effect of contraception fertility inhibition across all socioeconomic groups in 2010. Although such a finding was not revealed when looking contraceptive inhibition effect without background variables. Nonetheless, the decomposition of contraception effect by socioeconomic variables could partly explain fertility stalling noted above. This finding is consistent with findings from ZIMSTAT Census (2012) Nuptiality and Thematic report which found that fertility stalling could have been caused by fertility postponement in earlier during periods of economic hardships and the fertility rebounded when the economy stabilized in 2009 (ZIMSTAT 2015).

Moreover, the contraceptive fertility inhibition gap between and within each socioeconomic group is narrowing. However, women with no education are lagging behind, e.g. in 2015, they had contraception inhibition effect of 42% when other socioeconomic variables are below 64%. This is in line with the van de Walle's (1992) concept of numeracy discussed above. Similar studies also highlight the importance of mass education in fertility decline in developing countries (Goujon, Lutz and Kabat 2015). The narrowing of the contraceptive fertility inhibition gap between socioeconomic variables might suggest better future prospects for fertility decline. The strengthening contraception fertility inhibiting effect among women of low socioeconomic classes means that it is possible for fertility to decline in the low socioeconomic development environment. This is consistent with findings in other developing countries in Latin America (Bongaarts 2014) and also a phenomenon witnessed in the French fertility revolution (Pavlík and Hampl 1975). These results suggest empowerment of women might be necessary through female education and access to resources.

The study has a number of strengths. The study used a large, nationally representative population-based sample conducted over 30 years. The data has a high response rate above 90%, standardized surveys as enabling comparisons across countries and periods. The decomposition of proximate determinants was done in order to understand sources of fertility changes over time.

The study also had a weakness. The index of abortion was taken as one throughout the analysis, i.e. induced abortion was assumed to have no significant fertility inhibition effect on fertility due to lack of data. It is possible that excluding abortion in the model could have affected the estimation of total fecundity. Furthermore, studies have shown that abortion rates in sub-Saharan Africa are growing rapidly (Alazbih, Tewabe and Demisse 2017; Remez, Woog and Mhloyi 2014). Another limitation of the data used is that it failed to include reproductive data of males. Moreover, the surveys collected retrospective data on women's birth history. It is possible that such respondents might suffer from recall bias which might affect the accuracy and validity of the data. Thus predicted *TFR* will normally differ from the observed *TFR* because of underreporting of births; misreporting of behaviours measured by the indices or omission of proximate factors that help determine fertility levels in the population under study.

5. Conclusion

In this paper, we applied the Bongaarts model to assess and compare changes in the relative importance of proximate determinants of fertility for six ZDHSs conducted in Zimbabwe. The results indicate that the fertility inhibition effects of contraception are the most important than the inhibition effects of marriage, and postpartum infecundability in Zimbabwe. Moreover, the results show that the contraceptive patterns vary positively with education, wealth quintiles and areas of residence. There was a gradual erosion of post-partum infecundability through the analysis period and of marriage after 1999. It is plausible that the fertility stalling could have been caused by the cumulative effects of marriage and postpartum infecundability. The findings of this study have important policy implications. The strengthening of fertility inhibition effects of contraception, late marriage, and prolonged breastfeeding must be promoted. There is a need to research further on proximate determinants of fertility according to age groups as such research can illuminate the age-specific reproductive contributions of fertility.

References

Alazbih, N. M., Tewabe, G. N., Demissie, T. D. (2017): Contraception and fertility transition in AMHARA national, regional state of Ethiopia: An application of Bongaarts' model. Fertility Research and Practice 3(1), 12, https://doi.org/10.1186/s40738-017-0039-8.

Birdsall, N., Sinding, S. W. (2001): How and why population matters: New findings new issues. Oxford University Press, https://doi.org/10.1093/0199244073.003.0001.

Bongaarts, J. (1978): A framework for analyzing the proximate determinants of fertility. Population and Development Review 4(1), 105–132, https://doi.org/10.2307/1972149.

Bongaarts, J. (1982): The fertility-inhibiting effects of the intermediate fertility variables. Studies in Family Planning 13(6–7), 179–189, https://doi.org/10.2307/1965445.

Bongaarts, J. (2006): The causes of stalling fertility transitions. Studies in Family Planning 37(1), 1–16, https://doi.org/10.1111/j.1728-4465.2006.00079.x.

Bongaarts, J. (2008): Fertility transitions in developing countries: Progress or stagnation? Studies in family planning 39(2), 105–110, https://doi.org/10.1111/j.1728-4465.2008.00157.x.

Bongaarts, J. (2014): The impact of family planning programs on unmet need and demand for contraception. Studies in Family Planning 45(2), 247–262, https://doi.org/10.1111/j.1728-4465.2014.00387.x.

Bongaarts, J., Potter, R. E. (1983): Fertility, biology, and behavior: An analysis of the proximate determinants. Academic Press.

Caldwell, J. C., Caldwell, P. (2002): The fertility transition in sub-Saharan Africa. Paper presented at the Conference on Fertility and the Current South African Issues of Poverty, HIV/AIDS and Youth, Pretoria, South Africa.

Chesnais, J. (1992): The demographic transition: Stages, patterns, and economic implications. Clarendon Press.

Chitereka, J., Nduna, B. (2010): Determinants of unmet need for family planning in Zimbabwe. Harare: Zimbabwe National Family Planning Council and Liverpool School of Tropical Medicine.

Chola, M., Michelo, C. (2016): Proximate determinants of fertility in Zambia: Analysis of the 2007 Zambia Demographic and Health Survey. International Journal of Population Research 2016, https://doi.org/10.1155/2016/5236351.

Coale, A. J. (1973): The demographic transition reconsidered. Paper presented at the Congres De l'IUSSP, 1, 53–72.

Davis, K., Blake, J. (1956): Social structure and fertility: An analytic framework. Economic Development and Cultural Change 4(3), 211–235, https://doi.org/10.1086/449714.

Dyson, T. (2013): Population and development: The demographic transition Zed Books Ltd.

Finlay, J. E., Mejía-Guevara, I., Akachi, Y. (2018): Inequality in total fertility rates and the proximate determinants of fertility in 21 sub-Saharan African countries. PloS One 13(9), e0203344, https://doi.org/10.1371/journal.pone.0203344.

Garenne, M. (2013): Situations of fertility stall in sub-Saharan Africa. African Population Studies 23(2), https://doi.org/10.11564/23-2-319.

Goujon, A., Lutz, W., Samir, K. (2015): Education stalls and subsequent stalls in African fertility: A descriptive overview. Demographic Research 33, 1281–1296, https://doi.org/10.4054/demres.2015.33.47.

Gould, W. T. (2015): Population and development. Routledge, New York, USA.

Guilkey, D. K., Jayne, S. (1997): Fertility transition in Zimbabwe: Determinants of contraceptive use and method choice. Population Studies 51(2), 173–189, https://doi.org/10.1080/0032472031000149896.

Letamo, G., Letamo, H. N. (2001): The role of proximate determinants in fertility transition: A comparative study of Botswana, Zambia and Zimbabwe. Southern African Journal of Demography 8(1), 29–35.

Lutz, W., Goujon, A., KC, S., Kabat, P. (2015): The link between structural adjustment programs, education discontinuities and stalled fertility in Africa. IIASA Interim Report. IIASA, Laxenburg, Austria: IR-15-007.

Mahjabeen, T., Khan, I. (2011): Analyzing Bongaarts' model and its applications in the context of Bangladesh. Paper presented at the Proceedings of the 19th International Congress on Modelling and Simulation, 3052–3058.

Majumder, N., Ram, F. (2015): Explaining the role of proximate determinants on fertility decline among poor and non-poor in Asian countries. PloS One 10(2), e0115441, https://doi.org/10.1371/journal.pone.0115441.

Malthus, T. R. (1798): An Essay on the Principle of Population, ed. A. Flew. Aylesbury: Penguin Books.

Mhloyi, M. M. (1986): Fertility determinants and differentials: The cases of Kenya and Lesotho. Zambezia 13(2), 81–107.

Mhloyi, M. M. (1988): The determinants of fertility in Africa under modernization. In: African Population Conference/Congres Africain de Population, Dakar,

Senegal, November/Novembre 7–12, 1988. Liege, Belgium, International Union for the Scientific Study of Population (IUSSP), 1988, 2.3.1-22.
Ministry of Health and Child Care (2016): Zimbabwe National Adolescent Fertility Study, Harare: MoHCC Technical Report authored by Naomi N. Wekwete, Simbarashe Rusakaniko and George Zimbizi (Consultants) Harare Zimbabwe.
Mturi, A., Kembo, J. (2011): Falling fertility and increase in use of contraception in Zimbabwe. African Journal of Reproductive Health 15(2), 31–44.
Muhwava, W., Muvandi, I. (1994): Breastfeeding, contraceptive use, and fertility in Zimbabwe: A further analysis of the demographic and health survey. Paper presented at the Macro International (Ed.), Fertility Trends and Determinants in Six African Countries, DHS Regional Analysis Workshop for Anglophone Africa, Macro International, Calverton, Maryland, 181–215.
Müller, H., Muller, H., Mekgwe, P., Mhloyi, M. M. (2013): Values and development in Southern Africa African Books Collective. Codesria, https://doi.org/10.2307/j.ctvk8vz4z.
Notestein, F. W., (1945): Population – The long view. In: Food for the world. T. Schulz (Ed.), Chicago: University of Chicago Press.
Palamuleni, M. (1996): Proximate determinants and fertility in Malawi. Tanzanian Journal of Population Studies and Development 3(1–2), 65–71.
Palamuleni, M. (2017): Determinants of fertility decline in Namibia: An analysis of the proximate determinants. Bangladesh E-Journal of Sociology 14(2), https://repository.nwu.ac.za/handle/10394/27757.
Pavlík, Z., Hampl, M. (1975): Differentiation of demographic systems according to development and rank with special regard to the third world ECPS. European Demographic Information Bulletin 11(4), 130–139.
Remez, L., Woog, V., Mhloyi, M. (2014): Sexual and reproductive health needs of adolescents in Zimbabwe. Issues in Brief (Alan Guttmacher Institute) 3(3), 1–8.
Rutaremwa, G., Kabagenyi, A., Wandera, S. O., Jhamba, T., Akiror, E., Nviiri, H. L. (2015): Predictors of modern contraceptive use during the postpartum period among women in Uganda: A population-based cross-sectional study. BMC Public Health 15(1), 262, https://doi.org/10.1186/s12889-015-1611-y
Sayi, T. S., Sibanda, A. (2018): Correlates of child marriage in Zimbabwe. Journal of Family Issues 39(8), 2366–2388, https://doi.org/10.1177/0192513X18755198.
Singh, K. K., Suchindran, C. M., Singh, K. (1993): Effects of breastfeeding after resumption of menstruation on waiting time to next conception. Human Biology 65(1), 71–86.
Sibanda, A. (1999): Reproductive change in Zimbabwe and Kenya: The role of the proximate determinants in recent fertility trends. Social Biology 46(1–2), 82–99, https://doi.org/10.1080/19485565.1999.9988989.
Terceira, N., Gregson, S., Zaba, B., Mason, P. (2003). The contribution of HIV to fertility decline in rural Zimbabwe, 1985–2000. Population Studies 57(2), 149–164, https://doi.org/10.1080/0032472032000097074.
United Nations, Department of Economic and Social Affairs, Population Division (2017). World Population Prospects: The 2017 Revision, DVD Edition.
Van de Walle, E. (1992): Fertility transition, conscious choice, and numeracy. Demography 29(4), 487–502, https://doi.org/10.2307/2061848.
ZIMSTAT. (2015): Zimbabwe Statistical Agency (ZIMSTAT) Census 2012 nuptiality and fertility thematic report. Harare, Zimbabwe.
ZIMSTAT and ICF International. (2016): Zimbabwe demographic and health survey 2015: Final report. (2015 Ed.). Rockville, Maryland, USA: Zimbabwe National Statistics Agency (ZIMSTAT) and ICF International. Retrieved from https://dhsprogram.com/pubs/pdf/FR322/FR322.pdf.

Investigating the geotouristic risks of spa springs with emphasis on natural and human dangers in the study area between Sarein and Ardebil in the northwest of Iran

Mehdi Feyzolahpour*, Jamshid Einali, Hassan Gasemlu

Department of Geography, University of Zanjan, Iran
* Coresponding author: feyzolahpour@znu.ac.ir

ABSTRACT
Geotourism and geoparks provide good opportunities for rural development and reduce unemployment and migration. It attracts local communities for employment in geoparks and tourism marketing in the form of investments in ecotourism, rural tourism and health geotourism. Geotourism is closely related to the geology of treatment. One of these is the spa spring that has the therapeutic potential and plays a significant role in attracting domestic and foreign tourists. For this purpose the status of 11 hot springs in the geographical range between Ardabil city and Sarein city has been investigated in terms of parameters such as discharge, temperature, pH and anions and cations. However, this geographical area also has a number of natural and human hazards the most important of which is the occurrence of killing earthquakes. The earthquake of February 28, 1997 killed nearly 1000 people and destroyed many villages. The presence of spa springs on or near the faults and the establishment of recreational facilities in the area have threatened investment in the area. In this regard the status of clay, silt and sand percentage, Liquefaction Limit and Plasticity PI index in the study area were studied. It was observed that if in clay formations the amount of dough is about 40 to 50 percent and soil moisture reaches 25 to 35 percent there is a possibility of demolition of buildings and asphalt. According to the results of soil physics analysis, Sarein city is susceptible to soil infiltration and liquefaction during earthquake. In the final section the sanitary condition of the pools was investigated in terms of the possibility of dermatophyte fungi and 284 samples were examined. It was observed that due to the complete disinfection of pools with chlorine no dermatophytes were observed.

KEYWORDS
spa springs; earthquake; dermatophytes; geotourism; Sarein tourist area

Received: 11 May 2018
Accepted: 19 August 2019
Published online: 12 December 2019

Feyzolahpour, M., Einali, J., Gasemlu, H. (2019): Investigating the geotouristic risks of spa springs with emphasis on natural and human dangers in the study area between Sarein and Ardebil in the northwest of Iran. AUC Geographica 54(2), 194–206
https://doi.org/10.14712/23361980.2019.17

1. Introduction

The growing population of the world especially in developing countries, economic poverty and the low level of national income of the countries has caused the population of these countries to be in a state of economic strain. In examining the causes of these countries' retardation it was found out that despite the availability of natural resources and the of potential income sources they have accepted poverty and deprivation and in fact low levels of knowledge and technology and lack of awareness of how to exploit existing potential has increased the problems. One of the sources that can greatly increase revenues in the current era and by increasing GDP increase income on the international scene is the tourism industry, especially ecotourism which is a good source of revenue for most of the countries in the world. So now the income from tourism and ecotourism in countries like France, Spain, China and Germany is more than the national income in some of the underdeveloped countries. On the other hand the tourism industry will have problems for each region without careful planning and attention to ecological, local, cultural and social capabilities. In fact, the planned and perfect tourism system will advance to the proper use of the environment and various environmental, cultural, historical and similar resources in the region (Edgell et al. 2008). And this shows the position of strategic planning for the sustainable use of tourism resources especially ecotourism. Among many studies that have been done to develop ecotourism development strategies Johansen and Williams have been studying ecotourism development in a national South African wetland park (Johansen and Williams 2008).

In recent years researchers have dealt with many different aspects of ecotourism including Weaver and Lawton's article on ecological, economic, cultural and social effects of ecotourism and methods of quality control as well as ethics in it (Weaver and Lawton 2007). Jaafar and Maideen also study ecotourism products and services provided in four islands in Malaysia and issues related to the economic sustainability of small and medium-sized huts in these areas (Jaafar and Maideen 2012). Ecotourism encompasses a wide range of nature-based tourism activities. Examples include visiting national parks and protected areas, pristine natural areas, watching birds, circulating in natural environments, hiking, mountain climbing, visiting natural caves, studying plants and animals and ecological exploitation. This part of ecotourism is the interface between ecotourism and geotourism. In 1992 at the United Nations Conference on Environment and Development in Rio de Janeiro in command 21 geotourism was referred to and today is recognized as an appropriate tool for sustainable development. Undoubtedly the importance of geotourism is to emphasize a kind of holistic management perspective in areas such as the history of geological phenomena, ecosystems, land use, natural tourism and environmental education and sustainability (Kim et al. 2008). As a new option geotourism emphasizes

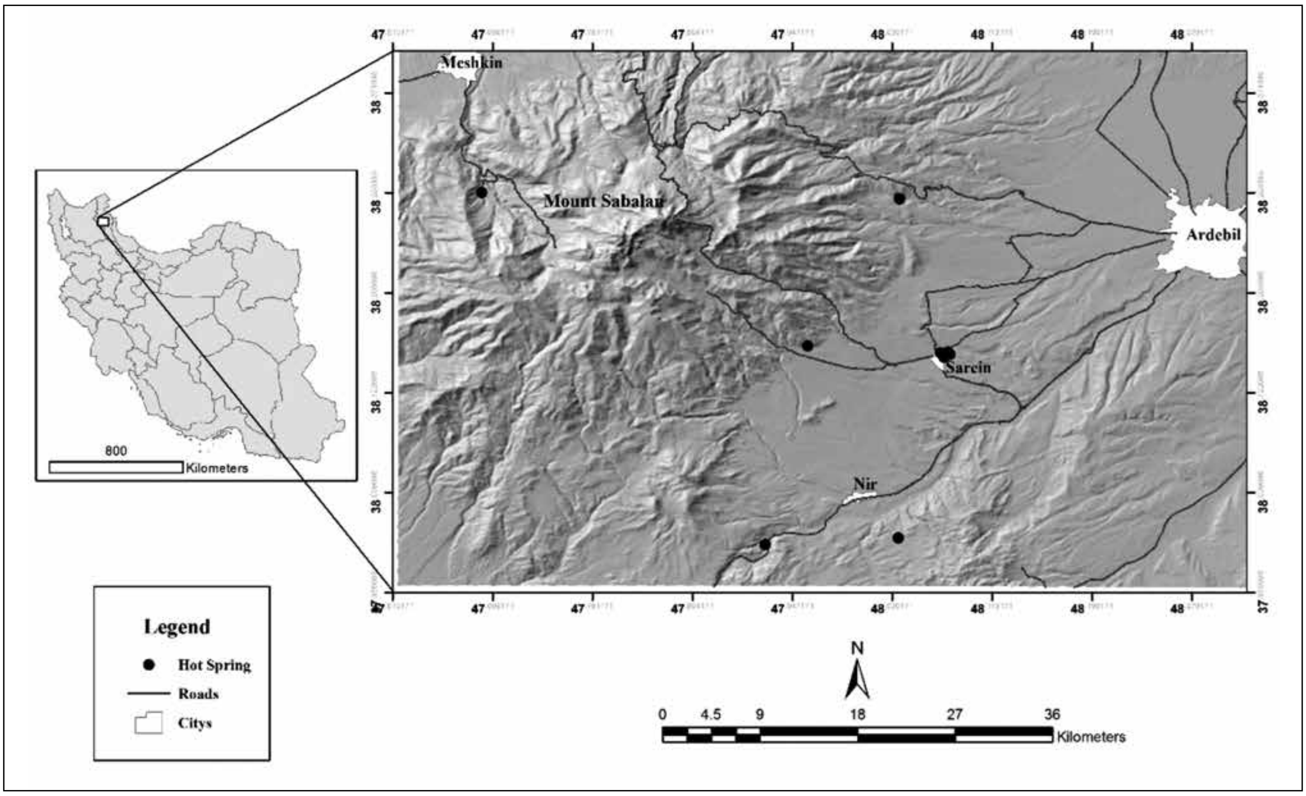

Fig. 1 Location of the study area in terms of location of hot springs.

not only all human and natural characteristics but also the function of each location. This form of tourism can play an important role in national development and diversification of the regional economy through planning based on the recognition of opportunities and limitations of geotourism. Geotourism is a form of cultural and environmental tourism that can be developed in the region with important geological works and is based on the preservation and expansion of the geological heritage that has enjoyed great growth in the world. It is expected that by the next decades the number of nature travellers which now accounts for 7 percent of the total number of travellers in the world will reach more than 20 percent. In 1977 in the UNESCO Earth Sciences Department the geopark program was officially launched (Zouros 2004). Panizza and Piacente (1993) refer to geomorphology as geomorphic assets in their research. Carton (1994) has studied geomorphology and geomorphologic processes that refer to geomorphologic locations as geomorphologic commodities. Hooke (1994) developed strategies to maintain and sustain the dynamics of geomorphologic sites. Grandgirard (1997) worked on the site conservation and geomorphology in his Ph.D. thesis and referred to the geosites as geosciences in his research.

As a region located in the desert and mountainous belt, Iran has special geological and geomorphological conditions. On the other hand, in different geological periods much of Iran was influenced by Paleozoic orogenic movements such as Caledonian and Hercynian before the Mesozoic era such as Laramide and the Cenozoic period such Pyrenean, Savian and Pasadenian which resulted in the rise of Iran. These geological movements have caused volcanic activity in the third period especially during the Eocene period and during the Quaternary period. This has led to the formation and activation of volcanoes in Iran. In some parts of the country karst and calcareous structures were established which resulted in the distribution of various caves in Iran. Glacial forms formed during the Pleistocene can be found in the northern, northwestern and western parts of Iran.

One of the issues that is encountered in the tourist attractions especially geotourism is the topic of hot springs which have been scattered in different parts of Iran and have been considered for its therapeutic efficacy. With more than 113 mineral water springs of good quality Iran can be one of the hotspots for tourist with natural landscapes and therapeutic applications but practically did not succeed in doing so. Investigation and recognition of mineral waters in Iran began in the second half of the 19th century by foreign tourists and foreign scientific delegations and tested on a number of mineral springs and their properties. In 1928 the first mineral spring was used in the north east of Tehran. In 1949 according to a plan by the Plan and Budget Organization the study of

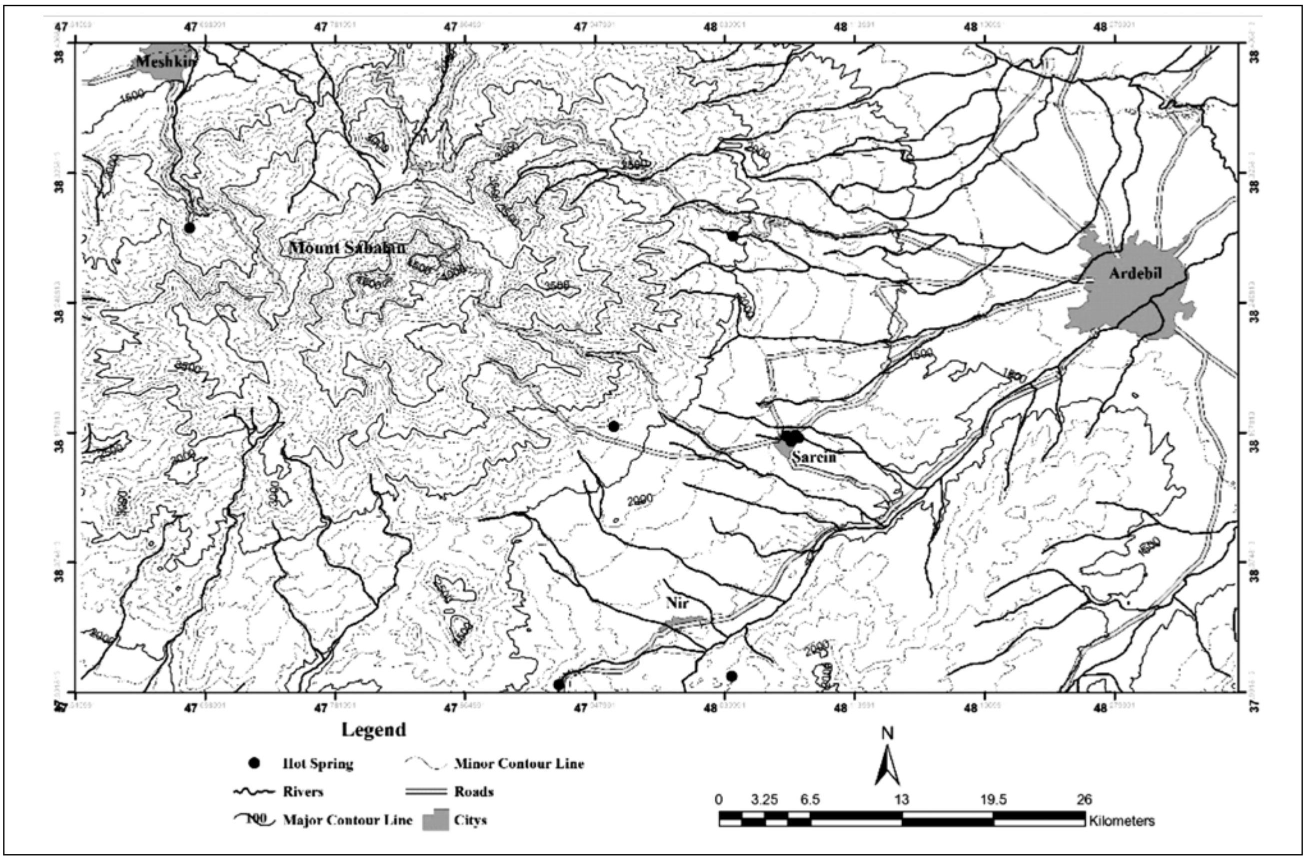

Fig. 2 Topography of the studied area.

the mineral waters of Mahallat and Ardebil was carried out. Determination of the mineral properties of Iran's mineral waters was initiated in 1961 by a team of hydrology from the Faculty of Pharmacy at the University of Tehran and continued until 1969 in which the physical, chemical and microbiological properties of many mineral springs were identified. It is worth noting that these studies have been ongoing and have resulted in the identification and investigation of about 350 mineral water springs in 40 regions of Iran.

Considering the scientific evidence of the therapeutic uses of springs on the one hand and the existence of hundreds of hot mineral springs in the geographical range of Iran on the other hand it is possible to plan and manage efficiently within the tourism industry in addition to the goals of spending leisure time using health applications.

The city of Sarein has become a geotouristic attraction due to the presence of hot springs. The presence of these conditions has attracted a lot of investment in the city and has produced significant positive and negative effects. Therefore, in this research the upcoming risks in the development of the city and tourist centers were examined. Due to the earthquake in the city of Sarein, the dispersion of faults in the area and their relationship with recreational units was discussed. In the biological survey the health status and pathogenicity of the hot springs pools are addressed. The probable germs in these pools were examined.

2. The study region

Sarein Tourist City is located in the south-east of the famous volcanic heights of Sabalan (4811 m) (Fig. 1). The western part of the city is located at higher altitudes than other areas due to its exposure to the Sabalan slopes. The eastern part of the city is located in a deeper region and hot springs are located especially in this area. The formation of Sarein city is completely in line with the natural environmental phenomenon.

The presence of highlands in the northwest and the alluvial valley in the east and faults in the city limits and its edges as well as the passage of the river from the city center completely affected the city's geomorphology. The peripheral lands of Sarein city are devoted to agricultural lands surrounded by pastures. Wasteland of about 84.5 hectares in the city of Sarein until 2005 accounting for 34.7% of the continuous urban texture has recently reached about 60.8 hectares. Considering the rapid development of the city it is expected that in the not-so-distant future of Sarein there will be plenty of uneven topographic levels at the surface of the Sabalan range. The general slope of the city of Sarein is in accordance with the state of the northwestern and southwest land.

The city is located at variable topographic levels between the 1640 m high and 1740 m elevation with an area of 430.36 hectares in the eastern slopes of the

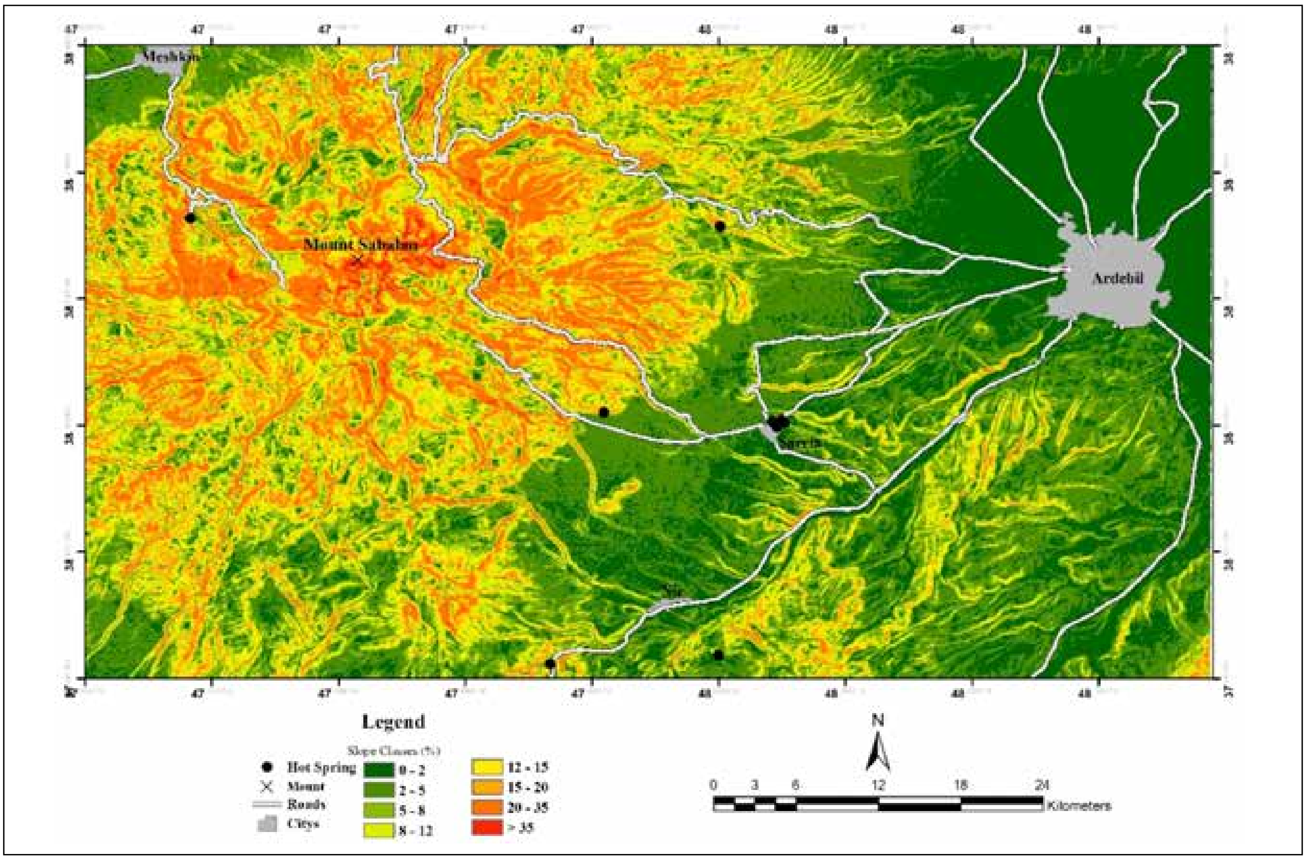

Fig. 3 Slope values of the study area.

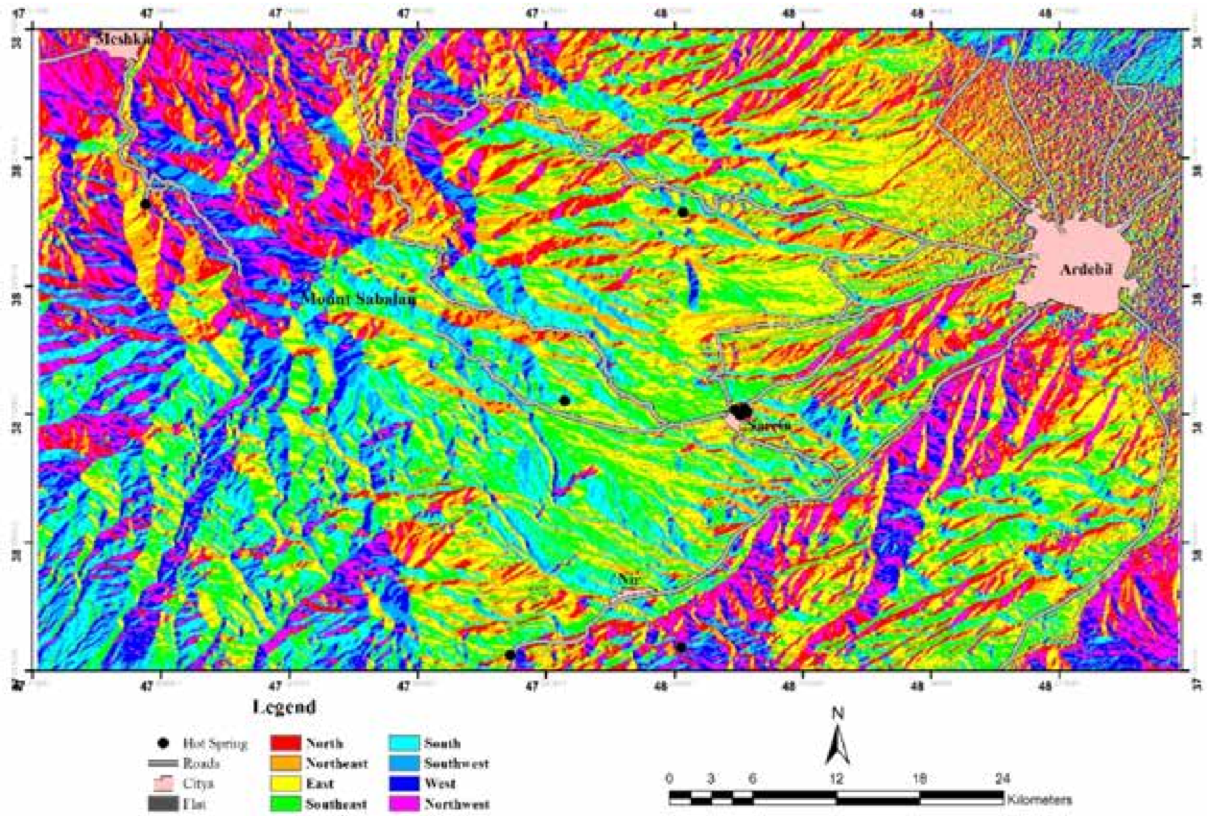

Fig. 4 Slope direction of the study area.

Sabalan Highlands. Therefore, the future expansion of the city to the east and north will be associated with many hydro-geomorphological problems because in the northern part the topographic condition is intensified in terms of the gradient and intensity of the earth's roughness. Parts of the north, northeastern and southwest of the city are developing at steep slopes which are not suitable for physical development of the city.

The tourist attractions of the studied area are important in comparison with tourist attractions in the world from a number of points. This study area is located in the southwest of the world's largest lake (the Caspian Sea). In the very short distances there are two types of seaside and mountainous climates which has caused 3 outlooks of the sea, forests and mountains. In terms of natural landscape and natural morphology it is comparable to that of Switzerland. The importance of tourism in this area compared to the tourist areas of the world can be better understood when we consider that in one day it is possible to visit and use the tourist attractions and ecotourism features such as mineral water springs, ski resort, volcanic views, forest and mountain views (Estelaji et al. 2011).

Tab. 1 Physical and chemical status of the springs.

Title of springs	Discharge (m/s)	Water Temperature in °C	PH	Anion	Cation
Eyes-Water Spa Spring	0.2	21	5.2	HCO_3	Ca
Hot springs of Hemmat Ardebil Coffee House	2	40.5	9	HCO_3, Cl	Na, Ca
Artemis Momtaz Coffee House hot springe	2	47	6.1	HCO_3, Cl	Na
Five Sisters Springe (Besh Bajjilar) Ardebil	3	35	5.9	HCO_3, Cl	Na, Ca
Seven Blocks Spring	2	20	5.3	SO	Ca
Boshli Spring	2	49	6.9	HCO_3, Cl	Na, Ca
Ghinarech spring	4	65	6.1	HCO_3, Cl	Na, Ca
Isti Su Spring	1.5	42	5.7	HCO_3, Cl	Na, Ca
Ilanjigh Spring	1	49	6.2	HCO_3	K, Na, Ca
Springe Dip Siz Gol	2	30	6.3	HCO_3	Na, Ca
Saghezchi Spring	1	39	5.8	HCO_3, Cl	Na, Ca

One of the existing capacities in the Sarein geotouristic zone is the existence of many hot and mineral water springs in this area, described by Rashidi et al. (2012). (Table 1) Due to the therapeutic capability and geotouristic attractions some of these hot springs can be considered as an international pole.

3. Methodology

The two issues of natural hazards such as earthquakes, and the health and biological hazards were examined in the analysis of the hazards in this region. The geology of Sarein should be considered to analyze natural hazards by emphasizing on the earthquake. The surface formations of the region were first examined in terms of soil thickness and the capillary ascent of water. The soil liquefaction limit was determined in soil formations to explore the vulnerability to earthquake. Results of the study conducted by Abedini (2013) were used to meet this objective. The fluidity limit in formations was determined by sampling three areas of Sarein. Thus, the percentage of clay, sand, silt, liquefaction limit and plasticity PI index were measured using the granulometric results (Abedini 2013).

Also, the status of formations of the area in terms of proximity to the Sabalan volcano was studied to determine the resistance or instability of the area in earthquakes. In addition to the above cases, it was necessary to identify the faults and their positions in the region. Therefore, the position of the faults from Sarein was determined. The conditions of the town before and after the severe earthquake on February 28, 1997 were considered. In all these stages, the conditions of the hot springs were examined according to these parameters. The risks related to the biological agents in the hot springs of Sarein were identified in the second stage, Therefore, the fungal infections were emphasized. The results of Seyedmousavi et al. (2007) were used in relation to fungal infections. 284 samples were collected from 11 hot spring springs to find the type of microbes. Samples of fungi and bacteria were identified in them. It was tried to create a relationship between the dispersion of fungi and bacteria in spas. In this regard, the dermatophyte was highly considered.

4. Results

4.1 Natural hazards in the geotouristic region of Sarein city

The current range of the Sarein city considering the slope is composed of rocks, clay, marl and igneous projectiles. The depth of the soils in the city's boundaries and its margins extend from the semi-deep erosion of coarse grains in the northwestern part of the city to relatively deep gravel in its low-slope lands in the south and south-east of the valley. In the current headquarters of the city not only two faults are distinguishable from the east to the west and from the north to the south but the Sabalan semi-volcanic volcanoes (in the tectonic active hydrothermal stage) are located in the east. Surface formation in the area showed that the soils of low-gradient areas (range 3–5%) were of heavy texture (51.45% clay texture). In terms of thickness the soils of the eastern and southeastern part of the city are semi-deep and deep. Water climbing especially freezing climb upward damages roads and pavements in Sarein city as well as the foundation of structures. Soils that are capable of expansion are swollen or as a result of freezing and increasing volume the floor of the building and the surfaces of the asphalt are swelling. Nevertheless, the city boundaries are located on the path to the fault valley and on relatively deep-sea alluvial deposits. Therefore, in this fine-grained form with surface and permeating waters design issues and robust materials will arised. Deepening of the ground and the use of robust anti-earthquake materials for large structures and maintaining the water level with the principles of drainage is necessary. Due to the high ductility pattern in fine-grained urban formations the potential for swelling of structures in the region is mostly moderate to higher-moderate (Table 2). The successive rise of moisture levels in the wet seasons of the year will lead to inflation and fluidization in the surface

Tab. 2 Location of harvesting samples, granulometric results for determination of tissue texture and ductile tolerance (Abedini 2013).

Plasticity PI index	Liquefaction Limit	Percent of Silt	Percent of Sand	Percent of Clay	Sample collection site
33.32	35.30	26.4	30.3	43	South entrance
42.23	38.22	25.32	22.25	51.45	Eastern regions alonge the valley of the city limits
34.6	31.22	23.5	43.42	34.35	North slopes

Tab. 3 Ductility Symbol and Potential of Inflation (Abedini 2013).

Liquefaction Limit	0–10	10–20	20–35	More than 35
Plasticity PI index	Little	Average	A lot	Too much

Fig. 5 A view of a lake located on the Kaldera Mountain of the volcano Sabalan.

formations of fine-grained clay and in the long run will destroy the base of the buildings. In places where discontinuous materials can saturate with water they can create fluidity. In gradient and hilling area of Sarein, high rainfall, high relative humidity and the number of freezing days (145 days per year) the presence of deep and fairly deep surface formations makes it very probable that soil leakage occurs during earthquake. An earthquake with a magnitude of 5 is a limiting threshold for soil leakage and lateral expansion of soil formations and in an earthquake with a magnitude 6.5 causes mass movement, avalanches and severe earth avalanches. Abedini (2013) declares the fluid level of the texture of the formations in the three sample areas in Sarein as follows:

In clay formations if the amount of dough is about 40 to 50 percent and the soil moisture content reaches 25 to 35 percent it is possible to destroy the building and the asphalt. According to the results of soil physics analysis Sarein city is susceptible to soil swelling and fluidization during earthquake. In the meantime, the lithological and tectonic conditions of the area should also be investigated. The lithological resistance of geological and surface formations depends on the type, shape, color, manner of mineralization, moisture percentage, weathering and tectonization. A wide area of the region consists of periodic layers of green tuff, marl and freshwater lime cobblestone and lava stone. In the northern and northeastern parts of villages and volcanic villages this formation consists of Andesi stone, basalt and trachy basalt. Lithologically these formations are mostly semi-resistant formations. The major geological phenomenon in this area is the Mount Sabalan volcano. It is the result of volcanic activity of the central type and its eruption mechanism is similar to the Italian Stromboli volcano. Volcanic activity of the upper Sabalan is based on volcanic and sedimentary bases of Eocene with an average altitude of 2700 m. The length of this volcanic mass is about 60 km and its width are about 45 km. The main crankcase of this volcanic mountain is a funnel-shaped hollow. In the summertime as a result of melting of the snow a lake is formed inside this slope.

Sabalan lava flows out in several stages and covers a total surface of 1200 square kilometers. Sabalan's early activities were related to Eocene. But its current cones are due to the subsequent steps in the Pliocene and Quaternary respectively. Sabalan caldera appears to have been created in the second phase of activity in Pliocene. One of the most noticeable side effects associated with magmatic activity is the large spread of mudflows in Sabalan's northern pine forests. These drains result from the combination of volcanic ash with melting water in the cold period. Apparently in Würm glaciation and during the formation of glacial masses in Sabalan its explosive activity was also started. Volcanic activity has melt the glaciers and the water from the melting ice is also smeared with volcanic ash and brought them into mudflowers. These masses eventually moved along mud floods along the valleys and pushed to the surface of the plains leading to Meshkinshahr and Ardabil. Conglomerates, lahar, tuffs and volcanic gray formations located in the north and east of Sarein city are also related to the early Quaternary. Lithologically these formations are mostly resistant to semi-resistant. Part of the Sarein area is formed by the material of the alluvial terraces that extends almost in the margins of the western and southwest streams of Sarein and the villages of this area.

The residential areas of the study area are often located on these formations. The thickness of this unit is more than 400 meters and spread from Sarein to the south and south-east of the region. Lithologically these formations are insufficient and semi-resistant. Therefore these structures are risky when earthquakes occur. The most significant potential hazard of these steep slopes will be the occurrence of slip and creep on soft and loose tuffs in the Sarin area. The slip on steep slopes causes a heterogeneous alignment or movement of the pillars and separation of the building from the ground. Liquefaction in saturated sandy soils leads to tilting and collapsing of buildings.

In total the margin area of Sabalan volcanoes and Ardabil area in northwest of Iran have been reported with high potential of seismicity. Situated in the southern volcanic slopes of Sabalan it has also transformed it into a city with a potential earthquake hazard. Considering the seismic state of Iran and the vulnerability of cities to earthquakes one of the approaches taken by urban planners to deal with this phenomenon is addressing the issue of urban immunization and preventative measures to reduce the damage caused by the earthquake. In terms of fault structure, we can classify Sarein's range into three major faults:

The Balikhlichay fault is about 70 km from the northern lake of Ardebil and is covered by the southern Quaternary alluvium in the north-eastern plain of Ardebil and continues to Astara. One of the main

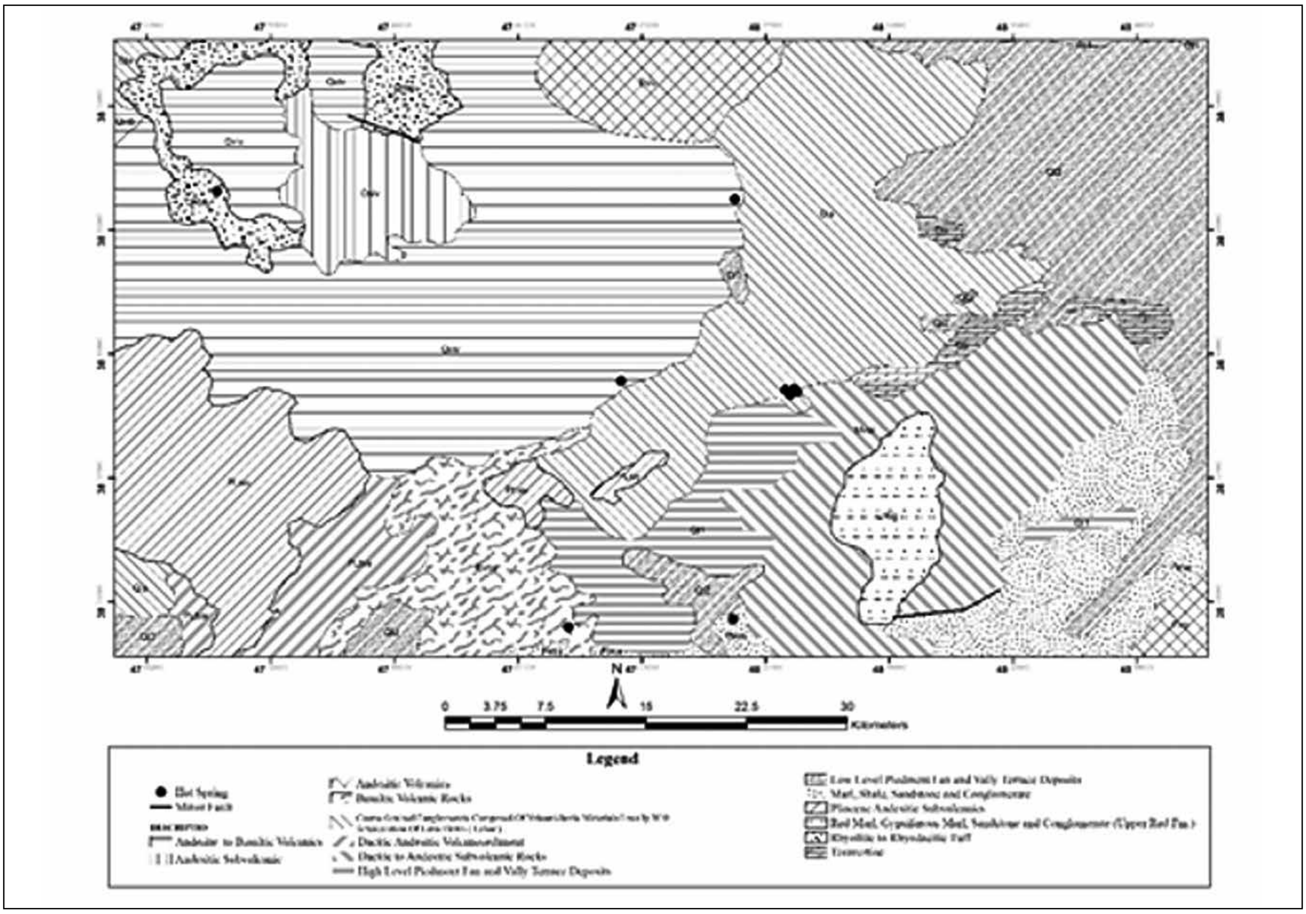

Fig. 6 Geological map of the study area with emphasis on the location of springs and faults.

signs of neo-terrestrial activity is the displacement of Miocene sediments and tilting of the Quaternary alluvium. A lot of spa springs run from the site of this active fault. The Alvares fault is 15 kilometers long with a northwest process extending south-east of the village of Alvares to Caldera in Sabalan. An important part of the Darwish river runs along this fault. Villadera fault begins with the northern and southern trend of the village of Sardabeh (north of Sarein city). A branch of it along the primary route is divided into two parts in the north of the city of Sarein; one branch of which passes through the city of Sarein and is again divided into two branches in the commercial area of the city. One branch ends in the village of Wind Kalakhoran and the other branch ends about 3–4 km from the city. Several spa springs have been formed along the Sarein fault. The location of the city of Sarein despite its adjacent active faults and faults in the city are highly threatened by tectonic and seismicity. Any displacement of these faults will result in earthquakes with destruction and damage to the city of Sarein. On the other hand with the onset of the earthquake the likelihood of blockage of hot water will be very high. By changing the direction of the springs practically some water treatments will be closed and the characteristics of the city's tourist performance will be changed.

Ardebil province is located on five earthquake faults including Barghoosh in the western part of Ardebil highlands, Dasht Moghan fault in Parsabad, Ardabil fault along the northeast to the northwest of the province, Masooleh fault in the southeastern part of Khalkhal and Rudbar fault in Astara in Namin area. On February 28 1997 an earthquake of 6.1 magnitudes occurred at a depth of 15 km. The number of people killed was about 965 and the number of injured was estimated at 2600. In this earthquake about 36 000 people became homeless. Studies in the area indicate that the highest severity of the earthquake is located in southeastern Sabalan and the greatest damage was reported in villages Golestan and Shiran. The magnitude of the high aftershocks occurred a few days after the main shake causing more damages and the collapse of buildings that damaged in the main shock and caused significant damages to the city of Sarein. Field observations after the earthquake showed that gradient thickness and sediment type had a significant effect on the severity of degradation. In the region the landslide phenomenon has not been significantly influenced by the slope due to the low slope during the earthquake. In the event of an earthquake rural houses were mostly made of inadequate material of poor quality mud. Such buildings have been completely destroyed in areas close to the earthquake site and

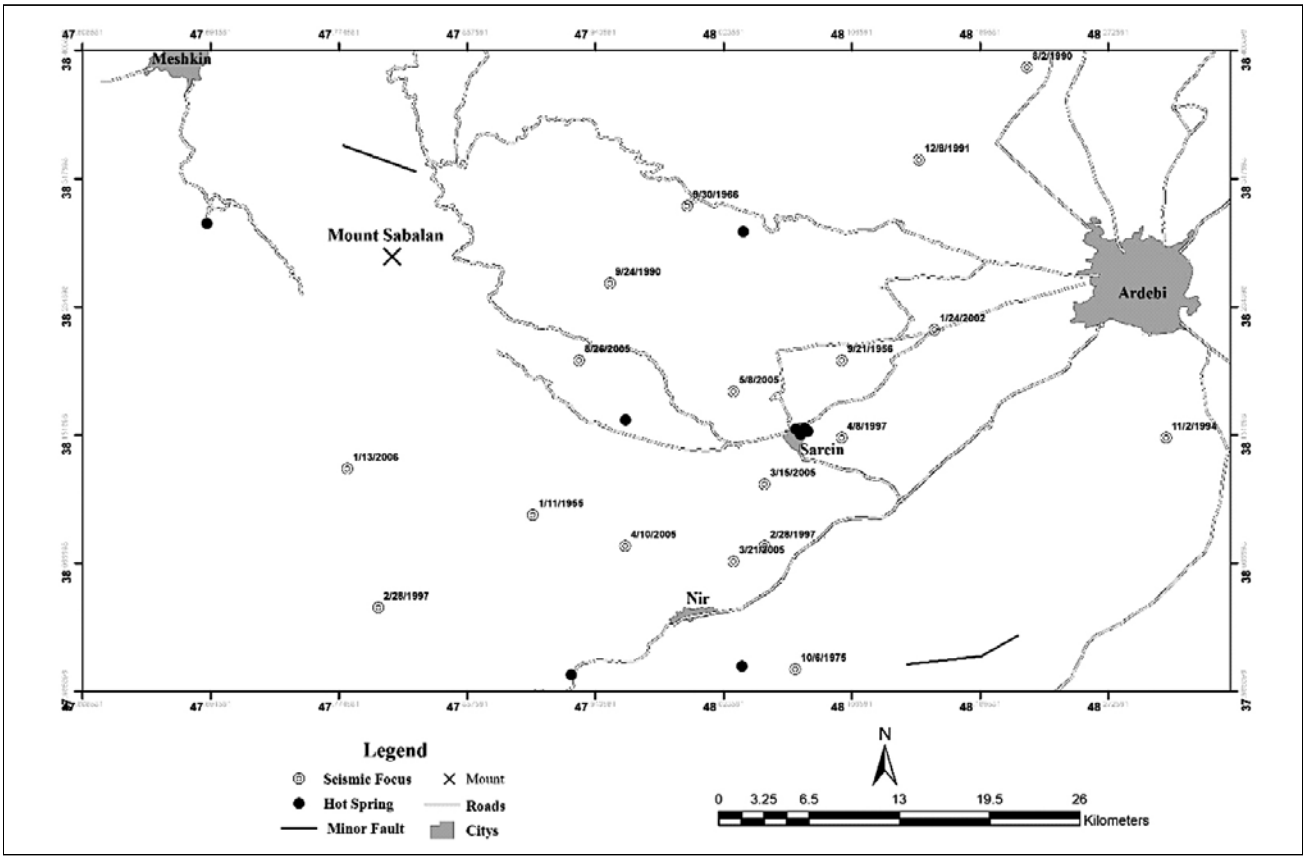

Fig. 7 Map of earthquake occurrence and location of springs in the studied area.

Fig. 8 An overview of the earthquake that occurred in 1997 in Sarein.

Fig. 9 An overview of the earthquake that occurred in 1997 in Sarein.

have been found to be deep-seated cracking in the distant regions and their local downturns. In most of the urban residential houses and some of the rural buildings except brick buildings the technical notes of the earthquake code were observed. The damage to such buildings has been less than that of mud-made houses. A few brick buildings were also found in nearby areas of the earthquake which due to non-observance of some technical points only suffered minor damage. The one that increased the depth of the disaster and increased the death toll was a severe snowfall at night of the earthquake and wolf attacks on the injured.

In spite of the severe earthquake and significant human damages within 20 years after the February 28, 1997 earthquake the city of Sarein has grown rapidly due to its natural attractions and has become one of the attractions of the Middle East and has witnessed rapid development. In this period mountaineering camps, ski resorts, spa pools and cable cars have been developed.

Another factor that threatens the city of Serein is the downfall of the earth. Earth subsidence is exacerbated in both cases naturally and sometimes also by the interference of humans on the earth and is an environmental hazard for urban construction. The center of the city is located from east to the west parallel to the fault valley. Today more than 10 to 15 floors of towers and modern hotels are rapidly expanding

Fig. 10 Growth and development of the city of Sarein after the earthquake of February 28, 1997. It is clearly seen in the image the Sabalan volcano that dominates the city and its construction.

Fig. 11 Another view of construction in the city of Sarein after the 1997 earthquake.

in the active tectonic zone. The core of the city is the location of fault division. In addition, the main stream along the fault line has been adapted. The site of the buried volcanic geological layers underneath the alluvial Quaternary alluvial deposits has created basis for formation of hot springs in the city. In the 1997 Ardabil earthquake by breaking part of the igneous section of the infrastructure hot water erupted in the middle of the city at an altitude of about 12 meters. It is expected that some earthquakes will dry up during earthquake due to physicochemical reactions within the earth or there may be created several hot water springs in the areas adjacent to them. So far after the start of huge and heavy construction the city of Sarein has not experienced a severe earthquake. An examination of all geological issues tectonics and hot water in the region indicates the instability of urban infrastructure.

According to Abedini (2013) most hot springs come out of the place and also adjacent places to faults. The existence of faults has led to the tectonic instability of the city lands. On the other hand the displacement of faults is one of the important factors in the gradual downsizing of the city on the path to the fault valley. The results of the sampling conducted by Abedini (2013) indicate a high percentage of clay (46%) silt (32%) and sand (25%). Therefore high presence of clay, high topographic slope and fault activity during earthquake could be risky. The higher the percentage of clay in the formations, the higher the amount of inflation, elasticity and fecundity. Continuous wetting of structures of foundation, chemical degradation, dissolution, decrease of internal resistance of materials and increase of shear stress cause bending, cracking the foundation and even falling it. Topographic and geological slopes with the high groundwater and hydrothermal velocities and trapping characteristics especially in low slopes and pits like the middle of the city play an important role in the risk-taking of structures. In the beginning the position of the city of Sarein was formed in relatively flat topographic lands but gradually with the expansion of physical space the city occupied more lands and encountered many hydro geomorphological bottlenecks. Considering the above-mentioned cases in the event of an earthquake exceeding 5.5 magnitude the city will suffer great damage

4.2 Analysis of hazards associated with biological agents in Sarein hot water springs

Given the increasing progress of human knowledge in controlling and eradicating fungal diseases including dermatophytosis in humans, the statistics still indicate that the disease is one of the most important health and therapeutic issues in the world and in Iran. Human fungal infections have increased considerably in recent years, one of the common causes of which is the increased exposure of people to the environment and public places infected with the transmission of disease to humans. The use of public places such as hot water pools is one of the ways in which superficial and cutaneous fungal infections spread to humans because they have a long life of up to 6 months (Rippon 1988).

Considering that to produce fungal skin disease other than the presence of pathogens in the environment other factors such as ambient temperature, relative humidity, age, occupation, living conditions and non-compliance with health standards are important for the development of fungal skin diseases (Rippon 1988) many researches and studies have been accomplished on the dermatophyte contamination in the environments and public places and their role in the transmission of the disease and how to prevent them in the world and in Iran.

In the study of pollutants in the southern barracks of the south of the country from 374 samples 90 cases (24.06%) of dermatophytes were isolated from the equipment and environment of the patients and 2 types of dermatophytes from one patient. They prove dermatophytes in sampling 7 bathrooms (63.63%) out of 11 baths (Emami et al. 1974).

Tab. 4 Service and operation status of hot water pools in Sarein tourist city in the summer of 2005 (Seyedmousavi et al. 2007).

Pool number	Pool status																
	Private	Governmental	Half private	Indoor	Outdoors	Lending				Shower situation		Dressing room		Disinfection	Service type		
						Slippers	Towel	Swimming suit	Hat	There must be	It does not have to exist	Solo	Share		Masculine	Feminine	Mixed
Pool No. 1	–	–	*	–	–	–	–	–	–	–	*	–	*	*	*	*	–
Pool No. 2	–	–	*	–	–	–	–	–	–	–	*	–	*	*	–	–	*
Pool No. 3	–	–	*	–	*	*	–	–	–	*	–	*	–	*	–	*	–
Pool No. 4	–	–	*	–	*	*	–	–	–	*	–	*	–	*	*	–	–
Pool No. 5	–	–	*	*	–	–	–	–	–	*	–	–	–	*	–	–	–
Pool No. 6	–	–	–	*	–	–	–	–	–	*	–	–	–	*	–	–	*
Pool No. 7	–	–	*	–	*	–	–	–	–	–	*	–	–	*	–	–	*
Pool No. 8	–	–	*	*	–	–	–	–	–	–	*	–	–	*	–	–	*
Pool No. 9	–	–	*	–	*	–	–	–	–	–	*	–	–	*	–	–	*
Pool No. 10	–	–	*	*	–	–	–	–	–	–	*	–	–	*	–	–	–
Pool No. 11	–	–	*	–	*	*	–	–	–	*	–	–	–	*	–	–	–

Tab. 5 Frequency distribution of fungi isolated from hot water pools of Sarein tourist city in summer 2005 according to type of pool (Seyedmousavi et al. 2007).

Title of Fungi	Pool number											Number of samples	Percentage
	1	2	3	4	5	6	7	8	9	10	11		
Aspergillus fumigatus	1	5	4	11	7	3	4	2	2	2	3	44	22.79
Aspergillus flavus	–	–	6	6	4	2	4	3	1	2	2	30	15.54
Aspergillus niger	2	3	2	5	5	4	3	1	3	2	–	30	15.54
SP penicillium	1	1	4	7	1	–	2	2	3	1	6	28	14.50
Mucor	–	1	2	1	1	–	2	–	–	–	1	8	4.14
Fusarium Oxysporum	–	–	1	1	1	–	1	–	1	1	2	8	4.14
Aspergillus	–	–	1	–	1	1	–	–	1	–	–	5	2.59
Geotrichum	1	–	3	1	–	–	–	–	–	–	–	5	2.59
Alternaria	–	–	–	–	–	–	–	1	–	1	–	4	2.07
Candida (fungus)	2	–	2	–	–	–	–	–	–	–	–	4	2.07
Scopulariopsis	–	–	1	–	–	–	–	1	–	1	–	3	1.55
Unknown	–	–	–	–	2	–	–	–	–	–	–	3	1.55
Mycelia sterilia	–	1	–	1	–	–	–	–	–	–	–	2	1.03
Rhizopus	1	–	–	1	–	–	–	1	–	–	–	2	1.03
Trichosporon	–	–	1	–	–	–	–	–	1	–	–	2	1.03
Acremonium	–	–	–	–	–	–	–	–	–	1	1	2	1.03
Curvularia	–	–	–	–	1	1	–	–	–	–	–	2	1.03
Trichotecium	–	1	–	–	–	–	–	–	1	–	–	2	1.03
Pseudoallescheria boydii	–	–	–	–	–	–	–	–	–	–	2	2	1.03
Chysosporium	–	–	1	–	–	–	–	–	–	–	–	1	0.51
Ulocladium	–	–	–	–	–	–	–	–	–	–	1	1	0.51
Sepedonium	–	–	–	–	–	–	–	–	–	–	1	1	0.51
Bipolaris	–	–	–	–	1	–	–	–	–	–	–	1	0.51
Paecilomyces	–	–	–	–	–	–	–	1	–	–	–	1	0.51
Stemphylum	–	–	–	–	–	–	–	–	–	–	1	1	0.51
Sreptomycin	–	–	1	–	–	–	–	–	–	–	–	1	0.51
Total	8	12	29	36	24	11	16	12	13	11	21	193	100

Since there has not been any study on the status of fungal growth in hot water pools in Sarein tourist town, therefore in this section the study of epidemiology of fungal diseases identification of fungal contamination and their sources, determination of dominant fungal species, identification of ways to transmit the disease and how to prevent them in hot water pools of Sarein tourist city are discussed.

In a study by Seyedmousavi et al. (2007) 284 samples were taken at different points in the Sarein Hot Springs Pools. In this study 284 specimens were collected from different parts of the pools 193 fungal plates (67.95%) and 43 positive plates (15.14%) and none of the 48 plates (16.9%) were extracted. Aspergillus fumigatus was isolated with 44 positive plates (22.79%) Aspergillus feltus and Aspergillus spp. with 30 positive plates (15.54%) and penicillium with 28 positive plates (14.50%) the most abundant fungi.

In this study no dermatophyte fungi nor any real dimorphic pathogens were isolated from carpet and water samples. One of the factors of the absence of human dermatophytes in the pools examined in this study is the presence of chlorine in the pool water and the spread of this water around the pool because there are more commuted and the contact of the legs and body of swimmers in this place has been more. On the other hand, given that the amount of chlorine in most pools is not standard one can not specifically comment because in talking to the authorities responsible for these pools they often did not know the exact amount of chlorine intake and expressed their dependence on the number of swimmers and water pollution and most of the pools were discharged every 24 hours and disinfected with perchlorin. Therefore, it is likely that another reason for the lack of dermatophyte agents in the pools is that the pool managers as well as the health authorities in the hot water pools of the Sarein tourist town pay more attention to the cleaning and control of the pools.

According to the results of this study it can be stated that it is possible that another reason for not finding dermatophytes in these pools is the high temperature of the mineral water so that the pool areas are heated above 40 degrees Celsius and the temperature inside the sauna room was 40 to 45 degrees Celsius. Therefore the high temperature of mineral water is considered as a preventative factor for the growth of pathogenic fungi. It should be noted that all findings of this section are the result of research by Seyedmousavi et al. (2007) which was conducted on Sarein hot springs.

5. Conclusion

This research investigates the attraction and hazards of the geothermal springs of Sarein district spa. Therefore, the research structure was divided into two sections of attraction and hazard. The hazards segment was also divided into two sections of the hazards associated with the earthquake and the risks associated with the disease outbreak. Regarding the potential of spa springs about 11 hot springs were selected and features such as discharge temperature pH and anion and cation values were extracted in each source separately.

In the context of natural hazards an analysis of the earthquakes in the area especially the earthquake of February 28, 1997 was carried out and accordingly it was necessary to draw geological maps, topography slope and direction of gradient and earthquake centers mentioning the location of the spa springs. Investigations showed that there was no reported earthquake occurring on this date due to the low slope of the occurrence area such as landslide or rock falls but due to the straw structure of buildings, high human casualties were observed. One of the most problematic measures in this area is the development of construction towards the high slopes in the mountain range of Sabalan.

In the final section of the study the risk factors related to biological factors in the pools of Sareen spa springs were analyzed. To this end with the emphasis on research by Seyedmousavi and colleagues (2007) of 284 samples no dermatophytic fungi were extracted from water and spa pools. The main reason for this is the complete disinfection of pools by chlorine. So the pools used by tourists do not contain any dermatophytes.

According to the above-mentioned items it can be accurately stated that the recreational and tourist area of Sarein has been high in terms of geo-tourism capabilities and has resulted in economic flourishing in the Sarein area.

References

Abedini, M. (2013): Investigation of depositional formation of Sarein Tourist Town site emphasizing topography, tectonic and climate to develop optimal urban land use. Geography and Environmental Planning Journal 49(1), 13–16.

Carton, A., Cavallin, A., Francavilla, F., Mantovani, F., Panizza, M., Pellegrini, G. G., Tellini, C. (1994): Ricerche ambientali per l'individuazione e la valutazione dei beni geomorfologici – metodi ed esempi. Il Quaternario Italian Journal of Quaternary Sciences 7(1/b), 365–372.

Edgell, D., DelMastro, A., Maria, S., Ginger, R., Swanson, J., (2008): Tourism Policy and Planning Yesterday, Today and Tomorrow. London: Routledge.

Emami, M., Asghari, M., Alilou, M. (1974): A Study of Dermathophytes Flora Among Three Army Bases in Southern Iran. Iranian Journal of Public Health 3(4), 199–205.

Estelaji, A., Allah Gholi Nejad, M. (2011): Planning for Sustainable Development with Emphasis on Tourism. Case study: Sarein and Surrounding Villages. Geography 9(30), 129–144.

Grandgirard, V. (1997): Géomorphologie, protection de la nature et gestion du paysage. Thèse de doctorat en géographie, Université de Fribourg.
Hooke, J. M. (1994): Strategies for conserving and sustaining dynamic geomorphological sites. In: O'Halloran, D. et al. (eds): Geological and landscape conservation. London: Geological Society, 191–195.
Jaafar, M., Maideen, S. A. (2012): Ecotourism-related products and activities, and the economic sustainability of small and medium island chalets. Journal of Tourism Management 33(3), 683–691, https://doi.org/10.1016/j.tourman.2011.07.011.
Johansen M., Williams, J. B. (2008): Designing of Strategic Planning for Ecotourism Development in Table Mountain National Park by Using of Freeman Method. Journal of Environmental Management 42(1), 23–34.
Kim, S., Kim, M., Park, J., Guo, Y. (2008): Cave Tourists, Characteristics, Motivations to Visit and the Segmentation of their Behavior. Asia Pacific Journal of Tourism Research 13(3), 299–318, https://doi.org/10.1080/10941660802280448.
Panizza M., Piacente S. (1993): Geomorphological assets evaluation. Zeitschrift für Geomorphologie, Suppl. Bd. 87, 13–18.
Rashidi M., Rameshat M., Safe A., Khoshhal Dastjerdi. J. (2012): Health Tourism in Iran (Emphasizing on ranking of therapeutic properties of selected springs in Ardabil province). Journal of Health 3(1), 49–57.
Rippon, J. W. (1988): Medical mycology: the pathogenic fungi and the pathogenic actinomycetes. Philadelphia: W. B. Saunders Publ.
Seyedmousavi, S., Fataei E., Hashemi S., Geramishoare M. (2007): Fungal Flora in Mineral Swimming Pools of Sarein. Journal of Ardabil University of Medical Sciences 7(2), 146–154.
Weaver, D. B., Lawton, L. J. (2007): Twenty years on: the state of contemporary ecotourism research. Tourism Management 28(5), 1168–1179, https://doi.org/10.1016/j.tourman.2007.03.004.
Zouros N. (2004): The European Geoparks Network. Geological heritage protection and local development. Episodes 27(3), 165–171, https://doi.org/10.18814/epiiugs/2004/v27i3/002.

Economic impacts of landslides and floods on a road network

Mike G. Winter[1, 2, *], David Peeling[3], Derek Palmer[4], James Peeling[3]

[1] Winter Associates, Kirknewton, Midlothian, U.K. formerly Transport Research Laboratory (TRL), Edinburgh, U.K.
[2] University of Portsmouth, Portsmouth, U.K.
[3] Transport Research Laboratory (TRL), Wokingham, U.K.
[4] Formerly Transport Research Laboratory (TRL), Wokingham, U.K., now Independent Transport Economist, U.K.
* Corresponding author: mwinter@winterassociates.co.uk

ABSTRACT

Even in the absence of serious injuries and fatalities, landslide and flood events can have significant socio-economic impacts. These include the severance of access to and from relatively remote communities for services and markets for goods; employment, health and educational opportunities; and social activities. The economic impacts can be classified as: direct economic impacts, direct consequential economic impacts, and indirect consequential economic impacts. In addition, the vulnerability shadow cast can be extensive, and its geographical extent can be determined by the transport network rather than the relatively small footprint of the event itself. Using a number of debris flow events and a flood event in Scotland this paper places values on the economic impacts of landslides and floods. It also demonstrates the widespread impact of the events by means of the vulnerability shadow that is cast.

KEYWORDS

landslides; floods; hazard; risk; economic; social

Received: 17 May 2019
Accepted: 15 October 2019
Published online: 18 November 2019

Winter, M. G., Peeling, D., Palmer, D., Peeling, J. (2019): Economic impacts of landslides and floods on a road network. AUC Geographica 54(2), 207–220
https://doi.org/10.14712/23361980.2019.18

1. Introduction

A series of debris flows associated with monthly average rainfall substantially in excess of the norm occurred in Scotland in August 2004. The resulting landslides affected important parts of the trunk (strategic) road network, linking both cities and smaller, remote communities. Notable events occurred (Figure 1) at the A83 between Glen Kinglas and north of Cairndow (9 August), the A9 north of Dunkeld (11 August), and the A85 at Glen Ogle (18 August). While there were no major injuries, 57 people had to be airlifted to safety at A85 Glen Ogle when they became trapped between two major debris flows.

The A83 Rest and be Thankful site, while not affected in August 2004, has been extremely active in recent years with multiple debris flow events and road associated closures; in particular events in 2007 (Figure 2), 2008, 2009, 2011, 2012, 2014, 2015, 2017 and 2018 had an adverse effect on the travelling public. As a result the area has become the focus of not only concern but also of extensive landslide risk reduction activities, in the form of both management and mitigation and. a study was commissioned to assess and make recommendations on potential landslide remediation actions at the Rest and be Thankful site (Anon. 2013a; Winter, Corby 2012).

The rainfall-induced landslides occur in thin deposits of weathering products and other debris overlying rock that frequently dips out of the slopes which may be at angles of up to around 36°. Detailed descriptions of events are available (Winter et al. 2006), as are approaches to hazard and risk assessment (Winter et al. 2005, 2009, 2013), and approaches to management and mitigation (Winter 2014a, 2016). The links with rainfall are described by Winter et al. (2010, 2019).

In the absence of serious injuries and fatalities to those involved, the primary impacts of these events were economic and social. Such impacts include the

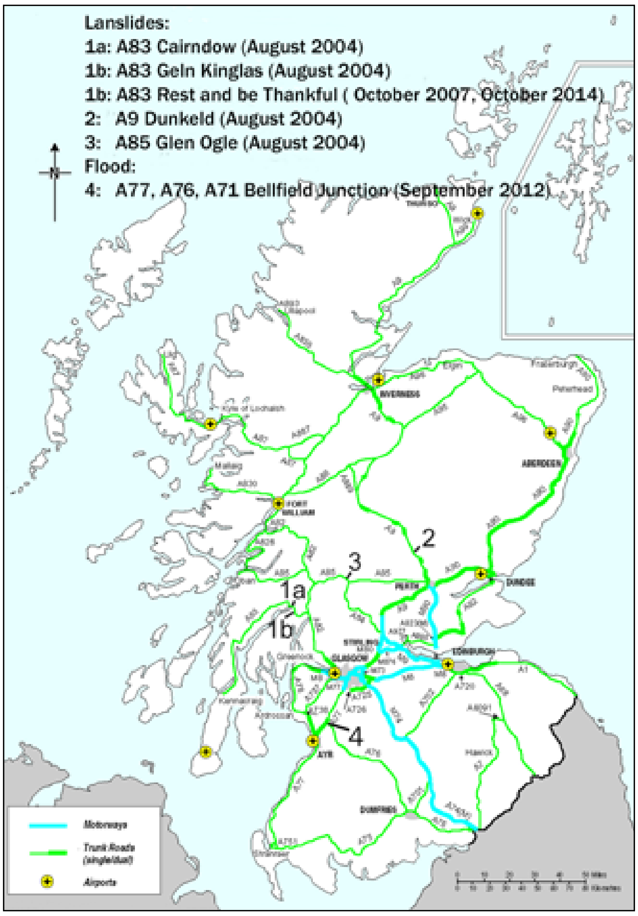

Fig. 1 Map showing the strategic road network in Scotland. The locations of the three main debris flow event groups that affected the trunk road network in Scotland in August 2004 (1, 2, 3) are shown (from Winter et al. 2009). The locations of the later A83, 2007 and 2014, landslides are coincident with the 2004 event at this scale and the A77, A76, A71 Bellfield Junction flood event from 2012 is also shown. The locations for the events correspond to those shown in Figures 4 (1b), 9 (2), 10 (3) and 11 (4).

Fig. 2 View of the debris flows above and below the A83 on the approach to the Rest and be Thankful (location approximately that of 1b in Figure 1). The head scar is at approximately 370 m Above Ordnance Datum (AOD), the A83 at 240 m AOD and the Old Military Road (OMR) at 180 m AOD.

cost of delays and diversion on transport networks and the severance of access to and from relatively remote communities for services and markets for goods; employment, health and educational opportunities; and social activities.

In 2004 for example, the A83, carrying up to 5,000 vehicles per day (all vehicles two-way, 24 hour annual average daily traffic, AADT) was closed for slightly in excess of a day, the A9 (carrying 13,500 vehicles per day) was closed for two days prior to reopening, initially with single lane working under convoy, and the A85 (carrying 5,600 vehicles per day) was closed for four days. The traffic flow figures are for the most highly trafficked month of the year (July or August). Minimum flows occur in either January or February and are roughly half those of the maxima reflecting the importance of tourism and related seasonal industries to Scotland's economy. Substantial disruption was thus experienced by local and tourist traffic, and goods vehicles.

This paper describes a study to assess the economic impacts of selected debris flow events in Scotland, based on the scheme set-out by Winter and Bromhead (2012). The impacts of floods can be assessed using the same set of principles and metrics and the impacts of one such event are also reported here.

2. Economic impacts

Due to the major contribution that tourism makes to Scotland's economy the impacts of road closures can be particularly serious during the summer months, during which period debris flows usually occur in July and August. Nevertheless, the impacts of any debris flow event occurring during the winter months, between October and November, and in January when debris flow usually occurs, should not be underestimated and events are arguably more frequent during the winter. Not surprisingly, the debris flow events described created a high level of interest in the media in addition to being seen as a key issue by politicians at both the local and national level. Indeed, the effects of such small events which may, at most, directly affect a few tens of metres of road cast a considerably broader vulnerability shadow (Winter, Bromhead 2012), which defines the geographical extent of the impacts of landslides and floods. The qualitative economic impacts of such landslide (and flood) events include:

- the loss of utility of parts of the road network,
- the need to make often extensive detours in order to reach a destination,
- the severance of access to and from relatively remote communities for services and markets for goods; employment, health and educational opportunities; and social activities.

The economic impacts of a landslide event that closes a road, and its associated vulnerability shadow, were summarized by Winter, Bromhead (2012), in three categories, as follows:

- direct economic impacts,
- direct consequential economic impacts,
- indirect consequential economic impacts.

Direct economic impacts: The direct costs of clean-up and repair/replacement of lost/damaged infrastructure in the broadest sense and the costs of search and rescue. These should be relatively easy to obtain or estimate for any given event, provided that records are still available.

Direct consequential economic impacts: These generally relate to 'disruption to infrastructure' and are really about loss of utility. For example, the costs of closing a road (or implementing single-lane convoy working with traffic lights) for a given period with a given diversion, are relatively simple to estimate using well-established models. The costs of fatal/non-fatal injuries and other incident accident costs may also be included here and may be taken (on a societal basis) directly from published figures. While these are set out for the costs of road traffic accidents, or indeed rail accidents, there seems to be no particular reason why they should be radically different to those related to a landslide or flood as all such incidents are likely to include the recovery of casualties from vehicles. Indeed, for events in which large numbers of casualties may be expected to occur, data for rail accidents may be more appropriate.

Indirect consequential economic impacts: Often landslide events affect access to remote rural areas with economies that are based upon transport-dependent activities, and thus the vulnerability can be extensive and is determined by the transport network rather than the event itself. If a given route is closed for a long period then confidence in, and the ongoing viability of, local business may be affected, for example. Manufacturing and agriculture (e.g. forestry in western Scotland), are a concern as access to markets is constrained, the costs of access are increased and business profits are affected and short-term to long-term viability may be adversely affected; in Jamaica a landslide on the B1 route in the Blue Mountains effectively severed local coffee production from the most direct route to the international market for that high value product (Figure 3). Perhaps of even more concern are the impacts on tourist (and other service economy) businesses. It is important to understand how the reluctance of visitors to travel to and within areas affected by landslides or floods is affected after an event that has received publicity and/or caused casualties and how a period of inaccessibility (reduced or complete) affects the short- and long-term travel patterns to an area for tourist services. Such costs form a fundamental element of the overall economic impact on society of such events. They are thus important to governments as they should affect the case for the assignation of budgets to landslide risk mitigation and remediation activities. However, these are also the most difficult costs to determine as they are generally widely dispersed both geographically and socially. Additionally, in an environment in which compensation might be anticipated, albeit often erroneously, those that have the best data, the businesses affected by such events, are also those that anticipate such compensatory events.

Typically analyses of the economic impacts of landslides and other natural disasters focus on the direct and/or indirect impacts, with the latter being generally analogous to direct consequential as set-out above (Schuster 1996; Highland 2006; Hearn et al. 2008; Chang, Nojima 2001; Bono, Gutiérrez 2011; Bil et al. 2015; Klose et al. 2015). Attempts to evaluate the indirect consequential impacts of such event are, at best, rare.

The vulnerability shadow cast can be extensive and its geographical extent can be determined by the transport network, including closures and diversionary routes, rather than the relatively small footprint of the event itself (Winter, Bromhead 2012). In the particular case of the event at the A83 Rest and be Thankful in October 2007 of the order of around 400 m^3 of material was deposited at road level with a footprint that closed a few tens of metres of the road (Winter 2014); the vulnerability shadow can be estimated to be of the order of 2,800 km^2 (total area approximately 3,500 km^2, 20% allowed for areas of sea) (Figure 4) which is, for

Fig. 3 Landslide on the B1 road at Section in Portland Parish, Jamaica. This event severed much of the local coffee production industry from the ports used to ship the product to market. (The picture is a photo-collage and some distortion is inevitable.)

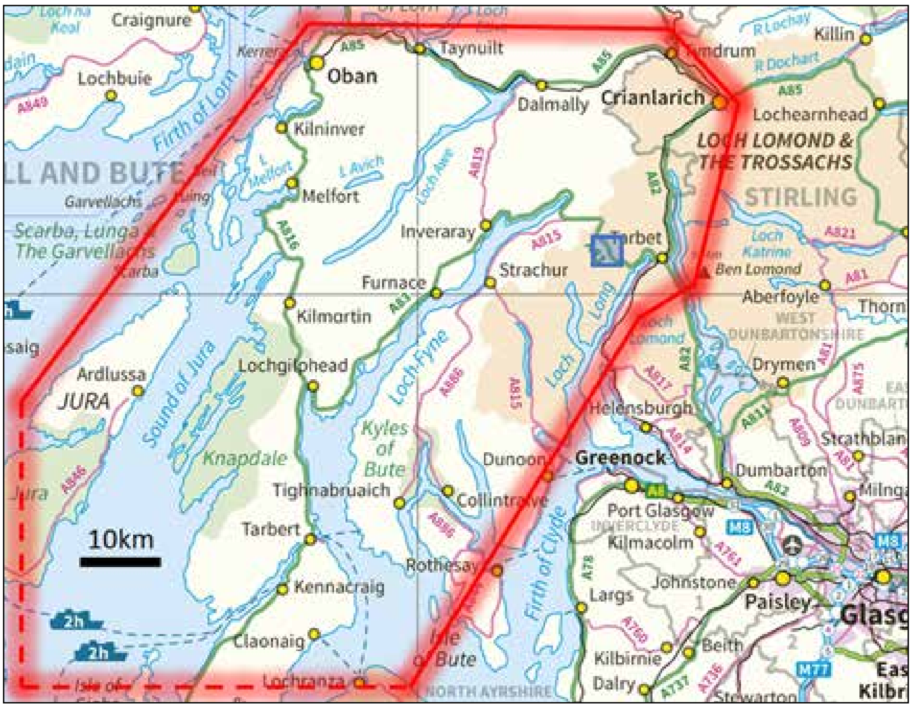

Fig. 4 A relatively small debris flow event (blue rectangle) closed the A83 at the Rest and be Thankful in October 2007; the vulnerability shadow that was cast (bounded in red) was extensive (Winter 2014a; 2014b). See Figure 1 (1b) for the event location in the wider geographical context. (Image based on OS Route Planner 2016 Map. © Crown Copyright. All rights reserved Scottish Government 100020540, 2019.)

the purpose of comparison, approximately two-and-a-half times the total land area of Hong Kong SAR.

The area has a population density of approximately 13 people/km^2 (www.argyll-bute.gov.uk) and the event thus had the potential to have had an economic impact upon up to approximately 36,400 people in Argyll, Bute, plus any transient (e.g. tourist) population.

It is instructive to make some simple comparisons with Hong Kong SAR, which has an average population density of around 6,500 people/km^2 (www.gov.hk). This dictates a much greater transport network density. Thus, and purely for the sake of comparison, in order to have an economic impact on the same number of people the vulnerability shadow cast need only be approximately 5.6 km^2 (2 km by 2.8 km, for example).

It is not suggested that the economic impacts would be similar for events with vulnerability shadows of these diverse sizes in Argyll, Bute and Hong Kong. However, it is clear that the low density/dispersed network in Argyll, Bute dictates a large vulnerability shadow while the much more dense/less dispersed network in Hong Kong means that the vulnerability shadow will be much less extensive, with the possible exception of events that affect critical infrastructure corridors, as more alternative routes will exist and will be more proximal to the event (Winter 2014b).

As part of this work (Winter et al. 2018), the vulnerability shadow has been evaluated using knowledge of the local transport networks and the socio-economic activity associated with the network that has been built up over a period of 30 years. This includes a holistic evaluation of major nodes, origins and destinations and includes both experience and knowledge gleaned from formal surveys (e.g. Winter et al. 2013).

The economic impact and the vulnerability shadow are concepts that apply equally to other discrete climate driven events, such as floods, that may close parts of the road network. Both landslides and floods are generally thought to be likely to increase in frequency as a result of climate change (Galbraith et al. 2005; Anon. 2011; Winter et al. 2010; Winter, Shearer 2013).

Notwithstanding the above, it is clear that for some events the hazard itself, and not the transport network or, more pointedly, its density, that determines the location, shape and extent (morphology) of the vulnerability shadow. It is therefore important to recognise that the morphology of the vulnerability shadow caused by other types of event (e.g. glacial lake outburst floods), may be determined by the nature of the hazard itself.

An example in which the hazard itself determines the vulnerability shadow is that of the Seti River debris flow in Nepal (Figure 5). On 5 May 2012 a major event caused significant erosion and deposition in the river channel over an approximate 40 km length. The event was thought to have resulted from a failed landslide dam. However, subsequent inspection of satellite imagery and aerial photography (Petley, Stark 2012; Petley 2014), and more detailed site inspection and investigation (Dahal , Bhandary 2013) led to a rather different conclusion; that the event was a debris flow initiated by part of a 22 Mm^3 rock avalanche originating on the slopes of Annapurna IV entering the upper stream channel at high speed. An estimated 71 people lost their lives at Kharapani, some 20 km north of Pokhara. The vulnerability shadow was constrained by the dimensions of the hazard flow within the stream channel, extending beyond these bounds only where infrastructure was damaged, including the footbridge at Kharapani.

Fig. 5 Residents of Kharapani located on the platform in the middle distance on the Seti River, Nepal, were among fatalities from the 5 May 2012 debris flow event. The abutment of the suspended footbridge is on the platform.

3. Methodology

3.1 Direct Economic Impacts

Direct economic impacts should be the most straightforward to determine. Indeed, this has generally proved to be the case with relatively recent events that occurred within the currency of existing network operation contracts. Thus, data relating to the 2007 and 2014 A83 Rest and be Thankful events were readily available from the company responsible for the operation of roads in this north-west part of Scotland at the times of enquiry, as was data for the 2012 flooding event.

Data on direct economic impacts was generally collected from the records of the companies responsible for operating the road network. However, data from less recent events such as the landslide events of 2004 (Winter et al. 2005, 2006, 2009) proved more difficult to obtain largely as both the operators and auditors had changed since the events occurred; as Highland (2006) points out past data are generally labour intensive to retrieve.

The passage of time and the consequent lack of readily available records limited the resolution and reliability of the data that can be obtained for the 2004 events. What data has been obtained was derived from high level reporting documents to Scottish Ministers and Senior Civil Servants and covers all three of the event groups from August 2004 (A83, A9 and A85). This data has been interpreted and broken down to the best of the ability of the original authors and editors of the Scottish Road Network Landslide Study reports (Winter et al. 2005, 2009).

3.2 Direct Consequential Economic Impacts

Direct consequential economic impacts are related to ‘disruption to infrastructure’ and concern loss of utility or the costs imposed on road users. For example, the costs of closing a road (or implementing single-lane working with traffic lights) for a given period with a given diversion, are relatively straightforward to estimate using well-established models. The costs of fatal/non-fatal injuries and incident damage only accidents are also included here and may be taken (on a societal basis) directly from published figures (Anon. 2013b). While these are set out for the costs of road traffic accidents (figures are also available for rail accidents) there seems to be no particular reason why they should be radically different to those related to a landslide or flood as both are likely to include the recovery of casualties from vehicles. Indeed, for events involving large numbers of casualties, it can be argued that data relating to railway accidents may be more appropriate. In all of the cases presented here the accidents have been non-injury accidents (damage only) and the numbers of vehicles involved have been estimated from contemporaneous photographs taken by the first author and others.

If a road is closed, either fully or partially, as a result of an event some or all of the users of that route will have to take an alternative, diversionary route, which may be significantly longer than the primary route. Even if no diversion is necessary, then the capacity of the road will likely be reduced (e.g. through a lane closure or the imposition of a speed limit) meaning that queues form, particularly at peak times, slowing the traffic flow. These effects can significantly increase road users’ journey times. It is also pertinent to note that such queues if formed adjacent to areas of high landslide susceptibility can significantly increase the risks to the affected road users.

The QUADRO (QUeues And Delays at ROadworks) model provides a method for assessing the costs imposed on road users while roadworks are being carried out, considering:

- Delays to road users: the change in road users’ journey times, priced using the value of their time (e.g. cost to their employer’s business of the time spent travelling during the working day) based on the type of vehicle, its occupants and trip purpose.
- Fuel carbon emissions: the change in carbon emissions due to vehicle fuel consumption, based on average figures per litre of fuel burnt and costed using estimated abatement costs (see STAG and WebTAG) (Anon. 2012a, 2012b).
- Accident costs: the change in the occurrence of accidents, in terms of the additional delay caused and the direct costs (e.g. property damage, police time and insurance administration).

The program contains a model for allocating traffic to the diversion route if the site becomes overloaded, representing both the road users that queue through the site and those that take an alternate route in the case of a partial closure. The details of QUADRO, including all assumptions made in its calculations, are provided in the manual (Anon. 2015).

In order to model a road closure in QUADRO, a diversionary route must be defined. The QDIV (QUADRO Diversion) tool was used to model the standard diversionary routes used by the road operator.

QDIV requires each diversionary route to be defined in terms of a set of links (each defined as rural, urban, sub-urban or small town) that can be combined in series and parallel to build up a network. For each event, a simplified diversionary network schematic was developed and Google Maps was used to measure the length of each link (Winter et al. 2018). Traffic data, represented as annual average daily traffic (AADT), were sourced using data from the relevant Road Administrations.

Where information was not available (e.g. lane and verge widths), the default values suggested in the QUADRO manual were adopted. Classified traffic counts (i.e. split into different vehicle types), and therefore the proportion of heavy vehicles, were only

Table 1 Site parameters for the direct consequential economic impact analysis.

Event	Number vehicles damaged	Traffic (AADT) (vehicles/day)[2]	HGVs (%)	Junction length (km)	Closure type(s)	Closure duration
August 2004: A83 Glen Kinglas to Cairndow (L)[1]	1	5,554	9	20	Full closure	2 days
August 2004: A9 N of Dunkeld (L)	5	13,864	18	18	Full closure, then shuttle working with convoy	2 days full 6 days convoy
August 2004: A85 Glen Ogle (L)	3	4,403	10	26	Full closure	4 days
October 2007: A83 Rest and be Thankful (L)	1	5,748	10	20	Full closure, then shuttle working[3]	15 days full 27 days shuttle[4]
September 2012: A77, A76, A71 Bellfield Junction (F)	0	6,400 to 13,800	6	0.3 to 4.9	A77: Full closure A76: Full closure then shuttle working A71 Full closure	25 hours 19 hours full 2 days shuttle 67 hours
October 2014: A83 Rest and be Thankful (L)	0	5,460	9.5	20 / 4[5]	Full closure (daytime) Full closure (night time) Convoy (daytime)[6] Convoy (night time)[6] Shuttle (daytime) Shuttle (night time) Shuttle (24 hours)	10 hours 28 hours 42 hours 56 hours 40 hours 42 hours 682 hours

[1] L = Landslide; F = Flood.
[2] Peak monthly figures, usually for August.
[3] Single-lane working with traffic light control.
[4] This figure represents the duration of the closure due to the instability and the immediate engineering works required to allow the reopening of the road. It is acknowledged that the road was subsequently subject to single lane working with traffic light control for a significantly longer period due to engineering works necessitated by the combination of this and subsequent events in the immediate vicinity.
[5] 20 km for full closure, 4 km for convoy working on the old Military Road (OMR).
[6] Using the OMR, the temporary diversion used when the A83 Rest and be Thankful road is closed.

available for some links; either the proportion from the closest link or a nominal 10% HGVs was assumed (Table 1).

It was assumed that all of the roads affected were rural all-purpose single carriageways with a speed limit of 96 km/h (60 mph), reduced to 48 km/h (30 mph) where part of the road remained open following the landslide or flood event, and that the length of the affected site in each case was 100 m. (The speed limit at the A85 site was reduced from the national speed limit of 96 km/h (60 mph) to 80 km/h (50 mph) in November 2015, more than a decade after the events and around the time that aspect of the study was completed.)

3.3 Indirect Consequential Economic Impacts

There is a wide range of possible approaches to estimating the indirect consequential economic impacts of landslides. These include:

- cost-benefit analysis,
- cost-effectiveness analysis,
- willingness to pay,
- multi-criteria analysis,
- methods based upon Transport Appraisal.

In addition, there are bespoke methods designed to address a particular set of circumstances (McLeod et al. 2005; Anon. 2013b) as described by Winter et al. (2018).

Surveys of businesses were undertaken in the areas of the 2004 A85 landslide; 2012 A77, A76, A71 Bellfield Junction flood; and A83 2014 landslide events. The surveys used questionnaires based on the Stated Preference approach that were mailed to respondents with follow up telephone calls to improve the response rate that varied between 20.8% (A5 Glen Ogle) 17% (A83 Rest and be Thankful) and 11.7% (A77, A76, A71 Bellfield Junction).

4. Results

4.1 Direct Economic Impacts

The available data is reported in Table 2, adjusted to 2012 prices.

Direct economic impacts include:

1. The direct costs of clean-up and the costs of search and rescue.
2. The repair/replacement of lost/damaged infrastructure in the broadest sense.

These might otherwise be described as 'emergency response' and 'remedial works', respectively as in Table 2.

Direct economic costs for the landside events range between approximately £250k and £1,700k. For the flood event in a more developed peri-urban part of Scotland the direct costs were relatively small (around £25k).

Table 2 Direct economic impacts (at 2012 prices).

Event	Emergency response	Remedial works	Total
August 2004: A83 Glen Kinglas to Cairndow	£395,043		£395,043
August 2004: A9 N of Dunkeld	£921,766		£921,766
August 2004: A85 Glen Ogle	£658,405		£658,405
October 2007: A83 Rest and be Thankful	£320,772	£1,372,629[1]	£1,693,401
September 2012: A77, A76, A71 Bellfield Junction	£16,756	£8,333	£25,088
October 2014: A83 Rest and be Thankful	£245,328	£0[2]	£245,328

[1] Comprises: debris barrier (including design, tender, award, construction and supervision) £425,446 (at 2012 prices); works prior to culvert replacement (including ongoing traffic management and carriageway protection, culvert design, geotechnical design and certification, ground investigation and diversion fibre optic telecommunications cable) £181,184; and culvert replacement (construction) £765,999. These works were undertaken in direct response to the October 2007 event and may thus be attributed to this event.

[2] While an extensive programme of remedial works has been undertaken at the wider A83 Rest and be Thankful site, some of which is focused on the location of the 28 October 2014 event, this was installed prior to the event and thus cannot be attributed to this event. In broad terms the remedial measures, including both management and mitigation measures (Winter 2014), worked as anticipated and would not have been expected to prevent a debris flow of this size from reaching the road.

4.2 Direct Consequential Economic Impacts

QUADRO enables the calculation of the costs of user delays and diversions, carbon emissions from vehicles and accidents associated with the road works, reporting the costs on the basis of an average day over a whole week (Table 3); the total costs for each site, taking into account the duration of the impacts, are summarised in Table 4. Implicit are assumptions regarding the costs of time (vehicle occupancy, journey purpose, and the value of time for both occupants and vehicles), vehicle operating costs (and associated carbon costs), and the value of accidents that occur within the section(s) of road under consideration (Anon. 2015) and the associated values are based on national statistics.

Careful consideration of the relative traffic levels, and closure type and duration (Table 4), reveals patterns that are broadly consistent with those that might be inferred intuitively, as follows:

- The costs of similar closures depend on traffic levels; costs being in proportion to traffic (A9 *cf.* A83 2004).
- Doubling the duration incurs higher costs, but may be reduced if the traffic levels are lower (A83 2004 *cf.* A85).
- A much longer duration increases the costs significantly (A83 2007).

Table 3 Incident accident costs (per vehicle) and QUADRO daily closure costs (at 2012 prices).

Cost (£)	August 2004: A83 Glen Kinglas to Cairndow	August 2004: A9 N of Dunkeld (Full closure/ shuttle working)	August 2004: A85 Glen Ogle	October 2007: A83 Rest and be Thankful (Full closure/ shuttle working)	September 2012: A77, A76, A71 Bellfield Junction (Full closure/ shuttle working)	October 2014: A83 Rest and be Thankful1 (Full closure/ convoy/ shuttle working)
Accident incident cost	2,520	2,520	2,520	2,520	2,520	2,250
Delay cost (daily)	84,071	270,885 / 135,339	71,679	88,040 / 461	1,548,624 / 94,363	168 to 83,427
Carbon cost (daily)	6,380	18,608 / 9,304	6,629	6,590 / 6	73,922 / 671	2 to 6,262
Accident cost (daily)	−4,360	−11,254 / −5,627	−4,494	−4,512 / 794	−38,568 / 7,564	−4,268 to 312

[1] The daily delay, carbon and accident costs for the October 2014 Rest and be Thankful event vary significantly, by several orders of magnitude, for the various types of closure described in Table 2 and accordingly only the range is reported here. Full details are given by Winter et al. (2018). The higher costs (lower daily accident costs) are for full closure and the lowest costs (highest daily accident costs) are for shuttle working.

Table 4 Total incident accident costs and QUADRO total closure costs (at 2012 prices).

Cost (£)	August 2004: A83 Glen Kinglas to Cairndow	August 2004: A9 N of Dunkeld (Full closure/ shuttle working)	August 2004: A85 Glen Ogle	October 2007: A83 Rest and be Thankful (Full closure/ shuttle working)	September 2012: A77, A76, A71 Bellfield Junction (Full closure/ shuttle working)	October 2014: A83 Rest and be Thankful (Full closure/ convoy/shuttle working)
Accident incident cost	2,520	12,600	7,560	2,520	0	0
Delay cost	168,143	1,218,460	286,718	1,333,020	3,080,542	173,956
Carbon cost	12,762	83,737	26,514	99,029	151,669	9,164
Accident cost	−8,721	−45,288	−17,974	−46,247	−71,309	29,466
Total	**174,703**	**1,269,508**	**302,817**	**1,388,322**	**3,160,902**	**212,587**

Fig. 6 Word map of responses from survey respondents: A85 Glen Ogle, 18 August 2004 landslide.

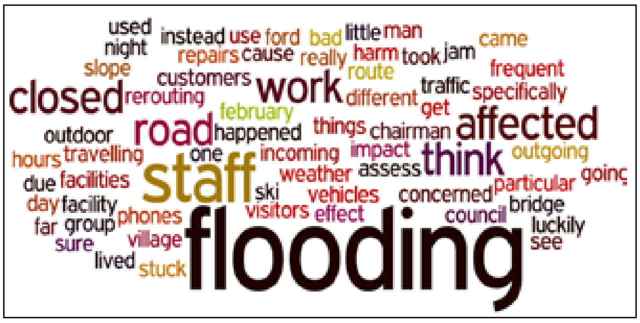

Fig. 7 Word map of responses from survey respondents: A77, A76, A71 Bellfield Junction Flooding: 21 September 2012.

Of particular interest are the negative costs (i.e. cost reductions) for traffic accidents during post-event diversions and/or restricted traffic flow. These reduced accident costs suggest a decrease in accident numbers and/or accident severity and seem most likely to be as a result of reduced traffic speeds leading to an increased opportunity to avoid accidents and lower severity when they do occur.

The landslide events were located in rural areas and their impacts are upon those areas and small towns and villages. Direct consequential costs range between approximately £180k and £1,400k for the landslide events. The latter costs are largely dependent upon the amount of traffic that uses the road and the duration of the disruption.

The flood event was located in a much more developed part of Scotland and on the edge of a town (Kilmarnock). The peri-urban flood location and much shorter event duration, places a different complexion on the direct consequential economic impacts which were more than twice those for the A83 2007 (c. £3,200k).

Notwithstanding this the impacts of the A83 event(s) should not be underestimated: those impacts were borne by a much smaller number of people over an extended period; the impacts on individuals and individual businesses seem likely to have been considerably greater. This part of the analysis also does not take account of the longer term indirect consequential economic impacts (see Section 4.3).

4.3 Indirect Consequential Economic Impacts

The surveys of businesses in the areas of these events provided cost information that could be interpreted in a number of ways and therefore gave a very wide range of potential results. The results did, however, provide useful qualitative information (Winter et al. 2018). For events of lesser impact, descriptors that relate to the hazard are used: 'landslide', 'flooding' and other words that describe the event itself are also to the fore (Figures 6 and 7).

In contrast responses to events of greater impact and or repetition such as at the A83 (Figure 8), at which a significant number of events and consequent closures have occurred over the past 20 years, tend to relate to the effects, risks, or impacts, that derive from the event. In this case the most frequently used word is 'road', with words such as 'closed', 'staff', 'visitors', 'due', 'access', 'tourism', 'minor' and 'island' also coming to the fore. These latter responses seemingly describe the consequences of the hazard, or the economic risks associated with the hazard, rather than the hazard itself, implying a greater economic impact or, at least, a greater awareness of the economic impact.

Fig. 8 Word map of responses from survey respondents: A83 Rest and be Thankful, 28 October 2014 landslide.

5. Vulnerability shadow

The vulnerability shadow for the October 2007 debris flow is described in Section 2 (Figure 4) and is estimated at around 2,800 km^2. This description holds for the event in 2014 also as the road closure was on the same link between the junctions with the A814 at Arrochar and the B828 immediately to the north of the Rest and be Thankful car park with no entry or exit routes to the A83 between.

While the 2007 and 2014 A83 debris flow events occurred to the east of the B828 (serving Lochgoilhead), the August 2004 events at the A83 were located to the west of the B828 in Glen Kinglas and further to the west, beyond the A815 (serving inter alia Dunoon) to the west of Cairndow (around 5 km to the north-west of the 2007/2014 events shown in Figure 4). Notwithstanding this, the differences in the diversions for the two sets of events were subtle and it is broadly considered that the extent of the vulnerability shadow was not dissimilar.

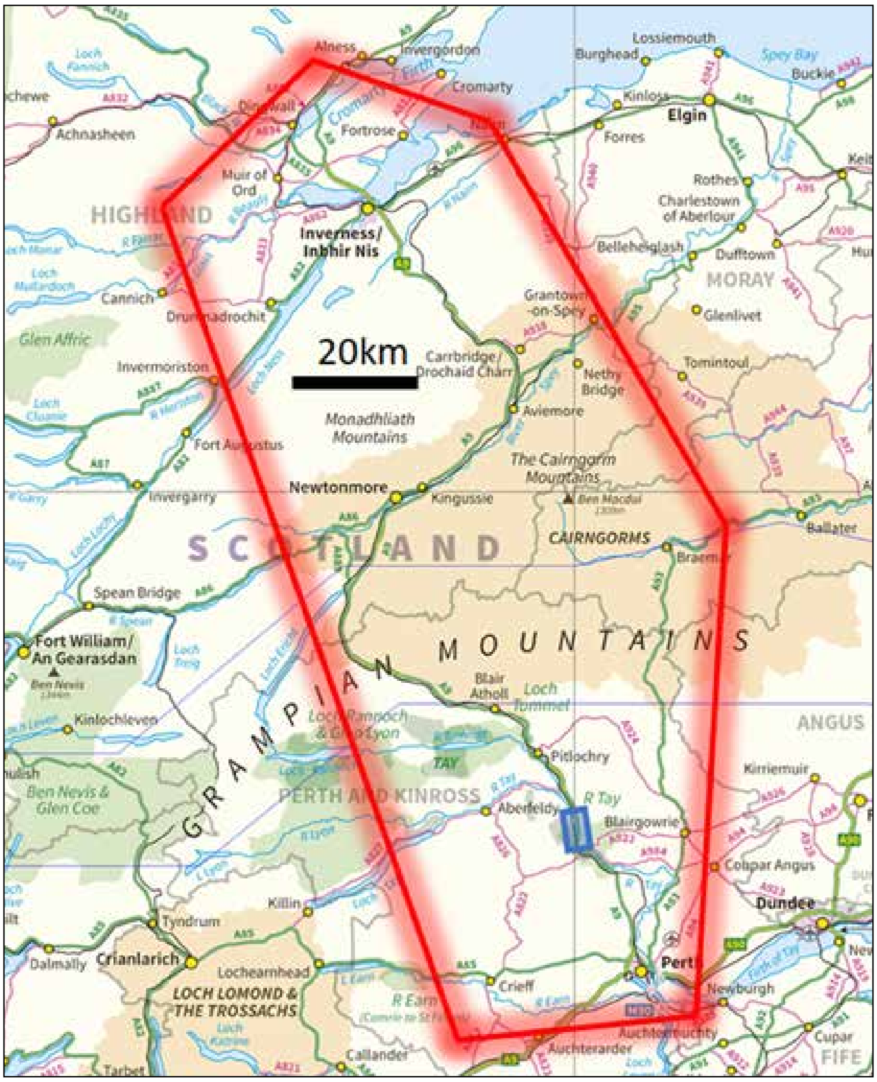

Fig. 9 Three main debris flow events (blue rectangle) closed the A9 N of Dunkeld on 11 August 2004; the vulnerability shadow that was cast (bounded in red) was extensive and reflects the importance of the A9, particularly to the communities to the north. See Figure 1 (2) for the event location in the wider geographical context. (Image based on OS Route Planner 2016 Map. © Crown Copyright. All rights reserved Scottish Government 100020540, 2019.)

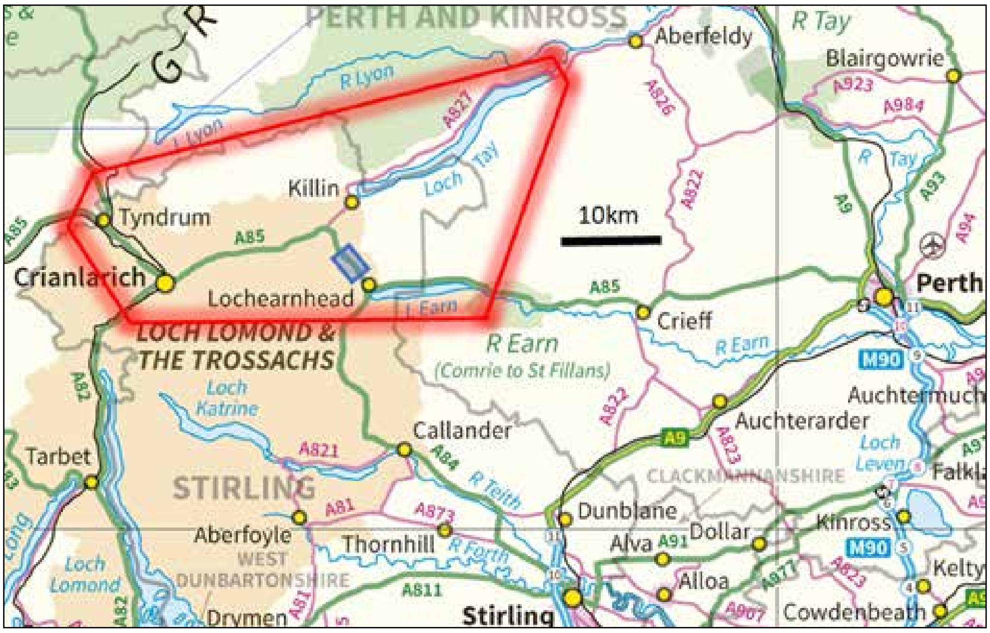

Fig. 10 Two debris flow events (blue rectangle) closed the A85 in Glen Ogle on 18 August 2004; the vulnerability shadow that was cast (bounded in red) was limited by the reasonably good range of alternative routes in the area. See Figure 1 (3) for the event location in the wider geographical context. (Image based on OS Route Planner 2016 Map. © Crown Copyright. All rights reserved Scottish Government 100020540, 2019.)

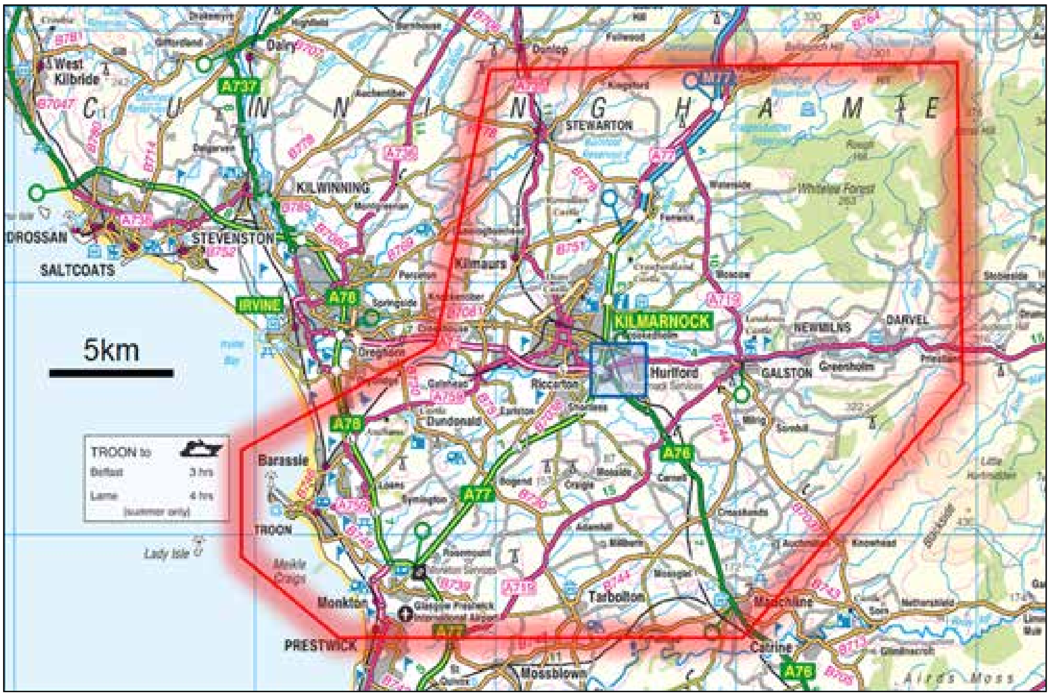

Fig. 11 Flooding at the A77, A76, A71 Bellfield Junction (blue rectangle) closed all three roads on 21 September 2012; the area of the vulnerability shadow that was cast (bounded in red) was limited by the availability of diversionary routes. See Figure 1 (4) for the event location in the wider geographical context. (Image based on OS 1:250,000 mapping. © Crown Copyright. All rights reserved Scottish Government 100020540, 2019.)

Those for the other 2004 events were estimated at around 8,000 km^2 (A9: Figure 9) and 800 km^2 (A85: Figure 10). The vulnerability shadow cast for the A9 debris flow events to the north of Dunkeld (11 August 2004) is significant (Figure 9), reflecting the importance of the A9 as the primary north-south route in Scotland, leading to the most northerly city of Inverness (population around 47,000) and numerous smaller communities along the route and to the north of Inverness, and the relative paucity of alternative routes and their commensurate length. The A85 Glen Ogle vulnerability shadow cast for the 18 August 2004 debris flow is, by comparison, relatively small (Figure 10). This, in turn, reflects the relatively central position of that area in Scotland and the existence and relative efficiency of alternate routes through and around the area.

The vulnerability shadow cast by the 21 September 2012 flooding event at the A77, A76, A71 Bellfield Junction (Figure 11) was smaller still at around 500 km^2 (the nearby town of Kilmarnock has a population of approximately 46,000). The road network in this area is much denser, and the population density is much higher, than that in the other areas studied and this leads to reasonable alternative routes, albeit with significant associated disruption to traffic.

6. Discussion

Highland (2006) (see also Schuster 1996; Schuster, Highland 2007; Highland 2012) concluded that records of landslides and their associated costs were inadequate and that a standardised approach to landslide loss inventory would reduce the cost, and improve the usability and availability of landslide loss information. Further work by Highland (2012) deals in more detail with both direct economic impacts and also some aspects of direct consequential and indirect consequential economic impacts, although these latter two are usually described in a qualitative manner. Highland describes decreased economic activity in some areas and increased economic activity in other areas (e.g. as a result of the Glenwood Canyon Rockfall) and both higher than usual flight costs and an increase in the incidence of charter aeroplane use between key locations on either side of the event.

A related study (Klose et al. 2015) collected data and modelled the costs incurred due to landslides that affected the federal road network in the Lower Saxon Uplands in NW Germany. The approach used the data collected at a local level to extrapolate direct costs for the region, using the results of a susceptibility assessment and an infrastructure exposure model. However, this study deals only with direct economic impacts and this degree of extrapolation across the network was not considered appropriate to the situation in Scotland where landslides occur relatively infrequently.

Hearn et al. (2008) deal with both direct and direct consequential economic impacts, but their approach to direct consequential impacts appears to be based on the assumption that all vehicles (and drivers and passengers) that would normally use the road for the period of the closure will wait for it to be reopened. While this might be appropriate for Laos, where the study was conducted, and reflects the morphology of the road network it does not account for either diversions or restrictions of traffic flow as are more normal in Europe.

The lack of a robust database of landslide events makes a quantitative national economic analysis difficult. Almost 40 landslides are reported in the media in the UK annually (38 in 2010) (Gibson et al. 2012). Anon. (2013a) noted that between 1 January 2007 and 31 October 2012, the A83 at the Rest and Be Thankful was closed five times as a result of landslides and these and other landslides have been enumerated and described (e.g. Winter et al. 2005, 2006, 2009) extensively in recent years. Despite this Dobbie et al. (2011) consider that landslides are rare in Scotland. Whatever the rarity or otherwise of such events, better data is critical to the effective management of risks, as highlighted for flood hazards and risks by the Association of British Insurers (Anon. 2010).

Importantly, this current work confirms Highland's (2006) assertion that past data for direct economic impacts are generally labour intensive to retrieve. The experience here has been that as people move on both knowledge and experience are lost but, even more critically, as contracts pass to new organizations data and information about events is lost.

The modelling of direct consequential economic impacts has not been attempted before and this throws up some interesting but rather unsurprising conclusions – not least that the total costs depend primarily on the amount of traffic affected and the duration for which it is affected. Notwithstanding this, it is important to recognize that where traffic levels are lower the network may represent a lifeline route with limited, complex/lengthy and/or no alternative routes.

Similarly, indirect consequential economic impacts are rarely evaluated. The surveys conducted for this work provided important and very useful qualitative data as set-out in Section 4.3 and represented in Figures 6,7 and 8. Indeed, some of the comments from the A85 Glen Ogle survey confirmed Highland's (2006) assertion that economic impacts need not only be negative, that some can be positive:

– "More people rented bikes to go and visit the landslide site. The landslide itself had little impact, it was the associated bad weather that was the problem - the main street was flooded" according to one retailer.
– Hoteliers also seem to have benefited as "[they were able to] put up people in the guest house who had been trapped in the glen (and who had been airlifted out) free of charge. The same happened in different places around Killin," which apparently "improved the image of the town."

Notwithstanding this an assessment that included indirect consequential impacts was reported as part of a route study of the A83 (Anon. 2013a) returning a figure of £286k per landslide at the A83 Rest and be Thankful. While this provides useful context, it is not entirely clear how much of that total would be accounted for by direct consequential impacts that were also included or for how long the impacts were assumed to endure.

The vulnerability shadow proved to be a useful tool to understand and articulate the extent of the socio-economic vulnerabilities and, indeed, the populations potentially affected.

The vulnerability shadows determined from the various events demonstrate that their extent is determined by the density of the network, and the linked availability and suitability of alternative routes. Clearly there is also a strong correlation between these factors and the population served by the routes and affected by the events; as the density of the network, and availability and suitability of alternative routes reduces so does the population served. This must of course be set against the fact that the lower density networks are often lifeline routes with few, if any, alternatives and their importance to a smaller population is amplified. The authors are not aware of this or a similar tool being used in other studies.

7. Summary and conclusions

This paper presents the results of a study to develop methods of obtaining data on the economic impacts of landslides and the associated extent of those impacts. The economic impacts of landslides are considered in three categories: direct economic impacts, direct consequential economic impacts, and indirect consequential economic impacts. This approach is also applicable to events that reflect relatively discrete closures including climate-driven flooding.

The work presented herein includes data for five Scottish landslide events that occurred between 2004 and 2014. Direct costs range between approximately £250k and £1,700k for the landslide event, while those costs for the flood event are relatively small. Direct consequential costs range between around £180k and £1,400k for the landslide events. The latter are largely dependent upon the amount of traffic that uses the road and the duration of the disruption. For a flood event in a more developed peri-urban part of Scotland although the direct costs were small but the direct consequential costs (c. £3,200k) much greater than for any of the landslide sites considered. It is also worth noting that flood event was of a relatively short duration compared to the high cost landslide event.

Work on indirect consequential impacts has provided valuable qualitative insights although meaningful quantitative data proved rather elusive.

The vulnerability shadow proved to be a useful tool to understand and articulate the extent of the socio-economic vulnerabilities and the populations potentially affected. They also aid in understanding and communicating the inter-relations between the density of the network, the linked availability and suitability of alternative routes.

Acknowledgements

Transport Scotland's funding of this work is gratefully acknowledged.

This paper is published with the permission of Transport Scotland and TRL Limited.

References

Anon. (2010): Fighting flood risk together. Association of British Insurers, London.

Anon. (2011): Scottish road network climate change study: UKCP09 update. Transport Scotland, Edinburgh. (Accessed February 2014: http://www.transportscotland.gov.uk/.)

Anon. (2012a): Scottish Transport Appraisal Guidance. Transport Scotland, Edinburgh. (Accessed February 2014: http://www.transportscotland.gov.uk/stag/home.)

Anon. (2012b): TAG UNIT 3.3.5: The Greenhouse Gases Sub-Objective, August. Department for Transport, London. (Accessed February 2014: http://www.dft.gov.uk/webtag.)

Anon. (2013a): A83 Trunk Road Route Study: Part A – A83 Rest and be Thankful. Final Report. Report prepared by Jacobs for Transport Scotland, 212p. (Accessed February 2014: http://www.transportscotland.gov.uk/road/maintenance/landslides.)

Anon. (2013b): Reported road casualties Scotland 2012. Transport Scotland, Edinburgh. (Accessed September 2015: www.transportscotland.gov.uk/statistics/reported-road-casualties-scotland-all-editions.)

Anon. (2015): The QUADRO 4 Manual, Part 2: The valuation of costs in QUADRO. (Accessed September 2015: http://tamesoftware.co.uk/manuals/manuals.html/.)

Bono F., Gutiérrez E. (2011): A network-based analysis of the impact of structural damage on urban accessibility following a disaster: the case of the seismically damaged Port Au Prince and Carrefour urban road networks. Journal of Transport Geography 19, 1443–1455, https://doi.org/10.1016/j.jtrangeo.2011.08.002.

Bíl, M., Vodák, R., Kubeček, J., Bílová, M., Sedoník, J. (2015): Evaluating road network damage caused by natural disasters in the Czech Republic between 1997 and 2010. Transportation Research Part A: Policy and Practice 80, 90–103, https://doi.org/10.1016/j.tra.2015.07.006.

Chang, S.E., Nojima, N. (2001): Measuring post-disaster transportation system performance: the 1995 Kobe earthquake in comparative perspective. Transportation Research Part A: Policy and Practice 35, 475–494, https://doi.org/10.1016/S0965-8564(00)00003-3.

Dahal, R. K., Bhandary, N. P. (2013): Excursion guidebook for Pokhara Valley area. Unpublished.

Dobbie, K. E., Bruneau, P. M. C., Towers, W. (Editors) (2011): The state of Scotland's soil. Natural Scotland, Scottish Government, Edinburgh. (Accessed December 2012: www.sepa.org.uk/land/land_publications.aspx.)

Galbraith, R. M., Price, D. J., Shackman, L. (Eds.) (2005): Scottish road network climate change study. Scottish Executive, Edinburgh.

Gibson, A. D., Culshaw, M. G., Dashwood, C., Pennington, C. V. L. (2012): Landslide management in the UK the problem of managing hazards in a 'low-risk' environment. Landslides 10(5), 599–610, https://doi.org/10.1007/s10346-012-0346-4.

Hearn, G., Hunt, T., Aubert, J., Howell, J. (2008): Landslide impacts on the road network of Lao PDR and the feasibility of implementing a slope management programme. Proceedings, International Conference on Management of Landslide Hazard in the Asia-Pacific Region (Ed: Chigira, M.), 187-195. The Japan Landslide Society, Tokyo.

Highland, L. M. (2006): Estimating landslide losses – preliminary results of a seven-state pilot project. US Geological Survey Open File Report 2006–1032, USGS, Reston, VA, https://doi.org/10.3133/ofr20061032.

Highland, L. M. (2012): Landslides in Colorado, USA: impacts and loss estimation for the year 2010. US Geological Survey Open File Report 2012–1204, USGS, Reston, VA, https://doi.org/10.3133/ofr20121204.

Klose, M., Damn, B., Terhorst, B. (2015): Landslide cost modelling for transportation infrastructures: a methodological approach, Landslides, 12, 321–334, https://doi.org/10.1007/s10346-014-0481-1.

MacLeod, A., Hofmeister, R. J. Wang, Y., Burns, S. (2005): Landslide indirect losses: methods and case studies from Oregon, State of Oregon, Department of Geology and Mineral Industries Open File Report O-05-X, Portland, OR.

Petley, D. N. (2014): The Seti River debris flow in Nepal – what was the role of the smaller landslide downstream? (Accessed February 2014: blogs.agu.org/landslideblog/2014/02/07/seti-river/.)

Petley, D. N., Stark, C. (2012): Understanding the Seti River landslide in Nepal. (Accessed February 2014: blogs.agu.org/landslideblog/2012/05/23/understanding-the-seti-river-landslide-in-nepal/.)

Schuster, R. L. (1996): Socioeconomic significance of landslides, Landslides – Investigation and Mitigation: Transportation Research Board Special Report 247, 36–75, Washington, DC.

Schuster, R. L., Highland, L. M. (2007): The Third Hans Cloos Lecture, Urban landslides: socioeconomic impacts and overview of mitigative strategies. Bulletin of Engineering Geology and the Environment 66, 1–27, https://doi.org/10.1007/s10064-006-0080-z.

Winter, M. G. (2014a): A strategic approach to landslide risk reduction. International Journal of Landslide and Environment 2, 14–23.

Winter, M. G. (2014b): The vulnerability shadow cast by debris flow events, Engineering Geology for Society and Territory, Volume 6: Applied Geology for Major Engineering Works (Eds: Lollino, G., Giordan, D., Thuro, L., Carranza-Torres, C., Wu, F., Marinos, P. Delgado, C.), 641–644, Heidelberg: Springer.

Winter, M. G. (2016): A strategic approach to debris flow risk reduction on the road network. Procedia

Engineering 143, 759–768, https://doi.org/10.1016/j.proeng.2016.06.121.

Winter, M. G., Bromhead, E. N. (2012): Landslide risk: some issues that determine societal acceptance. Natural Hazards 62, 169–187, https://doi.org/10.1007/s11069-011-9987-1.

Winter, M. G., Corby, A. (2012): A83 Rest and be Thankful: ecological and related landslide mitigation options. Published Project Report PPR 636, Transport Research Laboratory, Wokingham.

Winter, M. G., Shearer, B. (2013): Climate change and landslide hazard and risk – a Scottish perspective. Published Project Report PPR 650, Transport Research Laboratory, Wokingham.

Winter, M. G., Macgregor, F., Shackman, L. (Eds.) (2005): Scottish Road Network Landslides Study. The Scottish Executive, Edinburgh.

Winter, M. G., Heald, A., Parsons, J., Shackman, L., Macgregor, F. (2006): Scottish debris flow events of August 2004. Quarterly Journal of Engineering Geology and Hydrogeology 39, 73–78, https://doi.org/10.1144/1470-9236/05-049.

Winter, M. G., Macgregor, F., Shackman L. (Eds.) (2009): Scottish road network landslides study: implementation. Transport Scotland, Edinburgh.

Winter, M. G., Dent, J., Macgregor, F., Dempsey, P., Motion, A., Shackman, L. (2010): Debris flow, rainfall and climate change in Scotland. Quarterly Journal of Engineering Geology and Hydrogeology 43, 429–446, https://doi.org/10.1144/1470-9236/08-108.

Winter, M. G., Harrison, M., Macgregor, F., Shackman, L. (2013): Landslide hazard assessment and ranking on the Scottish road network. Proceedings, Institution of Civil Engineers (Geotechnical Engineering) 166(GE6), 522–539, https://doi.org/10.1680/geng.12.00063.

Winter, M. G., Shearer, B., Palmer, D., Peeling, D., Peeling, J., Harmer, C., Sharpe, J. (2018): Assessment of the economic impacts of landslides and other climate-driven events. Published Project Report PPR, TRL, Wokingham.

Winter, M. G., Ognissanto, F., Martin, L. A. (2019): Rainfall thresholds for landslides: deterministic and probabilistic approaches. Published Project Report PPR 901, Transport Research Laboratory, Wokingham.

European imbalances and shifts of global value chains to the Central European periphery: role of institutions

Pavel Hnát, Ondřej Sankot*

University of Economics, Prague, Faculty of International Relations, Department of World Economy, Czechia
* Corresponding author: ondrej.sankot@vse.cz

ABSTRACT

This article deals with the topic of European imbalances. They are defined as large and persistent differences in the current account position of European countries, which are closely connected to the emergence of the financial crisis and the subsequent sovereign debt crisis in 2008. A build-up in current account deficits had been observed from the mid-1990s, namely in two peripheral regions of the EU. However, little attention was paid to the potential differences between the Southern and Central European peripheries of the EU. The emergence of large and persistent current account deficits in Southern Europe was accompanied by a significant shift in gains from global value chains. The aim of this paper is to evaluate the factors that co-determined the changes in the geographic structure of GVCs in Europe. These changes decreased GVC income in Southern Europe, increased it in Central Europe and contributed to the build-up of account imbalances in Southern Europe. Despite the fact that Central Europe was among the deficit regions in European imbalances, the four Central European countries substantially increased their gains from global value chains as well as GVC participation. The shift in GVC activity towards Central Europe between 1995 and 2011 was driven not only by total labour costs but also by better regulatory quality. At the same time, TNCs switching from Southern to Central Europe had to accept worse quality contract enforcement.

Keywords
global value chains; institutions; European imbalances; Central Europe

Received: 9 April 2019
Accepted: 13 November 2019
Published online: 19 December 2019

Hnát, P., Sankot, O. (2019): European imbalances and shifts of global value chains to the Central European periphery: role of institutions. AUC Geographica 54(2), 221–231
https://doi.org/10.14712/23361980.2019.19

1. Introduction

The financial crisis in 2008 and subsequent sovereign debt crisis revealed the underlying structural weaknesses of the European integration project. These weaknesses require major action and increased coordination in economic policies. With most attention being paid to structural differences within the euro area, the political agenda has also slowly addressed the significant structural differences between Europe's East and West. The legacy of the Eastern enlargement has thus increasingly come under scrutiny. Moreover, its policy implications are very carefully studied, namely with respect to the labour market and migration. During the build-up to the financial crisis in Europe (data will be examined from 1995 to 2011 in this paper, referring to the early years of the build-up and its escalation on the eve of the financial crisis and initial years of the crisis itself), the problem of Europe's Eastern periphery was not obvious or was neglected. Policy debate as well as theoretical explanations of why the crisis hit Europe so hard mostly followed structural issues within the euro area. They left Central European countries aside and paid the greatest attention to Southern Europe, i.e. Greece, Italy, Portugal and Spain. From today's perspective it is, however, obvious that there are also significant policy challenges arising from Central European countries' trade and labour market integration. Since trade integration has increased significantly in Central Europe, we have decided to focus on the role of changing global value chain (GVC) participation obvious in Central Europe during the build-up of European imbalances. European imbalances, which were mostly seen as an interaction between Europe's competitive North and structurally weaker South, also have a significant East-West component.

Structural changes of GVCs in Europe can be demonstrated by changes in German outward FDIs. In 1991, 12% of German investments in developed OECD countries went to Southern Europe. By 2007, this share had fallen to 7%. Simultaneously, German investment stock in Czechia, Hungary, Poland and Slovakia reached the same level as the stock in the whole of Southern Europe, despite the fact that the overall GDP of those Central European countries in 2007 represented only 20% of Southern Europe's (OECDstat 2018). There has been a gradual GVC shift from Europe's South to Europe's East which are both current account deficit regions from the European imbalances perspective. The aim of this paper is to evaluate the factors that co-determined the changes in the geographic structure of GVCs in Europe which reduced GVC income in Southern Europe, increased it in Central Europe and contributed to the build-up of account imbalances in Southern Europe. In this paper, Southern Europe (the Southern European periphery) will be represented by Greece, Italy, Portugal and Spain, while Central Europe (the Central European periphery) by Czechia, Hungary, Poland and Slovakia.

European imbalances, which have been studied in line with global imbalances (see Dunaway 2009; ECB 2008; or O'Brien, Williams 2007 for example), are defined as large and persistent differences in current account positions of different European countries (or economic models). These imbalances underpinned the build-up to the financial crisis of 2008 in Europe. OECD (2010) data reveal a current account balance of between −5% and 7% of national GDP in 1995 and of between −14% and 8% of GDP in 2008. Additionally, current account position only underpinned changing debtor-creditor relations with Greece, Portugal and Spain cumulating more than 70% of GDP in international liabilities by 2008, while major current account surplus countries became major creditors. Growing current account surpluses in the North are mostly attributed to the creation of the euro, which made the core countries' investment rise in the euro area peripheries (IMF 2009). The prevailing rigidities of Europe's core goods and labour markets, which limited investment opportunities (Dunaway 2009), also contribute to the imbalances. The peripheries' deficits were mostly explained by declining competitiveness in the new economic climate caused by the adoption of a common currency (Wyplosz 2010). Falling competitiveness was fuelled by growing differences in wages and productivity between the periphery and the core (IMF 2009) as well as by the inappropriate fiscal policy response to greater financial market integration and low interest rates (Jiránková and Hnát 2012 or Šíma 2016). In general, the greatest attention was paid to Southern Europe and its growing wages accompanied by limited productivity growth and increased borrowing opportunities from current account surplus parts of the euro area. This fuelled sovereign debt and increased the expansionary effects of the common monetary policy (Wyplosz 2010; IMF 2009).

Little or no attention was paid to the fact that after the Eastern Enlargement, the EU absorbed another periphery at its Eastern border. Even though its persistent current account deficits clearly made it part of the core-periphery issue in the EU, the very different structure of Central European countries' current accounts was not seen as an important aspect for further analysis. Yet, the structure of the most trade-integrated parts of Central Europe shows striking differences from the predominant trend in Southern Europe's balance of payments. A positive trend in trade and services balances, which is more than offset by ever larger profit repatriations, reveals an important trend in the market. The GVC integration of Central European countries demonstrates a marked difference from the structural and productivity issues of Southern Europe.

A similar flaw can be identified in the growing literature on global value chains, which is becoming an increasingly popular focus for international trade analysis, particularly after Baldwin's Second Unbundling. Such works are largely dedicated to the most developed economies or to the development effects of trade and market integration and do not pay enough attention to the case of the (relatively) well-developed transition economies in Central Europe. Some literature studying the case of Central Europe has emerged, but most of it focuses on the automotive industry: Pavlínek 2004, 2018, 2019; Pavlínek and Ženka 2011; Humphrey et al. 2000; Sturgeon et al. 2008. Other sectors have only been studied selectively, like Smith et al. (2014) for the apparel industry in Slovakia, or Plank and Staritz (2013); Sass and Szalavets (2014) for the electronics industry in Hungary. For a comprehensive study on industrial upgrading in the region, see Vlčková et al. (2015).

Guzik and Micek (2008) compiled country-industry case studies from the region. These studies offer valuable data which supports our hypothesis that there has been a major structural shift between the two current account peripheries resulting in significant consequences for today's Europe.

Given the generally high proportion of the EU's share in Central Europe's trade and investment, a significant increase in the region's GVC participation must have had effects in Europe's other periphery, i.e. Southern Europe, which has so far benefitted from European trade integration. Changing gains from GVCs and current account position between the two peripheries indicate a more important structural shift in EU trade integration, this has mostly been neglected due to a predominant focus on the euro and its institutional weaknesses. To reveal such a structural shift, our analysis will connect economic growth with trade integration in GVCs to show which growth factors influenced the changing GVC patterns the most.

2. Literature review

Offshoring and internationalisation of production, which was related to the technological progress and overall liberalisation of the world economy, contributed to the so-called slicing up of the value chain (Dicken 2015). This has been described as global supply/commodity chains (Gereffi 1994), global value chains (Porter 1985) or global production networks (Coe et al. 2008). In this paper, we use GVC income data from Timmer et al. (2013) and we therefore follow the concept of GVCs. Value creation in GVCs is crucial for economic growth (Coe et al. 2008). As the costs and benefits of trade are distributed unevenly (Baldwin 2006), the objective of countries is to improve their positions in GVCs (Henderson et al. 2002; Humphrey and Schmitz 2002), which makes GVCs a dynamic system.

Since the beginning of its economic transition, Central Europe has consistently attracted significant amounts of inward FDIs, creating a specific system dependent on manufacturing exports through producer-driven networks (Myant and Drahokoupil 2012). International investors mainly utilised a relatively skilled and cheap labour force, the geographic location, investment subsidies, integration into EU structures and relative political stability (Pavlínek et al. 2009; Pavlínek and Ženka 2016; Myant and Drahokoupil 2012). Central European suppliers became increasingly integrated into European value chains, as proven by Domanski and Gwosdz (2009), Jürgens and Krzywdzinski (2009). On the other hand, being dependent on FDIs, Central European countries are prone to potential value transfer to the economies from which the capital initially came (Dishinger et. al. 2014). According to Pavlínek and Ženka (2016), foreign-owned companies create and capture more value than the lower tier domestic companies in Central Europe. Actual gains from GVCs in Central Europe therefore need to be empirically examined.

On the one hand empirical research clearly shows that there have been significant shifts in GVC activities in the pre- and post-crisis eras, but on the other it demonstrates that GVCs and global production remain regionally concentrated. Both statements lead to the suggestion that there might be a major shift in how European companies organise their cross-border production in Europe, possibly explaining at least some of the current account developments and competitiveness shifts. As claimed by Degain, Meng, Wang (2017: 45) "cross-border production sharing activities of complex GVCs increased every year" before the 2000-01 period. "And there was a dramatic expansion of GVCs, especially those with complex production-sharing activities" between 2003 and 2008. That being said, key theorist of global value chains, Richard Baldwin (2014), links his Second Unbundling with the prevailing regionalisation of the global economy: "The world economy is not global; it remains regionally segregated, such as Factory Asia, Factory Europe, and Factory North America. What matters is not value (added) but jobs, especially good jobs." Within Europe, the Eastern enlargement has played a major role since, according to Degain, Meng, Wang (2017: 61), "Eastern European countries have developed intensive bilateral trade linkages in industrial inputs with other European countries. Joining the EU and adopting its regulations have been conducive to the development of these ties within European GVCs. Czechia, Hungary and Poland, the largest players in intraregional trade in manufacturing inputs among European economies, accounted for more than 11% of intra-Europe exports in intermediate goods in 2015, a share that has more than quadrupled since 1995." There is already literature proving that Germany in particular, has gradually weakened its trade ties with Southern Europe in favour of Central Europe (Simonazzi et al. 2013). Similar conclusions were also drawn by the German Bundesbank (2011). Pavlinek

(2019) proves in a study on the automotive industry that geographic restructuring in the industry has been taking place since the late 1990s. While automotive production in France, Italy and Portugal is decreasing, production in the so-called integrated periphery, including Central Europe, is gradually increasing. This trend is explained by the labour cost advantage (Pavlínek, 2019), which is supported also by Chiappini (2012). For Germany, the Eastern enlargement was an opportunity to outsource low-skilled processes abroad, import the necessary inputs from Central European cost-efficient economies and keep the mid-skilled processes on domestic soil (Coricelli and Wörgötter 2012). Keeping only the final stages of production in Germany contributed to the sharp fall in German unit labour costs, increasing the competitiveness of the German industry (Marin 2010) and contributing to more efficient corporate labour division (Walker 1989). Apart from lower labour costs, Pavlinek (2018) also identifies geographic proximity to large markets, membership in regional trade agreements, and investment incentives as Central Europe's key assets.

In contrast, Italian companies inclined to the delocalisation of entire production processes to Southeast Europe (German Bundesbank 2011) leave much less scope for increasing GVC income in Southern Europe. In this context, understanding a country's current participation in value chains is central to ensuring that its industrial and trade policies can facilitate sustainable productivity gains and increased quality employment.

This paper follows on from Brumm et al. (2016), one of the few studies evaluating the effects of participation in global value chains on current account imbalances. Brumm et al. (2016) came to the conclusion that economies which demonstrate higher GVC participation also exhibit larger current account surpluses, as the positive effect on trade balance surpasses the negative impact on the balance of primary incomes. This is true especially for backward participation, which affects current account balance much more than forward participation. The positive effect of downstream participation is also confirmed by Haltmaier (2015), who, however, rejects the notion that there is any positive effect of upstream participation. As most of the countries included in this analysis demonstrated higher backward GVC participation rates in 1995 and 2009 (table 1), GVC income will be considered as positively affecting the current account balance of an economy.

Tab. 1 GVC backward / forward participation indexes.

	1995	2009
Czechia	32.1 / 19.4	39.4 / 23.0
Hungary	26.6 / 15.2	39.9 / 16.7
Poland	15.4 / 17.5	27.9 / 20.5
Slovakia	35.6 / 20.7	44.3 / 17.9
Greece	13.2 / 17.6	23.1 / 19.9
Italy	21.9 / 16.8	20.1 / 21.7
Portugal	28.9 / 15.3	32.4 / 19.0
Spain	20.6 / 19.7	20.7 / 21.1

Source: OECD.stat (2018)

The ECB (2017) also affirms that there is a relationship between GVC participation (especially backward) and current account or trade balance. Nevertheless, the justification is slightly different. Economies participating more in GVCs demonstrate larger current account surpluses (or smaller current account deficits) resulting from the increased competitiveness of domestic producers. The higher competitiveness of companies actively participating in GVCs is caused by stronger competition in the global markets. Several empirical studies confirm the positive impact of GVC participation on firm-level productivity (e.g. Amiti and Wei 2009; Winkler 2010; Crino 2008). The relative increase of domestic competitiveness tends to be only temporary. In order to distribute increased consumption over time, domestic savings increase, which contributes to an improvement in the external trade position.

Institutional setup proved to be important for the participation of an economy in GVCs, according to Dollar et al. (2016). These researchers argue that countries demonstrating better institutional quality tend to have higher forward GVC participation, while countries with weaker institutions usually evince higher backward participation. Institutions are therefore able to influence international trade. Sturgeon et al. (2008b) confirm this idea by highlighting the importance of institutions, especially concerning the labour market, to the distribution of automotive value chains. Nunn and Trefler (2014) summarise the effects of economic institutions on comparative advantages of states, paying particular attention to institutions regulating contractual arrangements among trading parties. In GVCs, suppliers produce highly customised products that have higher value only for the anticipated purchaser. Should the contracts not be consistently enforced, the purchaser would be motivated to renegotiate the contract to increase its own benefit after the supplier has already invested in its production. The supplier hereby faces a hold-up problem (Williamson 1985), which can be reduced by high-quality contractual institutions, e.g. property rights and investor protection (Lavchenko 2006). As a result, economies with higher institutional quality demonstrate a high level of relation-specific investments. Nunn (2007) identified the most relation-specific industries, calling them "contract-intensive". According to Nunn, contract intensive sectors include, among others, automotive, computer equipment, telecommunications equipment and engines, which are among the most exported articles from Central Europe (Sankot and Hnát 2015). Nunn's findings at the firm-level are confirmed by Ma et al. (2010), Li et al. (2012) and Feenstra et al. (2012). Apart from contract enforcing institutions, the regulation of labour and financial markets also affects the comparative advantage of an economy. According to Beck (2002), an economy with a developed financial market has a comparative advantage in manufacturing. Becker et al. (2012) prove that high fixed costs in export-oriented industries require well developed financial markets. Tang (2012) argues that countries with highly protective labour markets demonstrate comparative advantage in goods intensive in firm-specific skills. On the other hand, countries with more flexible labour markets demonstrate comparative advantage in industries that are more volatile (Cunat and Melitz, 2012). Pavlínek 2002, 2018; Aláez, Gil and Ullibarri 2015 confirm in their sectoral study that flexibility of the labour market determines spatial distribution of the automotive industry. As the articles exported from Central Europe (e.g. electronic components or industrial machinery) are classified as volatile by Cunat and Melitz (2012), lower regulation will be considered a positive attribute, promoting a GVC shift towards Central Europe.

3. Empirical background

It must be noted that the research outline is markedly influenced by data availability. Since there are numerous attempts at linking economic growth with GVC trends, measuring value added and value captured by an individual economy is a major challenge. As claimed by Amador and Cabral (2013) or Vlčková (2015), consistent data on GVCs are still rare. Most reliable datasets can be newly obtained from several input-output models (OECD TiVA, WIOD, UNCTAD Eora, or GTAP) but, currently, they do not offer the most recent years. This is another reason why our focus is on the build-up stage of the crisis (using the mid-1990s and 2011 as critical years) in our attempt to reveal a structural change in Europe's GVCs.

In this paper, data for real GVC income are taken from the World Input-Output Database (WIOD), based on Timmer et al. (2013). Timmer et al. (2013) calculated real GVC income for all manufactures produced in selected countries, using the US CPI as a deflator. Table 2 depicts real income from GVCs in economies of the two EU peripheral areas in 1995, 2008 and 2011. The time selection is dependent on very limited data availability. Nevertheless, the period between 1995 and 2011 represents quite well the time frame during which European imbalances emerged and increased, as well as the initial period of the European sovereign debt crisis, which had a strong impact on the structurally weak economies in Southern Europe. In order to ensure the comparability of these figures among economies of different sizes, real GVC income is normalised by real GDP.

Tab. 2 Real GVC income (as a % of GDP).

	1995	2008	2011	Difference (95–11)
Czechia	0.104	0.195	0.182	0.078
Hungary	0.120	0.195	0.194	0.074
Poland	0.132	0.190	0.158	0.025
Slovakia	0.108	0.196	0.168	0.060
Greece	0.097	0.092	0.088	−0.009
Spain	0.136	0.116	0.110	−0.025
Italy	0.155	0.160	0.141	−0.014
Portugal	0.131	0.113	0.102	−0.029

Source: Timmer et al. (2013), UNCTADstat (2018)

In 1995, the difference between Southern and Central Europe, in terms of relative real GVC income, was not significant. Until 2008, the GVC income of Central European economies rose markedly, while the income of Southern European countries stagnated (Italy and Greece), or even decreased (Spain and Portugal). During the initial years of the sovereign debt crisis, GVC income in all Southern European economies dropped below the 1995 level, while GVC income in Central European countries remained well above the 1995 benchmark. In terms of real GVC income, Central European economies gained ground, while Southern European economies lost ground. Decreasing relative GVC income implies lower importance of international labour division for economic growth and the higher importance of domestic factors, e.g. household and government consumption, which should be reflected on the current account. The current account also incorporates profit repatriation, which in Central Europe between 1995 and 2008 increased significantly and pulled all current accounts down into deficit (UNCTADstat 2017). This would conceal the change in position of Central European countries within GVCs. Therefore, solely balances of trade will be taken into account.

Tab. 3 Balance of trade (% of GDP).

Country	1995	2008	2011	Difference (95–11)
Czechia	−0.057	0.020	0.047	0.105
Hungary	−0.056	−0.003	0.070	0.126
Poland	−0.043	−0.072	−0.041	0.002
Slovakia	−0.010	−0.029	0.000	0.009
Greece	−0.108	−0.187	−0.117	−0.008
Spain	−0.026	−0.085	−0.047	−0.021
Italy	0.024	−0.008	−0.016	−0.039
Portugal	−0.083	−0.142	−0.095	−0.012

Source: UNCTADstat (2018).

Table 3 depicts the development of trade balance for selected economies. Between 1995 and 2008 trade balances improved only in Czechia and Hungary. The deterioration of trade balances in Poland and Slovakia were, nonetheless, lower when compared to Spain, Portugal and Greece. Until 2011, trade balances deteriorated below the 1995 level in all Southern European economies, while in Central European economies, trade balance improved, albeit with very limited improvement in Poland and Slovakia.

When analysing the relation between balance of trade and income from global value chains, it is necessary to bear in mind the possible built-in bias of both approaches that might cause analytical discrepancies. Balance of trade is measured using standard indicators of international trade (neglecting the origin of the value added), while income from global value chains is measured using advanced decomposition techniques tracing value added of labour and capital needed for the production of final manufactured goods (Timmer 2013). For that reason, figure 1 focuses rather on the big picture, not aiming to provide accurate outcomes, which would be provable in a more advanced econometric model.

A comparison of both groups of countries is shown in figure 1. The X-axis represents the change in relative real GVC income between 1995 and 2011 (depicted also in table 2) while the Y-axis represents the change in trade balance (depicted also in table 3) in the same period.

Based on figure 1, we can observe a direct relationship between change in relative GVC income and change in balance of trade. While the position of Central European economies improved in both GVC income and trade deficit, the opposite holds for all Southern European economies.

4. Data sources, results and robustness

The motivation of TNCs to move production to post-socialist Central Europe was, apart from its geographic location, caused by the local low-cost but skilled labour (Jürgens and Krzywdzinski 2009; Pavlínek 2019; Chiappini 2012). The following analysis therefore takes into account not only the institutional setup but also the local labour costs.

To estimate determinants of structural changes of GVCs in Europe, we employ an OLS linear regression, which is widely used in GVC empirical studies, e.g. Baldwin and Yan (2014). Due to limited GVC data availability, a more advanced analysis of time series cannot be employed. In the analysis, three explanatory variables are used: labour costs, quality of regulatory environment and quality of contract enforcing institutions.

Labour costs cover the compensation of employees plus taxes minus subsidies in the sectors of industry, construction and services (Eurostat 2018). As wages demonstrated a dynamic development, especially in Central Europe, average available labour costs in the selected time frame are applied. Quality of institutions is evaluated by the Rule of Law multi-criterial

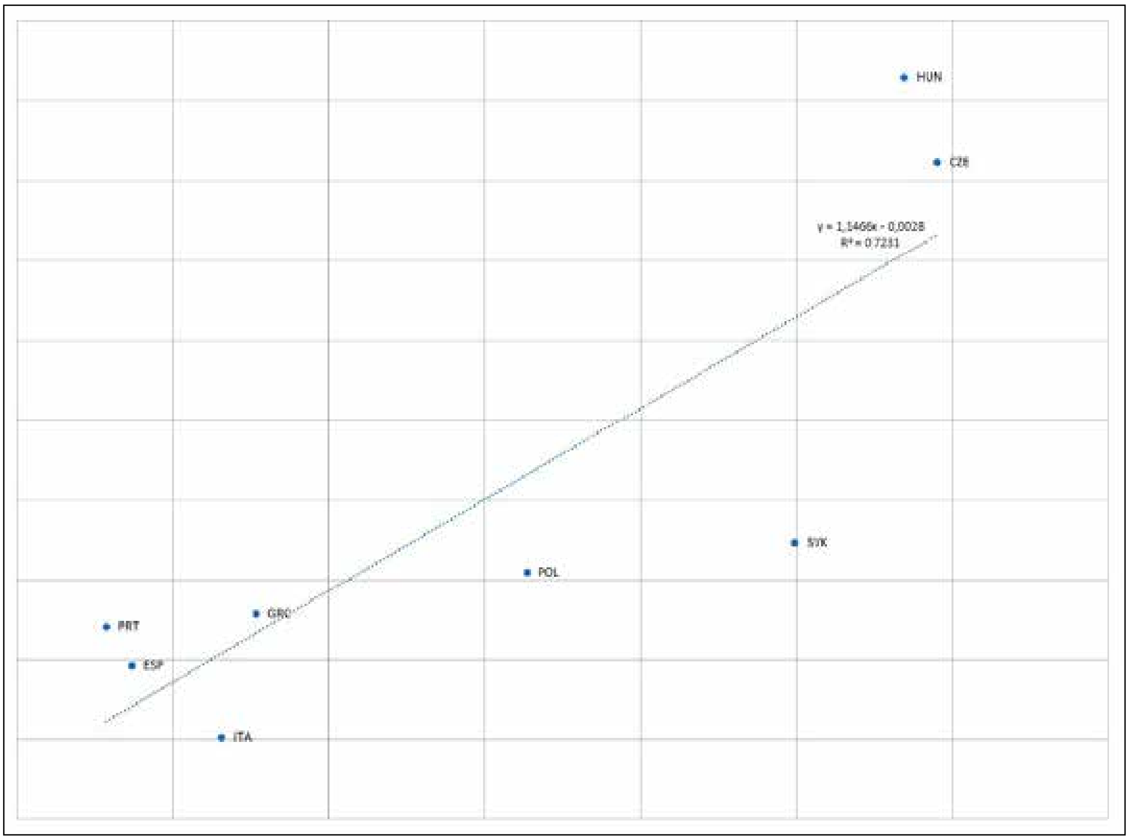

Fig. 1 Change in relative GVC income (x-axis) / change in balance of trade (y-axis) (1995–2011).
Source: Timmer et al. (2013), UNCTADstat (2018)

indicator, compiled by the World Bank (WGI 2017). This measure was developed by Kaufmann et al. and is used as a baseline of the ability to enforce contracts by Nunn (2007) and Lavchenko (2006). Rule of Law index covers a broad range of areas, e.g. property rights protection, quality of enforcement mechanisms, reliability of police services and judicial independence. As comparative advantage is, according to Nunn and Trefler (2014), co-determined by product and financial/labour markets regulation, we also include the multicriterial index of Regulatory Quality (WGI 2017). The regulatory quality index evaluates not only the regulatory environment, but also the effects of trade policy or tax policy on the domestic business environment. Rule of Law and Regulatory Quality indices are compiled from multiple sources, mainly surveys of independent providers (e.g. Gallup World Poll), business information providers (e.g. Economist Intelligence Unit), NGOs (e.g. Freedom House) and public sector organisations (EBRD Transition Report).

The explained variable, i.e. the structural changes in GVCs, is measured by the difference in real GVC income (per unit of GDP) between 1995 and 2011, as depicted in table 2. Economies included in the analysis are, for Southern Europe, Greece, Italy, Spain and Portugal, while Central Europe is represented by Czechia, Hungary, Poland and Slovakia. The estimated impact of selected variables on real GVC income is depicted in table 4.

Tab. 4 Impact of labour costs, regulatory quality and rule of law on GVC shift (1995–2011).

Variable	Coefficient (Standard error)
Labour costs	−0.003 ** (0.001)
Regulatory Quality	0.134 *** (0.026)
Rule of Law	−0.079 ** (0.029)

Sources: Eurostat (2018), WGI (2017)
** stands for p-value lower than 0.05,
*** stands for p-value lower than 0.01

Regression analysis confirms the statistically significant impact of lower labour costs and higher regulatory quality on the GVC shift to Central European countries. Adj. R-sq. of the model is equal to 0.86. TNCs were attracted by labour costs, which were on average, during the selected period, more than 55% lower in Central Europe than in Southern Europe. In addition, all Central European countries demonstrated better regulatory quality than Southern European economies (with the exception of Spain). This also proved to be a statistically significant determinant of the GVC shift towards Central Europe, denoting TNCs fleeing from the overregulated South to the slightly more liberal Central Europe. So far, the outcomes are in line with the available literature. Rule of law proved to be a significant determinant as well. However, the negative coefficient points to the fact that companies were moving their production to an environment with worse contractual institutions, which does not conform with the current theoretical and empirical literature.

A relatively similar result is achieved when the Rule of Law index is replaced by the Corruption Control index (WGI 2017). High levels of bribery and irregular payments are conspicuous signs of weak rule of law. Therefore, the Corruption Control index broadly covers the essence of weak rule of law and enforcement mechanisms in the economy. Meanwhile, the multicriterial Control of Corruption index consists of different input variables from the Rule of Law index, focusing on all kinds of irregular payments.

In order to further test the robustness of the results, we limit the examined period to 1995–2008. By neglecting the years 2009–2011, we omit the period of severe economic crisis. The European debt crisis particularly hit Southern Europe and affected both real GVC income as well as real GDP. Moreover, we replace the variables of the World Bank with indicators provided by the World Economic Forum (WEF). WEF uses data gathered independently, using its own surveys among executives worldwide. We apply the 'Institutions' index as a substitute for the 'Rule of Law' index and the 'Goods Market Regulation' index replaces 'Regulatory Quality'. The 'Institution' index evaluates the quality of both formal and informal institutions, i.e. the quality of state administration, protection of natural rights and also the ethical standards of companies. The 'Goods Market Efficiency' index evaluates, among other things, how easy it is to start a business, rules on FDIs, or what the effects of taxation are. Table 5 summarises the impact of the amended variables on real GVC income (per unit of GDP) between 1995 and 2008.

Tab. 5 Impact of labour costs, goods market regulation and institutional quality on GVC shift (1995–2008).

Variable	Coefficient (Standard error)
Labour costs	−0.004 ** (0.001)
Goods market regulation	0.085 *** (0.015)
Institutions	−0.074 *** (0.017)

Sources: Eurostat (2018), WEF (2009)
** stands for p-value lower than 0.05,
*** stands for p-value lower than 0.01

Outcomes of the modified regression in table 5 are comparable with the original model (table 4), adj. R-sq. of the modified model is equal to 0.92. The modified model confirms that GVC income in Central Europe increased despite the lower quality of domestic institutions. This, however, did not discourage

TNCs from moving their production eastwards, mainly because of the prospects of EU membership. The EU's Eastern enlargement managed to guarantee the necessary institutional standards. Moreover, regulatory quality in Central European countries improved between 1996 and 2008 (WGI 2017).

Furthermore, it has to be added that the results hold only for selected Central European countries, i.e. Czechia, Hungary, Poland and Slovakia. Once different countries (Bulgaria, Latvia, Lithuania or Romania) are incorporated in the model, the statistical significance of labour costs or rule of law disappears. The reason might be twofold. First, a less skilled labour force and greater geographic distance are not sufficiently compensated for by the decrease in labour costs. Alternatively, the trade-off between labour costs and quality of contractual institutions is only limited and TNCs are ready to accept a decrease in rule of law only to a certain extent. In the latter case, significantly lower institutional quality (e.g. in Bulgaria and Romania) could not be offset, even by substantially lower labour costs.

5. Limitations, conclusions and discussion

This empirical study demonstrates several limitations. Apart from the general drawbacks related to overall GVC data and their application, as discussed above, scholars are divided on the issue of how to measure soft variables like institutional quality. They also question whether the current data on institutions are relevant and generally applicable in econometric models, e.g. see Glaeser et al. (2004) or Voigt (2009) for details. As there is an overwhelming consensus that institutions matter, we employ the available and broadly used indicators, despite the ongoing academic discussion (e.g. Kaufmann et al. 2009 vs. Thomas 2009). Moreover, the limited amount of countries taken into account might reduce the statistical power of the regression models. However, the number of Southern European EU member states is factually given and results for other Central and Southeast European countries are discussed above.

With the limitations taken into account, the following conclusions can be drawn. Significant changes in GVC real income between 1995 and 2011 indicate a shift of GVC-related production from Southern Europe towards Central Europe. While Central European countries demonstrated improvement in GVC-related real income, income in Southern European countries diminished or stagnated. In this article, we argue that such a shift was reflected in the trade balances of both Central and Southern European countries. Therefore, this shift was also one of the contributors to the growing current account imbalances in the EU during the build-up to the European sovereign debt crisis. In this article, we identify total labour costs together with regulatory quality as reasons for the shift.

This is in accordance with other authors who have focused predominantly on increasing unit labour costs as a significant determinant for European imbalances. In a broader sense, soaring labour costs in Southern Europe together with the Eastern enlargement brought more cost-efficient competitors into the European Common Market. These factors undermined the existing comparative advantages in Southern European countries. Higher value-added production remained in the EU core, which exhibits higher total factor productivity. Simultaneously, lower value-added production moved to more cost-efficient new EU member states in Central Europe, leaving Southern Europe dependent on domestic demand. Subsequent changes in current account positions were further underpinned by soaring government expenditures and liquidity, initially through the misallocated investments of high-savings countries and later through the euro area's liquidity distribution mechanism, the TARGET2 system.

As regulatory quality proved to be one of the contributors to the GVC shift from Southern to Central Europe, the push for the liberalisation of product and labour markets in deficit-prone Southern Europe seems to be justified. During the years following the European sovereign debt crisis, current account deficits in Southern Europe almost disappeared. However, as Jirankova et al. (2015) demonstrate, this was predominantly due to a dramatic decrease in consumption in Southern European countries, i.e. due to the income adjustment mechanism of the balance of payments. This means that the preconditions for European imbalances have not yet been sufficiently tackled.

Even though this approach might provide a short- to medium-term solution, it would not address the issue of relatively lower competitiveness in Southern Europe when compared to Central Europe. Southern European countries have approved some measures to address their relatively lower competitiveness resulting from the unfavourable productivity-wages ratio. Such measures include reduced non-wage labour costs in Greece, increased flexibility of working time management in Portugal and increased space for firm-level bargaining to derogate from sectoral contracts in Spain and Italy (Buti and Turrini 2012).

Despite the actions taken recently, countries in Southern Europe are still lagging behind in terms of macroeconomic competitiveness, and not only behind the EU's core but also behind Czechia (WEF 2017). One of the major weaknesses of Southern Europe, besides the macroeconomic environment, is its financial market development, which suffers from the redirection of private cross-border capital flows towards surplus countries, i.e. from a direct effect of European imbalances. Should Southern European economies boost their productivity, they will have to attract foreign investments and private capital flows again. More robust financial markets would also accelerate and increase the benefits of conducted structural reforms. Completing the banking union would increase the trustworthiness of the financial sector. Further deregulation of goods and labour markets as well as improving the institutional environment are other preconditions to restore the attractiveness of Southern Europe for foreign investors. Foreign investments will later induce increased GVC income and, subsequently, also drive current account adjustment, which will not be based upon limited internal consumption.

Central European countries seem to benefit from the GVC shift eastwards. However, this has to be viewed only as a short-term victory. Foreign capital has been attracted predominantly to its cost-competitive labour, while quality of human capital and total factor productivity still do not match the standards of the EU's core. Therefore, Central Europe is actually prone to suffering from similar flaws as Southern Europe did before the start of the European debt crisis. The regulatory environment is currently improving in Portugal and Spain, while slightly deteriorating in Hungary and Slovakia. Moreover, contrary to general belief, Central European countries attracted TNCs despite relatively lower rule of law standards. This is, however, rather an exception to the rule. In our model, the statistically significant negative coefficient for rule of law does not hold universally for all the cost-efficient economies in Central and Southeast Europe, but only for Czechia, Hungary, Poland and Slovakia. These countries demonstrated only minor rule of law deficiencies. Rule of law and institutional quality in general, therefore, remain a crucial determinant of FDI allocation and the position of a country in GVCs. Unfortunately for Central Europe, rule of law is currently worsening in Hungary and Poland. In Hungary's case, it is also accompanied by increasing levels of corruption. Unfortunately, institutional quality is not universally improving in Central Europe, leaving the region dependent on still relatively low labour costs.

Cost-based comparative advantages will sooner or later be exhausted, as wage pressure in Central European economies is currently on the rise. Central European countries thus run a risk that they will share the same fate as Southern Europe. Wage increases unaccompanied by corresponding productivity and competitiveness gains might cause another shift of GVCs, leaving Central Europe in the same position as Southern Europe in 2009. Based upon WEF (2017) data, Central European economies, like their Southern European counterparts, suffer from comparably weak institutions, stringent labour market regulation and lower quality education systems. However, unlike Southern European countries, Central European economies possess a larger scope for fiscal policy to address well-known structural weaknesses. Political will and the stability of the institutional environment, therefore, remain the most relevant determinants of future economic development in Central Europe.

Acknowledgements

This research was financially supported by the IGA project No. F2/65/2019.

References

Aláez, R., Gil, C., Ullibarri, M. (2015): FDI in the automotive plants in Spain during the great recession. In: Foreign investment in eastern and southern Europe after 2008: Still a lever of growth? ed. B. Galgóczi, J. Drahokoupil, and M. Bernaciak, 139–70. Brussels: ETUI.

Amador, J., Cabral, S. (2014): Global Value Chains: A survey of drivers and measures. Journal of Economic Surveys 28(3), 401–593.

Amiti, M., Wei, S.-J. (2009): Service Offshoring and Productivity: Evidence from the US. The World Economy, 32(2), 203-220, https://doi.org/10.1111/j.1467-9701.2008.01149.x.

Baldwin, J., Yan, B. (2014): Global Value Chains and the Productivity of Canadian Manufacturing Firms. Economic Analysis (EA) Research Paper Series [online], available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.903.6431&rep=rep1&type=pdf.

Baldwin, R. (2006): Globalisation: The great unbundling(s). Paper prepared for the Finnish Prime Minister's Office for EU Presidency.

Baldwin, R. (2014): Global Value-Chain Training and Research Workshop, June 30 – July 11, 2014. University of International Business and Economics, Beijing, China.

Beck, T. (2002): Financial development and international trade: is there a link? Journal of International Economics 57(1), 107–131, https://doi.org/10.1016/S0022-1996(01)00131-3.

Becker, B., Chen, J., Greenberg, D. (2012): Financial Development, Fixed Costs and International Trade. Mimeo, Harvard Business School, https://doi.org/10.1093/rcfs/cfs005.

Brumm, J. et al. (2016): Global Value Chain Participation and Current Account Imbalances. ECB [online], available at: https://cepr.org/sites/default/files/(BRUMM_GEORGIADIS_GRAB_TROTTNER)_GVC_Participation_and_Current_Account_Imbalances.pdf.

Buti, M, Turrini, A. (2017): Overcoming Eurozone wage inertia. VoxEU.org, October, 6 [online], available at: http://voxeu.org/article/overcoming-eurozone-wage-inertia.

Buti, M, Turrini, A. (2012): Slow but steady? External adjustment within the Eurozone starts working. VoxEU.org, November, 12 [online], available at: http://voxeu.org/article/slow-steady-external-adjustment-within-eurozone-starts-working.

Chiappini, R. (2012): Offshoring and export performance in the European automotive industry. Competition and Change 16, 323–342, https://doi.org/10.1179/1024529412Z.00000000020.

Coe, N. M., Dicken, P., Hess, M. (2008): Global production networks: Realizing the potential. Journal of Economic Geography 8(3), 271–295, https://doi.org/10.1093/jeg/lbn002.

Coricelli, F., Wörgötter, A. (2012): Structural Change and the Current Account: The Case of Germany: OECD Economics

Department Working Papers, 940, OECD Publishing, https://doi.org/10.1787/5k9gsh6tpz0s-en.
Crinò, R. (2008): Service Offshoring and Productivity in Western Europe. Economics Bulletin 6(35), 1–8.
Cunat, A., Melitz, M. J. (2012): Volatility, labor market flexibility, and the pattern of comparative advantage. Journal of the European Economic Association 10(2), 225–254, https://doi.org/10.1111/j.1542-4774.2011.01038.x.
Degain, Ch., Meng, B., Wang, Z (2017): Recent trends in global trade and global value chains in Global value chain development report 2017: measuring and analyzing the impact of GVCs on economic development, 37–68, available at: http://hdl.handle.net/2344/00049246.
Dicken, P. (2015): Global shift: Mapping the changing contours of the world economy. London: Sage Publications.
Dischinger, M., Knoll, B., Riedel, N. (2014): There's no place like home: The profitability gap between headquarters and their foreign subsidiaries. Journal of Economics and Management Strategy 23(2), 369–395, https://doi.org/10.1111/jems.12058.
Dollar, D. et al. (2016): Institutions and Participation in Global Value Chains. Global Value Chain Development Report [online], available at: http://rigvc.uibe.edu.cn/docs/20160407201118816062.pdf.
Domanski, B., Gwosdz, K. (2009): Toward a more embedded production system? Automotive supply networks and localized capabilities in Poland. Growth and Change 40(3), 452–482, https://doi.org/10.1111/j.1468-2257.2009.00490.x.
Dunaway, S. (2009): Global Imbalances and the Financial Crisis. Council on Foreign Relations, Special Report, 44.
ECB (2008): A Framework for Assessing Global Imbalances. Occasional Paper, 78.
ECB (2017): ECB Economic Bulletin, 2, 2017. The impact of global value chain participation on current account balances – a global perspective.
Eurostat (2018): Statistics database [online], available at: http://ec.europa.eu/eurostat.
Feenstra, R., Hong, C., Ma, H., Spencer, B. J. (2012): Contractual Versus Non-Contractual Trade: The Role of Institutions in China. NBER Working Paper 17728, https://doi.org/10.3386/w17728.
Gereffi, G. (1994): The organisation of buyer-driven global commodity chains: How US retailers shape overseas production networks. In: G. Gereffi, M. Korzeniewicz (eds.), Commodity chains and Global Capitalism, Westport-London: Praeger.
German Bundesbank (2011): Developments in the exports of the four largest euro-area member states since the launch of monetary union. Monthly Report, July, 15–34.
Glaeser, E., La Porta, R., Lopez-de-Silanes, F., Shleifer, A. (2004): Do Institutions Cause Growth? Journal of Economic Growth, 9(3), 271–303, https://doi.org/10.1023/B:JOEG.0000038933.16398.ed.
Guzik, R., Micek, G. (2008): The impact of delocalisation in the European software industry. The Moving Frontier: The Changing Geography of Production in Labour-Intensive Industries, 229-254.
Haltmaier, J. (2015): Have Global Value Chains Contributed to Global Imbalances? International Finance Discussion Papers 1154 [online], available at: https://www.federalreserve.gov/econresdata/ifdp/2015/files/ifdp1154.pdf, https://doi.org/10.17016/IFDP.2015.1154.
Henderson, J., Dicken, P., Hess, M., Coe, N. M., Yeung H. W. C. (2002): Global production networks and the analysis of economic development. Review of International Political Economy 9(3), 436–464, https://doi.org/10.1080/09692290210150842.
Humphrey, J., Lecler, Y., Salerno, M. S. (eds.) (2000): Global strategies and local realities: The auto industry in emerging markets. Houndmills: Macmillan Press, https://doi.org/10.1057/9780333977712.
Humphrey, J., Schmitz, H. (2002): How does insertion in global value chains affect upgrading in industrial clusters? Regional Studies 36(9), 1017–1027, https://doi.org/10.1080/0034340022000022198.
IMF (2009). Global Imbalances: In Midstream? IMF Staff Position Note, https://doi.org/10.5089/9781462333387.004.
Jürgens, U., Krzywdzinski, M. (2009): Changing East–West Division of Labour in the European Automotive Industry. European Urban and Regional Studies 16(1), 27–42, https://doi.org/10.1177/0969776408098931.
Jiránková, M., Hnát, P. (2012): Balance of payments adjustment mechanism in the Euro area. Eastern Journal of European Studies 3(1), 67–86.
Jiránková, M. et al. (2015): Euro Area Imbalances – Macroeconomic Competitiveness as a Balancing Factor?. In: Antal, J. (ed.). Small States – Big Challenges: The Experience of the EU and Visegrad Region. Oeconomica, Prague, 9–25.
Kaufmann, D., Kraay, A., Mastruzzi, M. (2009): Response to "What Do the Worldwide Governance Indicators Measure?". World Bank [online], available at: http://info.worldbank.org/governance/wgi/pdf/KKMResponseEJDR2Final.pdf
Lavchenko, A. A. (2006): Institutional Quality and International Trade. The Review of Economic Studies [online], available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.561.1865&rep=rep1&type=pdf
Li, K., Wang, Y., Wang, Y. (2012): Judicial Quality, Contract Intensity and Firm Exports: Evidence from China. Mimeo, Nankai University.
Ma, Y., Baaozhi, Q., Zhang, Y. (2010): Judicial quality, contract intensity and trade: firm-level evidence from developing and transition countries. Journal of Comparative Economics 38, 146–159, https://doi.org/10.1016/j.jce.2009.09.002.
Marin, D. (2010): Germany's Super Competitiveness: A Helping Hand from Eastern Europe [online], available at: http://www.voxeu.org/article/germany-s-super-competitiveness.
Myant, M., Drahokoupil, J., (2012): International integration, varieties of capitalism and resilience to crisis in transition economies. Europe-Asia Studies 64(1), 1–33, https://doi.org/10.1080/09668136.2012.635478.
Nunn, N. (2007): Relationship-specificity, incomplete contracts, and the pattern of trade. Quarterly Journal of Economics 122(2), 569–600, https://doi.org/10.1162/qjec.122.2.569.
Nunn, N., Trefler, D. (2014): Domestic Institutions as a Source of Comparative Advantage. Handbook of International Economics, 4, https://doi.org/10.1016/B978-0-444-54314-1.00005-7.

O'Brian, R., Williams, M. (2010): Global Political Economy. 3rd edition. Palgrave Macmillan, Houndmills.
OECD (2010): Current Account Imbalances in the Euro Area: A Comparative Perspective. OECD, Paris.
OECD.Stat (2018): Statistics database [online], available at: https://stats.oecd.org.
Pavlínek, P. (2002): Transformation of the Central and East European passenger car industry: selective peripheral integration through foreign direct investment, Environment and Planning A 34, 1685–1709, https://doi.org/10.1068/a34263.
Pavlínek, P. (2004): Regional Development implications of foreign direct investment in Central Europe. European Urban and Regional Studies 11(1), 47–70, https://doi.org/10.1177/0969776404039142.
Pavlínek, P., Domański, B., Guzik, R. (2009): Industrial upgrading through foreign direct investment in Central European automotive manufacturing. European Urban and Regional Studies 16(1), 43–63, https://doi.org/10.1177/0969776408098932.
Pavlínek, P., Ženka, J. (2011): Upgrading in the automotive industry: firm-level evidence form Central Europe. Journal of Economic Geography 11(3), 559–586, https://doi.org/10.1093/jeg/lbq023.
Pavlínek, P., Ženka, J. (2016): Value creation and value capture in the automotive industry: Empirical evidence from Czechia. Environment and Planning 48(5), 937–959, https://doi.org/10.1177/0308518X15619934.
Pavlínek, P. (2018): Global Production Networks, Foreign Direct Investment, and Supplier Linkages in the Integrated Peripheries of the Automotive Industry. Economic Geography 94(2), 141–165, https://doi.org/10.1080/00130095.2017.1393313.
Pavlínek, P. (2019): Restructuring and internationalization of the European automotive industry. Journal of Economic Geography, 1–33, https://doi.org/10.1093/jeg/lby070.
Porter, M. E. (1985): Competitive advantage: Creating and sustaining superior advantage. London: MacMillan.
Plank, L., Staritz, C. (2013): 'Precarious upgrading' in electronics global production networks in Central and Eastern Europe: The cases of Hungary and Romania. Capturing the Gains Working Paper, 31, University of Manchester, https://doi.org/10.2139/ssrn.2259671.
Sankot, O., Hnát, P. (2015): Comparative advantage of V4 countries: Trends and changes between 2003 and 2013. In: Grešš, M., Mutual Relations between the Republic of Korea and V4 Countries in Trade and Investments.
Sass, M., Szalavetz, A. (2014): R&D-based integration and upgrading in Hungary. Acta Oeconomica 64(1), 153–180, https://doi.org/10.1556/AOecon.64.2014.S1.6.
Simonazzi, A., Ginzburg, A., Nocella, G. (2013): Economic relations between Germany and southern Europe. Cambridge Journal of Economics 37(3), 653–675, https://doi.org/10.1093/cje/bet010.
Šíma, O. (2016): Heterogenita mobilních výrobních faktorů jako narušení podmínky optimální měnové oblasti (příklad eurozóny). Politická ekonomie 64(3), 319–337, https://doi.org/10.18267/j.polek.1072.
Smith, A., Pickles, J., Buček, M, Pástor, R, Begg, B. (2014): The political economy of global production networks: Regional industrial change and differential upgrading in the East European clothing industry. Journal of Economic Geography 14(6), 1023–1051, https://doi.org/10.1093/jeg/lbt039.
Sturgeon, T. J., Memedovic, O., van Biesebroeck, J., Gereffi, G. (2008a): Globalisation of the automotive industry: Main features and trends. International Journal of Technological Learning, Innovation and Development 2(1–2), 7–24, https://doi.org/10.1504/IJTLID.2009.021954.
Sturgeon, T., Van Biesebroeck, J., Gereffi, G. (2008b): Value chains, networks and clusters: reframing the global automotive industry. Journal of Economic Geography 8, 297–321, https://doi.org/10.1093/jeg/lbn007.
Tang, H. (2012): Labor market institutions, firm-specific skills, and trade patterns. Journal of International Economics 87(2), 337–351, https://doi.org/10.1016/j.jinteco.2012.01.001.
Thomas, M. A. (2009): What Do the Worldwide Governance Indicators Measure? European Journal of Development Research 22(1), 31–54, https://doi.org/10.1057/ejdr.2009.32.
Timmer, M. P., Los B., Stehrer R., de Vries G. J. (2013): Fragmentation, Incomes, and Jobs in an Analysis of European Competitiveness. Economic Policy [online] 11(4), 613–661. Data available at: http://www.wiod.org/gvc, https://doi.org/10.1111/1468-0327.12018.
UNCTADstat (2017): Statistics database [online], available at: http://unctadstat.unctad.org/EN/.
Vlčková, J. (2017): Global production networks in Central European Countries: the case of the Visegrad Group. Praha: Nakladatelství Oeconomica, 2015, https://doi.org/10.18267/pu.2017.vlc.2197.8.
Vlčková, J. (2015): Can exports be used as an indicator of technological capabilities of countries? Geografie 120(3), 314–329.
Vlčková, J. (ed.). et al. (2015): How to Benefit from Global Value Chains: Implications for the V4 countries [online]. Praha, 09.06.2015. Praha, Nakladatelství Oeconomica.
Voigt, S. (2009): How (Not) to Measure Institutions (February 1, 2009), http://dx.doi.org/10.2139/ssrn.1336272
Walker, R. (1989): A requiem for corporate geography: new directions in industrial organization, the production of place and the uneven development. Geografiska Annaler. Series B. Human Geography 71, 43–68, https://doi.org/10.1080/04353684.1989.11879586.
WEF (2009): Global Competitiveness Report 2008–2009. WEF, Geneva.
WEF (2017): Global Competitiveness Report 2017–2018. WEF, Geneva.
WGI (2017): World Governance Indicators [online], available at: http://info.worldbank.org/governance/wgi/#home.
Williamson, O. E. (1985): The Economic Institutions of Capitalism. The Free Press, New York.
Winkler, D. (2010): Services Offshoring and its Impact on Productivity and Employment: Evidence from Germany, 1995–2006. The World Economy 33(12), 1672–1701, https://doi.org/10.1111/j.1467-9701.2010.01269.x.
Wyplosz, C. (2010): Germany, current accounts and competitiveness. VoxEU.org, March, 31 [online], available at: http://voxeu.org/article/germany-current-accounts-and-competitiveness.

Quantitative mapping of desertification risk using the modified MEDALUS model: a case study in the Mazayejan Plain, Southwest Iran

Reza Zakerinejad[1] , Masoud Masoudi[2, *]

[1] University of Isfahan , Faculty of Geographical and planning, Iran
[2] Shiraz University, Department of Natural Resources and Environmental Engineering, Iran
* Correspondence author: masoudi@shirazu.ac.ir

ABSTRACT

This paper presents the Modified MEDALUS (MMEDALUS) approach, a quantitative assessment of desertification, in the case study area located in the Southern part of Iran. Six main factors of desertification including: soil, climate, plant cover, management, erosion state and ground water situation were considered for the model approach. Then several sub-factors determining the quality of each main factor were quantified according to their quality and weighted on a scale between 1.0 and 2.0. We used a Geographic Information System (GIS) software to analyze and prepare the spatial distribution of the factor layers. Subsequently, the final desertification hazard map was prepared by combining the different MEDALUS factors in Arc GIS 10.3 in order to define the final hazard classes on the basis of hazard scores based on the geometric mean of the main factors. The MEDALUS and MMEDALUS models show the "Desertification Potential" that in turn was validated with the current state of desertification observed in the field. The results show that the applied MMEDALUS approach yield significantly better results than the MEDALUS model in the study area. The results also show that the areas under severe and very severe hazard are the most extensive classes in the desertification map. Thus, we illustrate that most of the study area is sensitive to desertification. However, we highlight that management, climate and water table qualities were the most important indicators affecting the desertification processes, while soil quality seems to play a minor role in our study area.

KEYWORDS

desertification; assessment; MEDALUS; MMEDALUS

Received: 28 August 2019
Accepted: 21 September 2019
Published online: 20 December 2019

Zakerinejad, R., Masoudi, M. (2019): Quantitative mapping of desertification risk using the modified MEDALUS model: a case study in the Mazayejan Plain, Southwest Iran. AUC Geographica 54(2), 232–239
https://doi.org/10.14712/23361980.2019.20

1. Introduction

Land degradation is a significant global environmental and socioeconomic problem (Taddese 2001; Miao et al. 2015). Drylands (arid, semi-arid, and dry sub-humid areas) cover approximately 40% of the Earth's surface (Deichmann et al. 2018) that are more sensitive to degradation (Zakerinejad et al. 2018). Desertification refers to land degradation caused by climate change and human activity in arid, semi-arid, and dry sub-humid areas (UNCCD 2014). In the early 1990s, desertification was defined as 'land degradation resulting from various factors, including climatic variations and human activities' (UNEP 1992). It is a prolonged type of land degradation which in space and time converts the productive ecosystem to a fragile one by two crucial factors, namely, climate and negative human activity. (Shoba, Ramakrishnan 2016). The character and intensity of desertification is closely related to environmental factors such as climate, soil characteristics, vegetation cover, and morphology. Desertification is also strongly linked to socio-economic factors, since human's behavior and their social and economic actions can greatly influence the evolution of numerous environmental characteristics. The United Nations Environmental Program (UNEP) estimates that 69% of the world's arid lands, excluding the very arid deserts, are under moderate to severe hazard of land degradation (Dregne 1991). This type of land degradation seriously threatens agriculture, natural resources and the environment (Lal 1998; Yang et al. 2003; Feng et al. 2010; Fleskens, Stringer 2014; Zakerinejad et al. 2018). Especially areas with arid and semi-arid climates are affected due to a lack of financial resources to cope with and mitigate the effects of soil erosion and desertification (Zakerinejad, Märker 2014; Zakerinejad, Märker 2015; Masoudi, Jokar 2017). Desertification was recognized as a severe problem already between the 1930s and 1960s in Iran. Iran having an arid to semi-arid climate with low precipitation and high evaporation rates compared to world averages shows a high vulnerability to land degradation and desertification. In the last decades over 20 per cent of the country is exposed to desertification. It has detrimental impacts on agricultural productivity and on ecological function (Zehtabian, Jafari 2002; Eliasson et al. 2003; Amiraslani, Dragovich 2009; Pan and Li 2013).

Recently, several methods of desertification and land degradation assessment have been applied. The FAO/UNEP (1984) introduces the "Provisional Methodology for Assessment and Mapping of Desertification Hazard" which evaluates the main factors affecting desertification processes. This method was the first major apporach that was developed to assess land degradation in arid and semi-arid regions. Some other important models are Global Assessment of Soil Degradation – GLASSOD (Oldeman et al. 1991), Assessment of Soil Degradation – ASSOD (Van Lynden, Oldeman 1997) and LADA (Ponce Hernandez, Koohafkan 2004). Another model specifically developed for the Iranian conditions is the Iranian Model of Desertification Potential Assessment (IMDPA) proposed by the Iranian Forests and Rangeland Organization (Ahmadi 2007; Masoudi, Zakerinejad 2010). 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. In total 35 indicators are used by the model (Masoudi, Zakerinejad 2011). An alternative model especially developed and applied in several parts of the Mediterranean is the MEDALUS (Kosmas et al. 1999) which identifies regions that are environmentally sensitive to desertification processes. The model evaluates the main quality layers (factors) including soil, climate, vegetation, and management. After assessing these factors or quality layers, the Environmental Sensitivity Index (ESI) is defined by combining the four quality layers. ESI is a composite indicator that can be used to get insights into the factors causing desertification risk at a certain point in the landscape. Since the model was often applied in Mediterranean regions it is also appropriate to be applied in the Iranian conditions (arid and semiarid regions). So, all the data considering the MEDALUS (Modified MEDALUS) layers were prepared in a geographical information system (GIS), and overlay in accordance with the developed algorithm which took the geometric mean to compile maps of desertification intensity.

The main aim of this study is the assessment of the most important factors affecting desertification in the study area using the MMEDALUS model. Moreover, we compare the MMEDALUS model with the original model (MEDALUS) and evaluate both with the current state of desertification observed in the field. For this purpose, the Mazayejan (MZY) plain, located in Southern part of Iran, for which enough data were available has been chosen.

2. Materials and methods

2.1 Study area

The study area is located in the MZY plain, in the Fars province in the Southern parts of Iran (54°34′ to 54°44′ E and 27°59′ to 28°5′ N). The study area shown in Figure 1 is part of the Zagros Mountains. The Zagros mountain range from the north-west along the west to south-eastern parts of Iran. Is highest peak reaching nearly 5000 meters. The study area covers an ca 20,000 ha and is drained by the MZY river. According to the national topography map (1 : 25,000; Iranian Cartographic Center 1994) the elevation is ranging from 693 m a.s.l. to a maximum

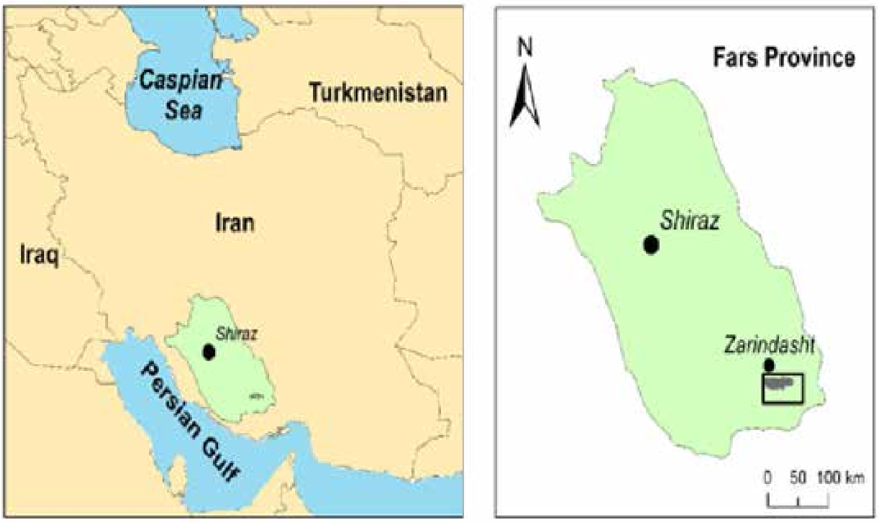

Fig. 1 Study area: MZY plain in Fars province Southwestern Iran.

altitude of 1,371 m a.s.l. The annual average rainfall is around 243 mm with a high inter-annual variability characterized by very dry summer months (June to September), followed by short period of heavy rainfall from December till March. Soil erosion as one of the most import types of land degradation and desertification is occurring frequently especially in the pediments of the mountain ranges of this area (Zakerinejad, Märtker 2014; Zakerinejad, Märker 2015).

2.2 Methodology

Desertification is a complex phenomenon that leads to the reduction of land productivity (Sepehr et al. 2007). In order to assess the degree of desertification we applied the MEDALUS and MMEDALUS models. The key indicators of the MEDALUS model allow to identify the Environmental Sensitivity Areas (ESA). Generally, ESAs represent areas whose socio-economic and ecological aspects are not sustainable for a particular landuse (Basso et al. 2004).

The data for this study such as the information on soils, climate and vegetation cover were collected in the field, from maps and from data available in the related reports published by the different departments of the Ministries of Jahade-Agriculture as well as by the Meteorological Organization of Iran. The MMEDALUS model is differing from the MEDALUS model since it considers two more factors, namely, the soil erosion state and the ground water situation (Table 1). Moreover, information about the organic matter, soil EC and evapotranspiration were integrated in the model. Furthermore, we adapted rainfall and their hazard scores for specific soil depth as shown in Table 1. Based on this method, 20 entities were identified in the study area. Each entity was considered as an individual study unit and the assessment of desertification was conducted for all of them.

Tab. 1 Classes and assigned weighing indices for the various parameters used for the assessment of soil quality (a), climate quality (b), vegetation quality (c), land management quality (d), erosion state quality (e), ground water state quality (f).

a. Soil quality index in the model

Slope (%)		Drainage		Soil Depth(cm)		EC (mmhos/cm)	
Description	Idex	Description	Index	Description	Index	Description	Index
<6	1	well drained	1	>75	1	4>	1
6–18	1.2	imperfectly drained	1.2	75–30	1.2	4–8	1.2
18–35	1.5	Poorly drained	2	15–30	1.6	8–16	1.4
>35	2			<15	2	16–32	1.6
						32–64	1.8
						>64	2
Texture				Organic matter (%)		Stone fragment cover (%)	
Description			Index		Index		Index
L, SCL, SL			1	>3	1	>60	1
LS, CL				2–3	1.2		
SC, SiL, SiCL			1.2	0.5–1	1.7	20–60	1.3
Si, C, SiC			1.6	<0.5	2	<20	2
S			2				

b. Climate quality index in the model

Rainfall (mm)		Aridity index (P/ETp)		Evapotranspiration (mm)	
Description	**Index**	**Description**	**Index**	**Description**	**Index**
>300	1	AI ≥ 1	1	<1500	1
150–300	1.5	0.1 < AI < 1	1.5	1500–2000	1.5
<150	2	AI ≤ 0.1	2	>2000	2

c. Vegetation quality index in the model

Fire risk		Erosion protection		Drought resistance		Plant cover	
Type of vegetation	**Index**	**Vegetation types**	**Index**	**Types of vegetation**	**Index**	**Plant cover (%)**	**Index**
bare land	1	gardens, evergreen rangelands	1	gardens, evergreen rangelands	1	>35	1
annual agricultural crops, cereals, grassland, shrubland	1.5	rangelands, permanent grasslands	1.3	rangelands, permanent grasslands	1.3	10–35	1
gardens, evergreen rangelands	2	annual agricultural crops, cereals, annual grasslands	1.6	annual agricultural crops, cereals, annual grasslands	1.6	<10	2
		bare land	2	bare land	2		

d. Management quality index in the model

Land use		Policy enforcement		Livestock pressure	
Type	**Index**	**Degree of enforcement**	**Index**	**Livestock pressure**	**Index**
agriculture lands	1	Complete: >75% of the area under protection	1	<1	1
good to moderate rangelands	1.3	Partial: 25–75% of the area under protection	1.5	1–2.5	1.5
poor rangelands	1.6	Incomplete: <25% of the area under protection	2	>2.5	2
bare land	2				

e. Erosion state quality index in the model

Water erosion		Wind erosion	
Description	**Index**	**Description**	**Index**
very low	1	very low	1
low	1.2	low	1.2
moderate	1.5	moderate	1.5
severe	1.7	severe	1.7
very severe	2	very severe	2

f. Ground water state quality index in the model

EC (mmhos/cm)		CL (mg/l)		SAR		Water Table Depth (cm)	
Description	**Index**	**Description**	**Index**	**Description**	**Index**	**Description**	**Index**
250>	1	250>	1	10	1	315>	1
250–750	1.2	250–500	1.2	10–18	1.3	285–315	1.5
750–2250	1.5	500–1500	1.5	18–26	1.6	285>	2
2250–5000	1.7	1500–3000	1.7	26	2		
5000<	2	3000<	2				

The MMEDALUS procedure in the first stage is based on six quality indicators (climate, soil, vegetation cover, management, erosion state and ground water situation). These six major layers were derived from the sub-indicator layers that reflect individual conditions attributed to a specific value according to Table 1. Subsequently, the six quality layers were combined to give a single desertification sensitivity layer (Environmentally Sensitive Index-ESI). Table 1 shows the indicators selected and the values attributed. These quality layer were then used in the GIS procedure to assess the final desertification intensity. A quantitative classification with values between 1.0 and 2.0 was used throughout the model for the indicators as well as for the final classification of desertification intensities. Value "1" was considered to areas of minor sensitivity, and value "2" was considered in areas with the major sensitivity. Values between 1 and 2 reflect a relative vulnerability.

The different quality layers are assessed using the following equations based on geometric means to integrate the individual sub-indicator maps.

Soil Quality Index (SQI) = (texture × electrical conductivity × rock fragment × depth × slope × drainage × organic matter)$^{1/7}$

Climate Quality Index (CQI) = (rainfall × aridity × evapotranspiration)$^{1/3}$

Vegetation Quality Index (VQI) = (fire risk × erosion protection × drought resistance × vegetation cover)$^{1/4}$

Management Quality Index (MQI) = (land use × policy enforcement × livestock pressure)$^{1/3}$

Erosion state Quality Index (EQI) = (water erosion × wind erosion)$^{1/2}$

Ground Water state Quality Index (WQI) = (CL × EC × SAR × water table depth)$^{1/4}$

2.3 Description of ESI to desertification

Based on the data obtained from the applied methodology for defining the ESI map to assess desertification intensities in the MZJ plain, the various types and subtypes of ESI can be described as following in terms of land characteristics and management quality. According to their value, each of the six quality indices (MMEDALUS) were classified as high, moderate or low as shown in Tables 2. Finally, all six quality indices were combined to calculate a single index of desertification severity using the following equation:

Final Equation = (SQI × CQI × VQI × MQI × EQI × WQI)$^{1/6}$

The ranges of desertification intensity (ESI), for the four classes is illustrated in Table 3.

Tab. 2 Classification of six quality criteria used in the MMEDALUS Model and also the percent areas which belong to each class.

Kind of quality criteria	Class	Kind of quality	Range	Area (%)
Soil	1	High	<1.13	0
	2	Moderate	1.13 to 1.45	89.87
	3	Low	>1.45	10.13
Climate	1	High	<1.5	0
	2	Moderate	1.5	0
	3	Low	>1.5	100
Vegetation	1	High	1 to 1.13	0
	2	Moderate	1.13 to 1.38	0
	3	Low	1.38<	100
Management	1	High	1 to 1.25	0
	2	Moderate	1.26 to 1.50	19
	3	Low	>1.50	81
Erosion	1	High	1 to 1.25	43.68
	2	Moderate	1.26 to 1.50	34.72
	3	Low	>1.50	21.6
Ground Water	1	High	1.2>	0
	2	Moderate	1.–1.38	48.06
	3	Low	>1.38	51.94

Tab. 3 Classes of desertification intensity and corresponding range of indices (Sepehr et al. 2007).

Qualitative classes	Low	Moderate	Severe	Very severe
Quantitative classes	1–1.22	1.23–1.37	1.38–1.53	1.54–2

Note: In this research MEDALUS method (Masoudi and Zakerinejad 2010) was done based on its characteristics, too.

2.4 Testing the Models

In order to evaluate quantitatively the accuracy of the obtained maps (MEDALUS and MMEDALUS), the maps were compared to observed desertification intensities in the field (ground truth). The ground truth map was prepared based on agricultural production conditions. In this research we prepare the map of current state of desertification mapped based on the ratio between current production to potential production (FAO/UNEP 1984). A low ratio values reflect increasing desertification intensity. Finally the spatial distribution of the productivity ratio was compared to the MEDALUS and MMEDALUS results.

3. Results and discussions

We applied the two model approaches as described above to derive the spatial distribution of the desertification intensities. As shown in Table 2, most of the study area has moderate quality soils (89.87% of the area) in terms of desertification risk followed by low quality soils (10.13%). High quality soils were not documented in the area. Moreover, the majority of this area is characterized by low climate quality (100% of the area). This is mainly due to the relative low amount of precipitation occurring in the area and the high bioclimatic aridity index. All the different types of vegetation growing in the study area are characterized as low quality (100%). Additionally, the majority of this area is described by low land management quality (81%) especially in grazing areas. Most of the area shows a high erosion quality (43.68% of the area). Furthermore, the majority of this area is characterized by low ground water quality (51.94% of the area). The results also show that management, climate and ground water quality were the most important indicators affecting desertification process while soil quality seems to be less important in the study area (Figure 2).

Based on the above describe methodology different desertification maps (Figure 3) have been produced for the MZY Plain. The current desertification map based on FAO/UNEP, 1984 is shown in the same figures for comparison. The desertification map derived with the MMEDALUS model shows that most of the territory is classified as severe and very severe desertification (84%) (Figure 4). These critical areas exist somehow all over the study area and coincide with areas of greater human activity or severely eroded soils and low ground water quality. These critical areas need management initiatives that can effectively promote a slow regeneration of the landscape as a measure to combat desertification.

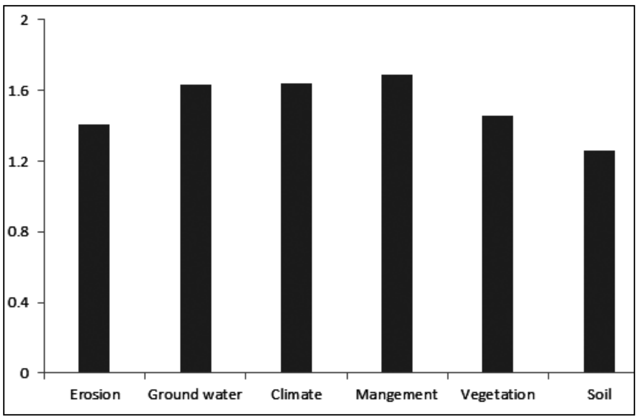

Fig. 2 Average of hazard degrees of main criteria used in the MMEDALUS Model.

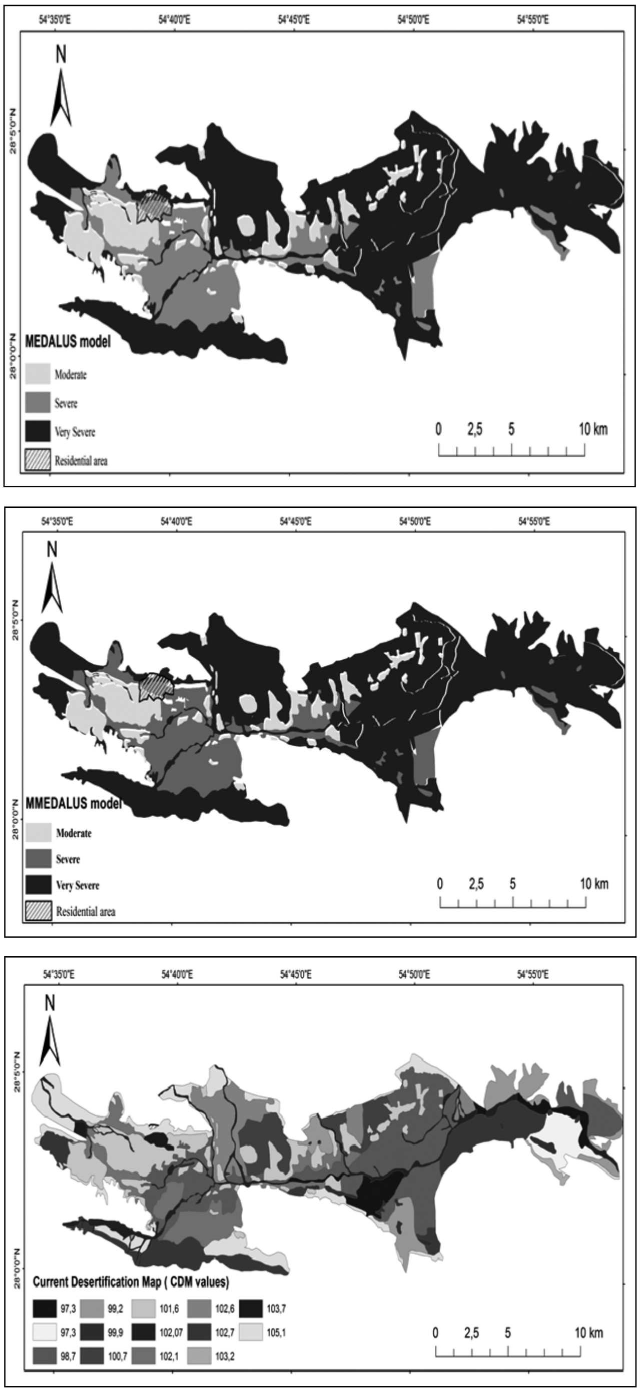

Fig. 3 Different desertification maps used for MEDALUS, MMEDALUS models and current desertification map.

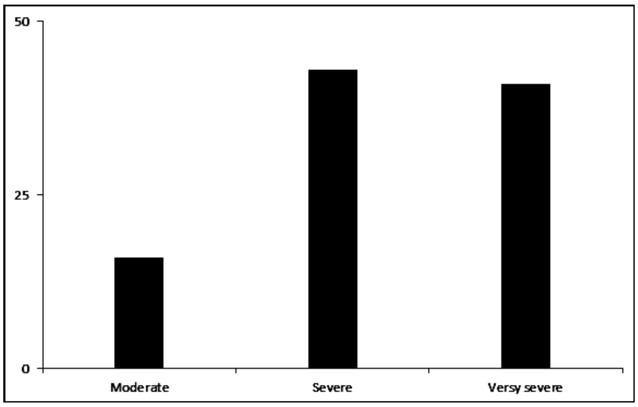

Fig. 4 Percent areas under hazard classes of desertification in the MZY Plain using MMEDALUS model.

In order to compare the different desertification maps with the agricultural productivity ratio (current desertification map), we use a correlation matrix. As illustrated in Table 4, there is an inverse correlation between current desertification map and the risk of desertification detected by MMEDALUS (0.05 level). It means by decreasing the ratio of current production to potential production the risk of desertification is increasing. While, there is no correlation between the reduction of production and the risk of desertification by MEDALUS in the study area. The 'ratio of current production to potential production' has been used to describe the current state of land degradation in the different models of land degradation or desertification assessment such as FAO/UNEP, GLASSOD and ASSOD. In both ASSOD and GLASSOD models, local experts assess 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. In other words, the estimates of ASSOD consider that a given amount of soil erosion is a more serious problem on poor and shallow soil than on a deep and fertile soil.

The developed model (MMEDALUS) attempts to assess and identify the factors affecting desertification. The results indicate the the MMEDALUS model outperforms the MEDALUS model in the study area. Results also revealed that the main reason for the better performance of the MMEDALUS assessment was to define a suitable criteria framework (e.g. adding ground water and erosion criteria). The result agree with other findings such as published by Sepehr et al. (2007) and Jafari, Bakhshandehmehr (2013).

4. Conclusion

In this study, MEDALUS and MMEDALUS models were mapped in the MZJ area. This MEDALUS model approach was adjusted to develop a regional model that could be adapted to the south and south west of Iran. The results obtained show that most of the area can be classified as having severe and very severe desertification susceptibility in accordance with the ground truth map of agricultural productivity ratio. Especially the MMEDALUS model shows a good accuracy with the ground truth map and hence, outperforms the MEDALUS model. The results also show that management, climate and ground water quality were the most important indicators affecting desertification process while soil quality seems to be less important indicators. The current state of the environmental conditions is caused by the arid climate, unproper management and also by the deterioration of the ground water. This highlights that particularly human activities play an important role in accelerating desertification process in the study area. The improvement of the MEDALUS approach by taking into account the groundwater and soil erosion state (MMEDALUS) generally allows for a better assessment of the desertification status of a landscape. However, in future studies we will focused on the integration of data related to water and soil salinity. In addition, considering some quantitative map of soil erosion may be useful for a further improvement of desertification modelling.

Finally, the limitation of desertification research assessment and monitoring in several developing countries (e.g., southeast of Iran, Iraq, Afghanistan, etc.) is the absence of historical data, including climatic, pedological (soil salinity, organic matter) information. Therefore applying simple models driven mainly by remote sensing data and available regional and national data is crucial to evaluate desertification on small spatial scales especially in remote areas.

References

Amiraslani, F., Dragovich, D. (2011): Combating desertification in Iran over the last 50 years: an overview of changing approaches. Journal of Environmental Management 92(1), 1–13, https://doi.org/10.1016/j.jenvman.2010.08.012.

Basso, F., Bove, E., Dumontet, S., Ferrara, A., Pisante, M., Quaranta, G., Taberner, M. (2000): Evaluating environmental sensitivity at the basin scale through the

Tab. 4 Correlation matrix between current desertification map and MMEDALUS and MEDALUS maps.

		MEDALUS	MMEDALUS
Ratio of current production to potential production	Pearson Correlation	−.029	−.185*
	Sig. (2-tailed)	.671	.033

* Correlation is significant at the 0.05 level.

use of geographic information systems and remotely sensed data: an example covering the Agri basin (Southern Italy). CATENA 40, 19–35, https://doi.org/10.1016/S0341-8162(99)00062-4.

Chenchouni, H. (2009): Integrating Biological Resources as Faithful Indicator in Climate Change Assessment: Case Study of Biodiversity Monitoring in eastern Algeria. Presented at the 3rd World Climate Conference, Geneva, Switzerland.

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

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

Fleskens, L., Stringer, L. C. (2014): Land management and policy responses to mitigate desertification and land degradation. Land Degradation and Development 25, 1–4, https://doi.org/10.1002/ldr.2272.

Eliasson, A., Rinaldi, F. M., Linde, N. (2003): Multicriteria decision aid in supporting decisions related to groundwater protection. Environmental Management 32, 589–601, https://doi.org/10.1007/s00267-003-2906-9.

Kosmas, C., Poesen, J., Briassouli, H. (1999): Key indicators of desertification at the Environmentally Sensitive Areas (ESA) scale. In: C. Kosmas, M. Kirkby and N. Geeson (Eds.), 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.

Lal, R. (1998): Soil erosion impact on agronomic productivity and environment quality: critical review. Plant Science 17, 319–464, https://doi.org/10.1016/S0735-2689(98)00363-3.

Ludwig, J. A., Wilcox, B. P., Breshears, D. D., Tongway, D. J, Imeson, A. C. (2005): Vegetation patches and runoff-erosion as interacting ecohydrological processes in semiarid landscapes. Ecology 86, 288–297, https://doi.org/10.1890/03-0569.

Masoudi, M., Jokar, P. (2017): A New Model for Desertification Assessment Using Geographic Information System (GIS): A Case Study, Runiz Basin, Iran. Polish Journal of Ecology 65(2), 236–246, https://doi.org/10.3161/15052249PJE2017.65.2.006.

Masoudi, M., Zakeri Nejad, R. (2010): Hazard assessment of desertification using MEDALUS model in Mazayjan plain, Fars province, Iran. Ecology, Environment and Conservation 16(3), 425–430.

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

Miao, L., Moore, J. C., Zeng, F., Lei, J., Ding J., et al. (2015): Footprint of Research in Desertification Management in China. Land Degradation and Development 26, 450–457, https://doi.org/10.1002/ldr.2399.

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, maps. Revised edition

Pan, J. H, Li, T. Y. (2013): Extracting desertification from LANDSAT imagery based on spectral mixture analysis and AlbedoVegetation feature space. Natural Hazards 25, 915–927, https://doi.org/10.1007/s11069-013-0665-3.

Ponce, H. R., Koohafkan P. (2004): Methodological framework for Land Degradation Assessment in Drylands (LADA). FAO report: Rome.

Shoba, P., Ramakrishnan, S. S. (2016): Modeling the contributing factors of desertification and evaluating their relationships to the soil degradation process through geomatic techniques. Solid Earth 7, 341–354, https://doi.org/10.5194/se-7-341-2016.

Sepehr, A., Hassanli, A. M., Ekhtesasi M. R., Jamali, J. B. (2007): Quantitative assessment of desertification in south of Iran using MEDALUS method. Environmental Monitoring and Assessment 134, 243–254, https://doi.org/10.1007/s10661-007-9613-6.

Taddese, Y. (2001): Land degradation: a challenge to Ethiopia. Environmental Management 27, 815–824, https://doi.org/10.1007/s002670010190.

UNEP (United Nations Environmental Program) (1992): World Atlas of Desertification, editorial commentary by N. Middleton and D. S. G. Thomas. Arnold: London.

UNCCD (2014): Desertification: The Invisible Frontline; UNCCD Publications: Bonn, Germany.

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.

Yang, D., Kanae, S., Oki, T., Koik, E. T., Musiake, K. (2003): Global potential soil erosion with reference to land use and climate changes. Hydrological Processes 17, 2913–2928, https://doi.org/10.1002/hyp.1441.

Zakerinejad, R., Maerker, M. (2015): An integrated assessment of soil erosion dynamics with special emphasis on gully erosion in the Mazayjan basin, southwestern Iran. Natural Hazards S25–S50, https://doi.org/10.1007/s11069-015-1700-3.

Zakerinejad, R., Maerker, M. (2014): Prediction of gully erosion susceptibilities using detailed terrain analysis and maximum entropy modeling: a case study in the Mazayejan plain, southwest Iran. Geografia Fisica e Dinamica Quaternaria 37(1), 67–76.

Zehtabian, G., Jafari, R. (2002): Evaluation of water resources degradation in Kashan area using desertification model. Journal of Environmental Studies 28, 19–30.

Morphostratigraphy of river terraces in the Eger valley (Czechia) focused on the Smrčiny Mountains, the Chebská pánev Basin and the Sokolovská pánev Basin

Břetislav Balatka[1], Jan Kalvoda[1, *], Tereza Steklá[1], Petra Štěpančíková[2]

[1] Charles University, Faculty of Science, Department of Physical Geography and Geoecology, Czechia
[2] Institute of Rock Structure and Mechanics, Academy of Sciences of the Czech Republic, Department of Neotectonics and Thermochronology, Czechia
* Corresponding author: kalvoda@natur.cuni.cz

ABSTRACT

The Eger (Ohře) River terraces originated in varied morphotectonic and climate-morphogenetic conditions that existed during the late Cenozoic evolution of the western part of the Bohemian Massif. In the area between the Smrčiny Mountains and the Sokolovská pánev Basin, these levels of the Eger River terrace system were identified (Table 1): the Pliocene terrace niveau B, the Cheb terrace (I), the Hradiště terrace (II), the Chvoječná terrace (III), Jindřichov terrace (IV), Nebanice terrace (V), Chocovice terrace (VI), Chotíkov Terrace (VII) and the recent flood plain (N). It was determined to be a morphostratigraphical system of 7 river terraces of Quaternary age. Older levels of fluvial sediments, occupying a still higher morphological position in the area between the Smrčiny Mountains and the Sokolovská pánev Basin, have been classified to the Pliocene. A comparison of terrace flights in the longitudinal profile of the Eger River between the Smrčiny Mountains and the Doupovské hory Mountains indicated that the Cheb terrace (I) in the Smrčiny Mountains is tectonically uplifted around 10 m in comparison with its level in the Chebská pánev Basin. In the Chlumský práh Horst area, the oldest Pleistocene terraces, which originated during the Tiglian stage, were uplifted by approximately 15 m. The Chebská pánev Basin originated at the intersection of the Eger rift and the Cheb-Domažlice fault zone and its river network is incised ca 40 m into the planation surfaces of the sedimentary basin. Both volcanic processes and frequent seismic activity in the region are associated with the Late Cenozoic tectonic movements. According to the current stratigraphical scheme of the Quaternary, the Eger terrace system was formed mostly by the Pleistocene (Table 2) during the Tiglian to the Weichselian stages.

KEYWORDS

river terraces; evolution of the Eger valley; neotectonic movements; Smrčiny Mountains; Chebská pánev Basin; Sokolovská pánev Basin

Received: 7 June 2019
Accepted: 20 November 2019
Published online: 20 December 2019

Balatka, B., Kalvoda, J., Steklá, T., Štěpančíková, P. (2019): Morphostratigraphy of river terraces in the Eger valley (Czechia) focused on the Smrčiny Mountains, the Chebská pánev Basin and the Sokolovská pánev Basin. AUC Geographica 54(2), 240–259
https://doi.org/10.14712/23361980.2019.21

1. Introduction

1.1 Theme and aims

The Eger River has a remarkable position among the streams of the Bohemian Massif because of the varied geological structure and palaeogeographical history of the whole catchment area. Conspicuous river valleys, terrace systems and related fluvial deposits originate in regions of neotectonic activity, combined with specific climate-morphogenetic processes (Balatka et al. 2015; Balatka and Kalvoda 2018). Geomorphological research on the Eger valley was aimed at landform evolution and morphogenetic evaluation of river terraces and further fluvial sediments. Particular attention was paid to the classification of the flight of Eger River terraces in the stratigraphical system of the Quaternary.

The aim of this research is to discover or verify locations of river terraces and other fluvial sediments in the Eger River valley between the Smrčiny Mountains and the Sokolovská pánev Basin (Fig. 1), to provide their morphogenetic evaluation and to suggest their correlation with the current stratigraphical classification of the Quaternary. The field research was focused on a documentation of the fluvial sediment exposures and a characterisation of the landform evolution in the Eger River valley and the surrounding area (Balatka and Kalvoda 2018). This geomorphological research confirmed the importance of the Eger terrace system in the assessment of the range of the Quaternary tectonic movements, especially in the area of the Chlumský práh Horst. The geomorphological analysis of the fluvial landscape in the Eger River valley enabled a creation of the longitudinal profile of the river terraces from the German border in the Smrčiny Mountains through Citice in the Sokolovská pánev Basin to Vojkovice in the Doupovské hory Mountains.

During the interpretation of the geomorphological analysis of the Eger River valley landforms, special attention was focused on the influence of neotectonic uplift on the terrace flights in the longitudinal profile. Primarily, the historical-genetic relationship between the older terrace levels evolved in the Chebská pánev Basin and the corresponding terrace levels in the Bohemian part of the Smrčiny Mountains and in the crystalline Chlumský práh Horst, which is asymmetrically extended in the Mariánské Lázně fault zone, was assessed.

The Eger River terraces originated during the varied morphotectonic and climate-morphogenetic conditions which existed during the late Cenozoic evolution of the western part of the Bohemian Massif. It is a substantial reason for a discussion about the morphostratigraphical correlation of the river terraces in the studied area of the Chebská pánev Basin and its neighbouring regions with the terrace system along the middle and lower course of the Eger River.

1.2 A brief review of earlier papers

The oldest works about the Eger valley were related to the Sokolovská pánev Basin, the Slavkovský les Mountains and the Doupovské hory Mountains (Wilschowitz 1917; Danzer 1922; Peter 1923). The Eger River terraces in the Chebská pánev Basin were addressed only by Engelmann (1920), who identified 5 terraces, whose relative surface heights above the valley floor reached 5 m, 10 m, 15 m, 20 m and 30 m. He also correlated the highest terrace flights in the Chebská pánev Basin with fluvial gravel in the Sokolovská pánev Basin at the relative height of 50–70 m and with the Eger terrace A in the lower valley of the Bílina River at the relative height of 170 m. However, this incorrect interpretation would imply that the oldest terrace of Eger would have a distinct divergence of 140 m in the upstream direction, which

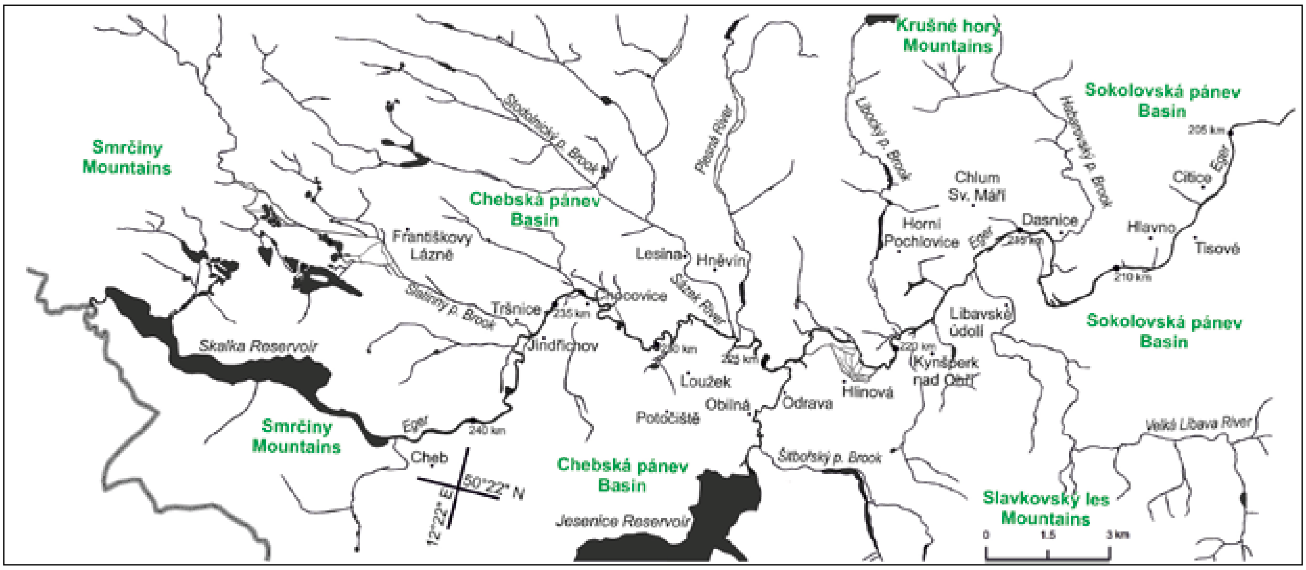

Fig. 1 Geographical position of the Eger (Ohře) valley between the Smrčiny Mountains and the Sokolovská pánev Basin.

is unrealistic. The studied area is also presented in Peter (1923), which describes 6 Eger River terraces between Kynšperk nad Ohří and Karlovy Vary. J. Peter identified the levels at the relative height of 50 m and 25 m as Quaternary terraces, while those at the relative height of 125 m, 100 m and 75 m as Pliocene terraces and the levels at the relative height of 175 m as terraces of Upper Miocene age. Nevertheless, the higher terrace levels at the relative height of 75 m to 175 m presented by Peter (1923) are in fact denudational plateaux and relics of local planation surfaces.

Complex geological research of the Chebská pánev Basin and parts of the Sokolovská pánev Basin, which took place in the second half of the 20th century, provided further substantial knowledge (Ambrož et al. 1958; Vrba 1959, 1981; Ambrož 1960; A. Kopecký 1960, 1966; Mazáč and Pokorný 1961; Kolářová 1965; Šantrůček et al. 1969, 1994). Ambrož et al. (1958) presented 10 terrace levels in the Chebská pánev Basin, whereas A. Kopecký (in Šantrůček et al. 1969) reported only 7 terrace levels in the same area, the highest one at the relative height of 30–35 m above the Eger valley floor. This author also correlated the highest terrace level with higher locations above the basin (relative height of 65 m) and in the area of former Chlum nad Ohří (relative height of 85 m). Their current position was explained by the young tectonic movements.

According to geological maps of the studied area (Škvor and Satran et al. 1974; Mlčoch et al. 1993; Müller et al. 1998), Eger fluvial sediments are of Pleistocene age. The complex studies of the Eger river terraces in the Chebská pánev Basin are presented by Kvaček (1987, 1989), who defined the localities of fluvial gravel-sand more accurately. This provided petrographic and granulometric analysis of the terrace sediments and documented them by several transverse profiles. Five terrace levels, the highest one in relative height of 17–25 m, were distinguished in this way. Their stratigraphical classification was derived from the older concepts mentioned above. Older publications concerning with the terrace system along the lower course of the Eger River are cited especially in the third chapter.

2. Methods

Geomorphological analysis of landforms was performed in focus on the palaeogeographical history of the studied region. Field works resulted in a detailed description of the identified localities with relics of fluvial sediments and in a set of graphical documentation. It especially concerns with a construction of a geomorphic sketch map, cross profiles and a complex longitudinal profile of the river terraces in the Eger valley. Geomorphological research was also supported by an evaluation of data presented in earlier regional papers about the fluvial accumulations.

The acquired data were used for the classification of the river terrace levels applying a method based on the reconstruction of the terrace levels in the longitudinal and transverse valley profiles. This complex method was already used by Q. Záruba in his work on the terrace system of the Vltava River (Záruba-Pfeffermann 1943; Záruba et al. 1977). The relics of the Eger River fluvial accumulations were recorded in the longitudinal profile, whose topographic base of the river level was measured in November 1949 with an average daily discharge of 7.32 $m^3 s^{-1}$ in Citice (Vodohospodářská kancelář ministerstva techniky v Praze 1950). The reconstruction of the terrace levels in the longitudinal profile was based on the methodological approach called the equilibrium profile (Krejčí 1939). This method of characterising the terrace system builds on the assumption that the palaeo-thalweg and the surfaces of each major terrace level maintained stable gradients that correspond to the longitudinal profiles (Balatka et al. 2015). In the presented longitudinal profile, relics of the original surface (not reduced by erosion) and the base of the river terrace correspond to the state of equilibrium of the valley floor evolution. It means that a stream of certain discharge does not erode nor accumulate, so the only process that takes place is a transport of carried material. This state may be disturbed – as a consequence of tectonic movements, changes in the discharge regime and sediment supply – in the direction of net erosion or in that of net accumulation.

The regional research was essentially conducted in accordance with the current state of the stratigraphical division of the Quaternary (e.g. Gibbard and Cohen 2008; Gibbard et al. 2010). Previously used names of glacial periods in the Quaternary are presented in this work only as a mandatory referential statement on the timing of the occurrence of certain fluvial accumulation according to older publications.

The geomorphological record of the Eger valley evolution and its river terraces was significantly disturbed by the current character of settlement and by an intensive economical exploitation of the area. For example, the fluvial relief of the western part of the Sokolovská pánev Basin was completely degraded by anthropogenic activity. Numerous locations of the river terraces were destroyed by brown coal mining. Therefore, the fluvial sediments were preserved only on the Paleogene sediments and crystalline rocks. Waste areas of the river terraces of the Middle Pleistocene age were extracted in the Chebská pánev Basin and the Mostecká pánev Basin.

3. Key morphological patterns of the Eger valley

The source of the Eger River is situated in the Smrčiny Mountains. The river enters Czechia at the river km 263 (measured from the Eger River mouth to the Elbe

Fig. 2 Western side of the Chebská pánev Basin (in a view from the north to the south) with mildly incised Eger valley and upper part of the Skalka dam. In the background the Smrčiny Mountains are seen. Photo: Petra Štěpančíková

River). It reaches the Františkolázeňská kotlina Basin near the mouth of the Libský potok Brook in the south-western corner of the Chebská pánev Basin (50°23′ N, 12°21′ E). Then to the confluence with the Elbe, the Eger River follows the significant depression of the Krušné hory Fault and the subsequent Dolnooharská tabule Plateau in the tectonically subsided zone of the Bohemian Cretaceous Basin at the base of České středohoří Mountains. The Eger River occupied the lowest parts of three basins situated at the base of the Krušné Hory Mountains, while simultaneously epigenetically eroded the neovolcanics in the Doupovské hory Mountains and the České středohoří Mountains as well as the uprising crystalline plate of the Chlumský práh Horst between the Chebská pánev Basin and the Sokolovská pánev Basin. Originally, the Eger River followed the Krušné hory fault in its full length, since the confluence with the Elbe River was situated in the place of the present-day Ústí nad Labem until the Middle Pleistocene (Balatka and Sládek 1962, 1976). The length of the Eger River currently reaches 302 km with a catchment area of 5 614 km^2. The confluence with the Elbe River is situated near Litoměřice, which is before the Elbe River antecedent valley and fault gap, which cuts through the neovolcanics of the České středohoří Mountains.

The Eger catchment belongs to the Saxo-Thuringicum, which is a part of the Bohemian Massif built mostly by the metamorphosed rocks and Variscan granitoids (Chlupáč et al. 2002). It includes metamorphosed Paleozoic rocks of the Thuringen-Vogtland, the crystalline complex of the Krušné hory Mountains and the Eger rift, Tertiary sediments and neoid volcanics (Mahel' et al. 1984). In the studied area, the Paleozoic rocks of the Thuringen-Vogtland are represented by quartz-mica schist, phyllitic mica schist with quartz layers, Cheb phyllite of Cambric and Ordovician age and biotite granite of Late Paleozoic age (Škvor and Satran et al. 1975; Müller et al. 1998). The Eger valley and its surrounding area in the Hazlovská pahorkatina Hills (west of Františkolázeňská kotlina Basin) are built by porphyric biotite granite and biotite hornstone. Phyllite and mica schist with layers of Vildštejn Formation protrude between Bříza and Hradiště (north-west of Cheb). The bedrock of the Chebská pánev Basin sediments is largely built by graywacke quartz-mica schists and white mica schists, which also surfaced in the Chlumský práh Horst and in the part of Eger valley leading to Hlavno.

Two depressions of the Eger Rift were created in the studied area, namely Chebská pánev Basin and Sokolovská pánev Basin. Chebská pánev Basin is filled by lacustrine and fluvial sediment of Eocene and Quaternary age (Fig. 2). Eocene clays, sand and gravel of the Starý Sedlec Formation fill the depressions of the crystalline complex and granitoids, which were affected by fossil weathering. The younger Nový Sedlec Formation is mostly of Oligocene age. The upper part of these layers contains volcanogenic sediments and lava bodies of olivine basalts (Ambrož 1958; Václ 1979). The Sokolov Formation was formed in the Lower Miocene. Its middle part contains a brown coal

Fig. 3 Landscape in the south-western part of the Chebská pánev Basin in a view from the Zelená Hora Mountain. Afforested ridges of the Slavkovský les Mountains (in the background) are delimited by fault- and denudation slopes along the Mariánské Lázně fault zone.
Photo: Petra Štěpančíková

seam, which can be several tens of meters thick. The thickness of the upper part of the Sokolov Formation reaches up to 170 m (Cypris Formation) and it is built mostly by bitumen clays and sandstones.

The Chebská pánev Basin is situated on the crossings of the Eger rift and the Cheb-Domažlice graben. It consists of the Tertiary sediments, whose thickness reaches up to 400 m and it is significantly disrupted by many faults of several tectonic systems, such as Krušné hory system, Sudety system, Český les system and Jizera system (Václ 1979; Dobeš et al. 1986). Earthquake epicentres are closely connected to this area, which is the most seismically active region of the Bohemian Massif (Babuška et al. 2010). The western boundary between the Chlumský práh Horst and Chebská pánev Basin is created by the fault-slope of Mariánské Lázně fault zone (Fig. 3). The vertical range of the Cenozoic tectonic movements along the Mariánské Lázně fault, namely the subsidence of the Chebská pánev Basin and the uplift of the Chlumský práh Horst, is assumed to be 300–400 m (Malkovský 1976, 1979). The analysis of the neotectonic evolution of the Chebská pánev Basin by Peterek et al. (2011) confirmed the significant role of the Upper Pliocene and Quaternary tectonic movements.

In the Chebská pánev Basin, the lacustrine and fluvial sedimentation of the clay, sand and gravel of the Vildštejn Formation took place in the Pliocene and the Lower Pleistocene (4.5–1.5 Ma, Špičáková et al. 2000). According to the paleoflora analysis by Bůžek et al. (1985), the prevailing climate of the last lacustrine sedimentation shows the transition between the warm temperate (mean annual temperature 12–14 °C) and cold temperate climatic zone (6–7 °C). A diatreme with pipe filling near Podhrad originated in the Upper Pliocene. The younger parts of the Komorní hůrka Hill (503 m), which is built by the pyroclastic rocks and effusion of melilitic olivine nephelinite, are of Pleistocene age (L. Kopecký 1978; Shrbený 1982; Gottsmann 1999), and are the same age as the fluvial sediments involved in this study. The Quaternary age of the youngest active phases of the Komorní hůrka Hill was confirmed by radiometric dating of its volcanic rocks, namely 0.85 ± 0.1 Ma up to 0.26 ± 0.05 Ma (Šibrava and Havlíček 1980) and 0.45–0.90 Ma (Wagner et al. 1998).

The Sokolovská pánev Basin has undergone a similar morphostructural evolution as the Chebská pánev Basin, since they used to be connected. In its longitudal direction, the Sokolovská pánev Basin is enclosed by the significant fault-slopes of the Krušné hory Mountains (Krušné hory Fault) and Slavkovský les Mountains (Eger fault), which define the edges of the Eger Rift. Separation of these basins was caused by the Neogene uplift of the Chlumský práh Horst. Thus, the Chebská pánev Basin was occupied by the paleo-lake in the Pliocene, while the Sokolovská pánev Basin was no longer occupied by it. The surface of the Sokolovská pánev Basin is also built by the oldest sediments of Staré Sedlo Formation as well as by the neovolcanic rocks (L. Kopecký 1978, 1985; Malkovský 1979, 1980). Conspicuously varied relief of the Sokolovská pánev Basin was caused by the numerous tectonic outcrops of kaolinically weathered granitoid bedrock. In the northern edge of the Slavovský les Mountains (between Loket and Doubí), the Eger River created an epigenetic and antecedent canyon.

4. Terrace system of the Eger River

One Pliocene terrace of niveau B and 7 terrace levels of the Quaternary age (Fig. 4) were discovered by the geomorphological analysis of the Eger River valley as well as by the evaluation of the previous research data and by the reconstruction method, which was used to assess the terrace system. The morphostratigraphical classification of the river terraces (Tables 1 and 2, Figures 5 and 6) was based on the parallel nature with the terrace system of the middle and lower Eger (Balatka and Sládek 1976; Balatka 1993; Tyráček 1995; Tyráček et al. 1985, 2004) and also on the correlation of the fluvial accumulations in the studied area with the Quaternary-geological system

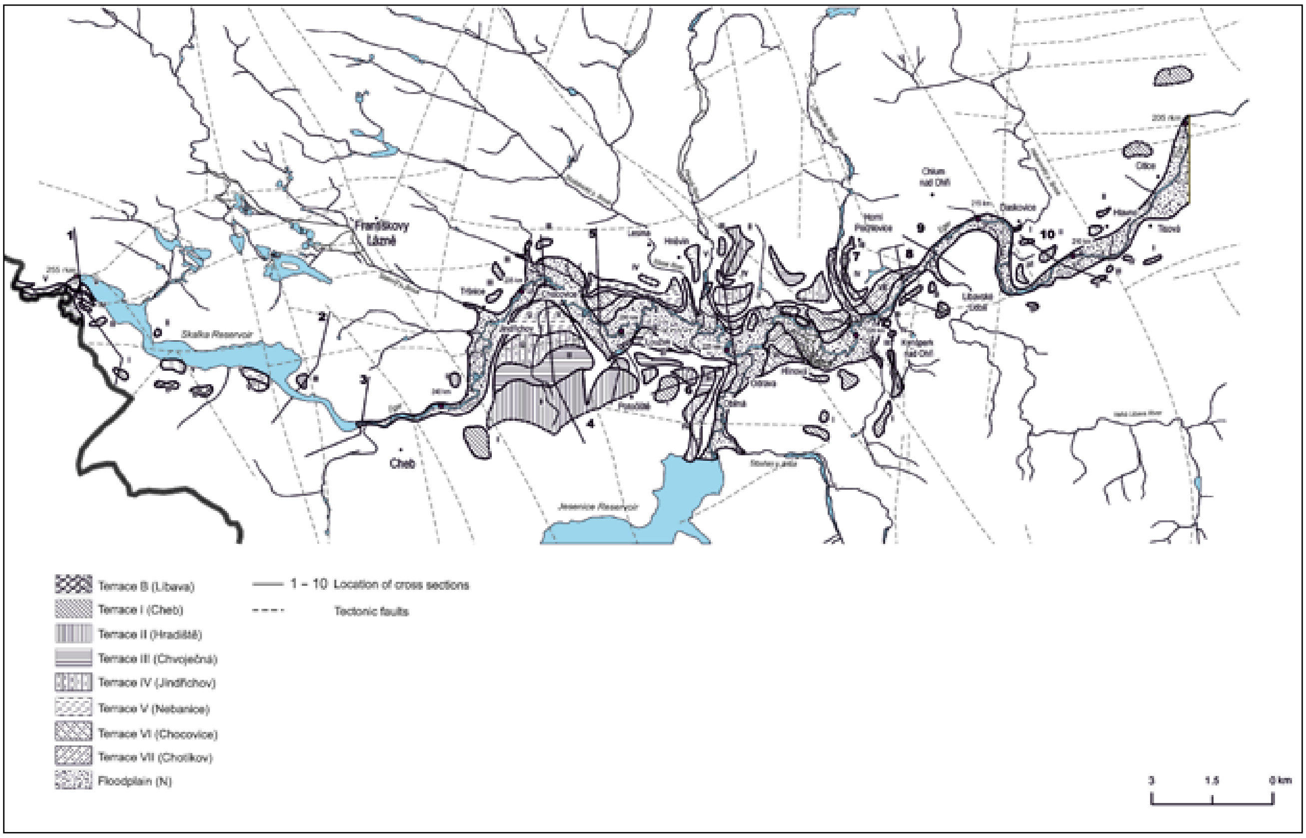

Fig. 4 Geomorphological sketch map of the Eger River terraces between the Smrčiny Mountains and the western part of the Sokolovská pánev Basin.

(e.g. Ambrož et al. 1958; Šantrůček et al. 1994; Balatka and Kalvoda 2008; Tyráček and Havlíček 2009; Balatka et al. 2010a,b, 2015).

The dependence of the terrace system structure on the morphotectonic and lithological conditions of the area, through which the river flows, can be clearly observed on the Eger River valley landforms. Since the Pliocene, the Eger River valley has been evolving in various kinds of relief, such as lowland, basin, upland, highland or mountain type of relief.

Neotectonic processes of the studied area took place in the Neogene and the Pleistocene, which was proven by the tectonic deformation of the Upper Pliocene – Lower Pleistocene Vildštejn Formation as well as by the structure and flights of the older terrace levels in the Eger River longitudal profile (Fig. 7).

The Eger River flows through the studied area from the state border between the mouth of the Hraniční potok Brook and the vicinity of Citice in the length of 60 km (between river km 266 and 206), while crossing three geomorphological areas of the Krušné hory sub-province: the Smrčiny Mountains, the Chebská pánev Basin and the Sokolovská pánev Basin. The Eger River enters the rugged relief of the Smrčiny Mountains between the Hraniční potok Brook and the Libský potok Brook in the Hazlovská vrchovina Hills (Balatka and Kalvoda 2006; Demek and Mackovčin et al. 2006). In this area, the Eger River created a valley with a wide floodplain and meandering channel (Fig. 2). However, it is presently hidden under the surface of the Skalka dam. The narrow erosion valley that was created by the Eger River immediately above Cheb is ca 85 m deep. On the other hand, the Eger River flows through a wide shallow valley with extensive flood plain and meandering channel in the Chebská pánev Basin between Cheb (river km ca 240) and the mouth of the Libava River (river km 216.8). This part of the Eger valley is incised 30–40 m into the planation surface of the oldest terrace, into that of the other fluvial deposits (Fig. 3) and also into the Vildštejn Formation of Pliocene and Lower Pleistocene age.

Between the mouth of the Libava River and Černý Mlýn (river km 209.2), the Eger River cuts into the crystalline complex of the Chlumský práh Horst creating a deep (ca 85–155 m) and asymmetrical valley. This epigenetic and antecedent valley intersects the morphologically significant zone of the Mariánské Lázně fault in the wider vicinity of Kynšperk nad Ohří. Near Černý Mlýn, the Eger River reaches the lower relief of the Tertiary sediments of the Sokolovská pánev Basin, namely the unit of the Svatavská pánev Basin. In the studied part of the Eger valley between the river km 265 and 148 (compare Figures 1, 4 and 7), the relics of Pliocene terrace of niveau B and heterogenous group of the Quaternary terraces were identified and documented. Their vertical distribution in the Eger River valley and morphostratigraphical classification is elaborated in Tables 1 and 2.

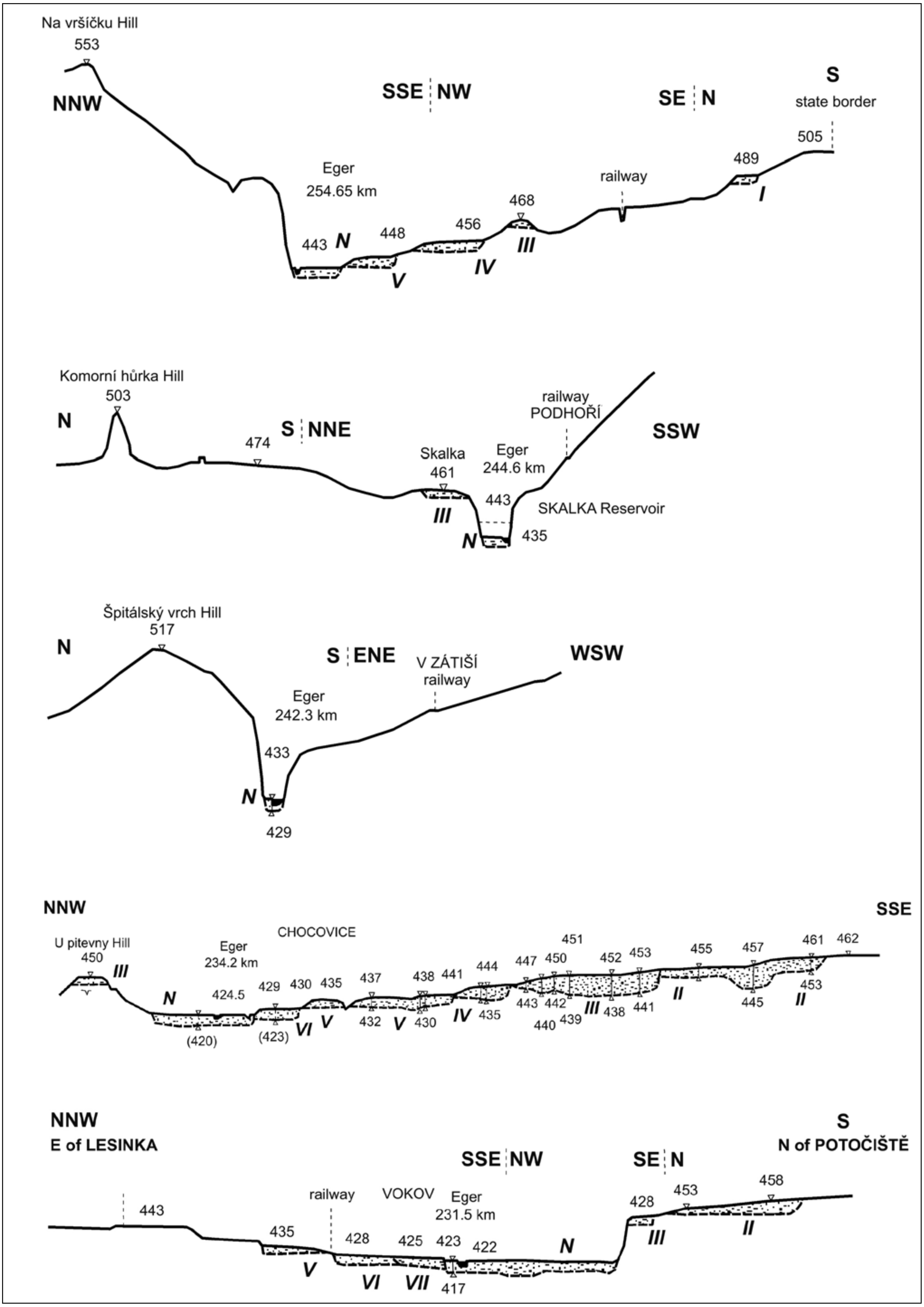

Fig. 5 Cross sections of the Eger valley and fluvial deposits between river km 254.65 and 231.50 (No. 1–5). Locations see in Fig. 4.

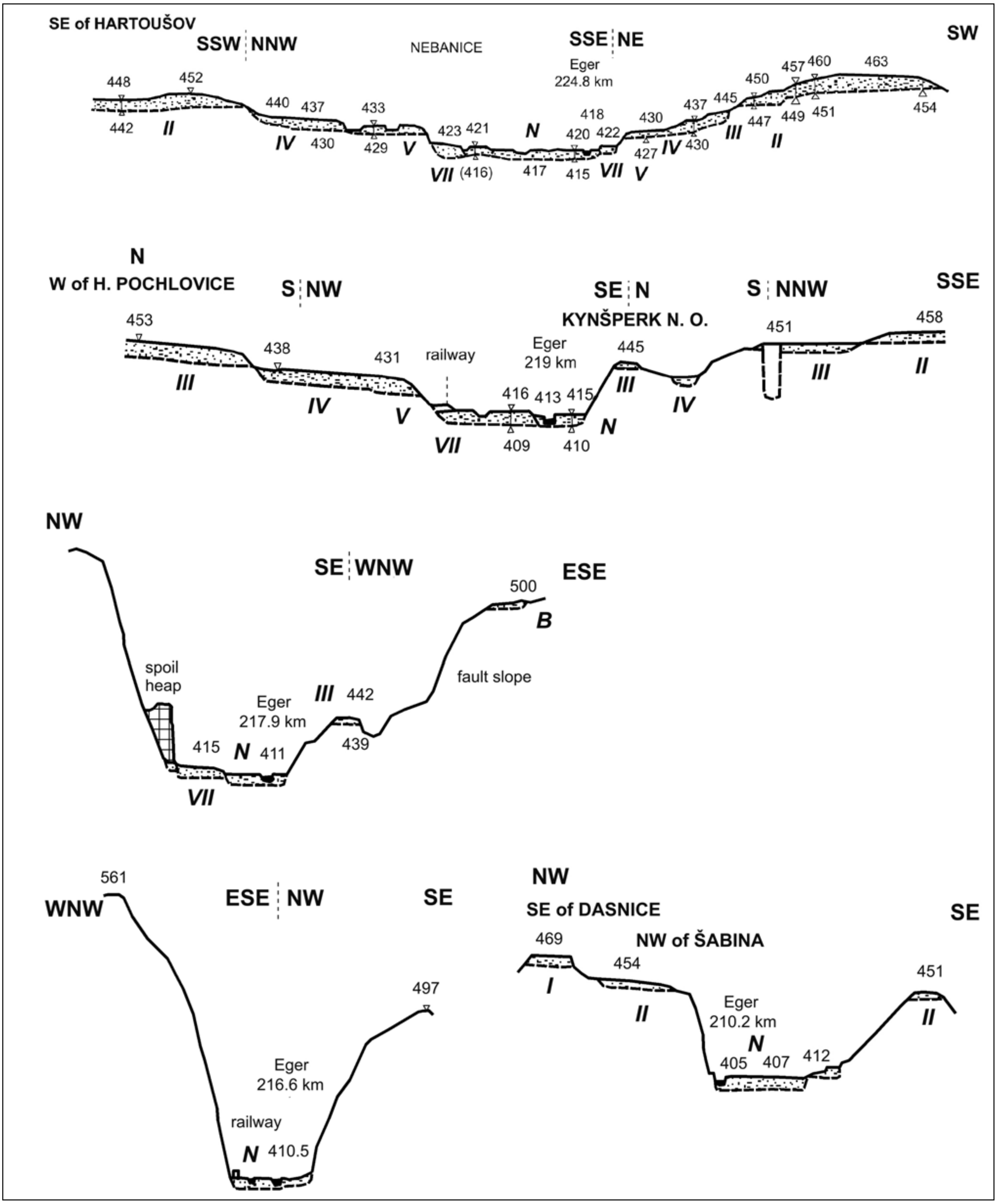

Fig. 6 Cross sections of the Eger valley and fluvial deposits between river km 224.80 and 210.20 (No. 6–10). Locations see in Fig. 4.

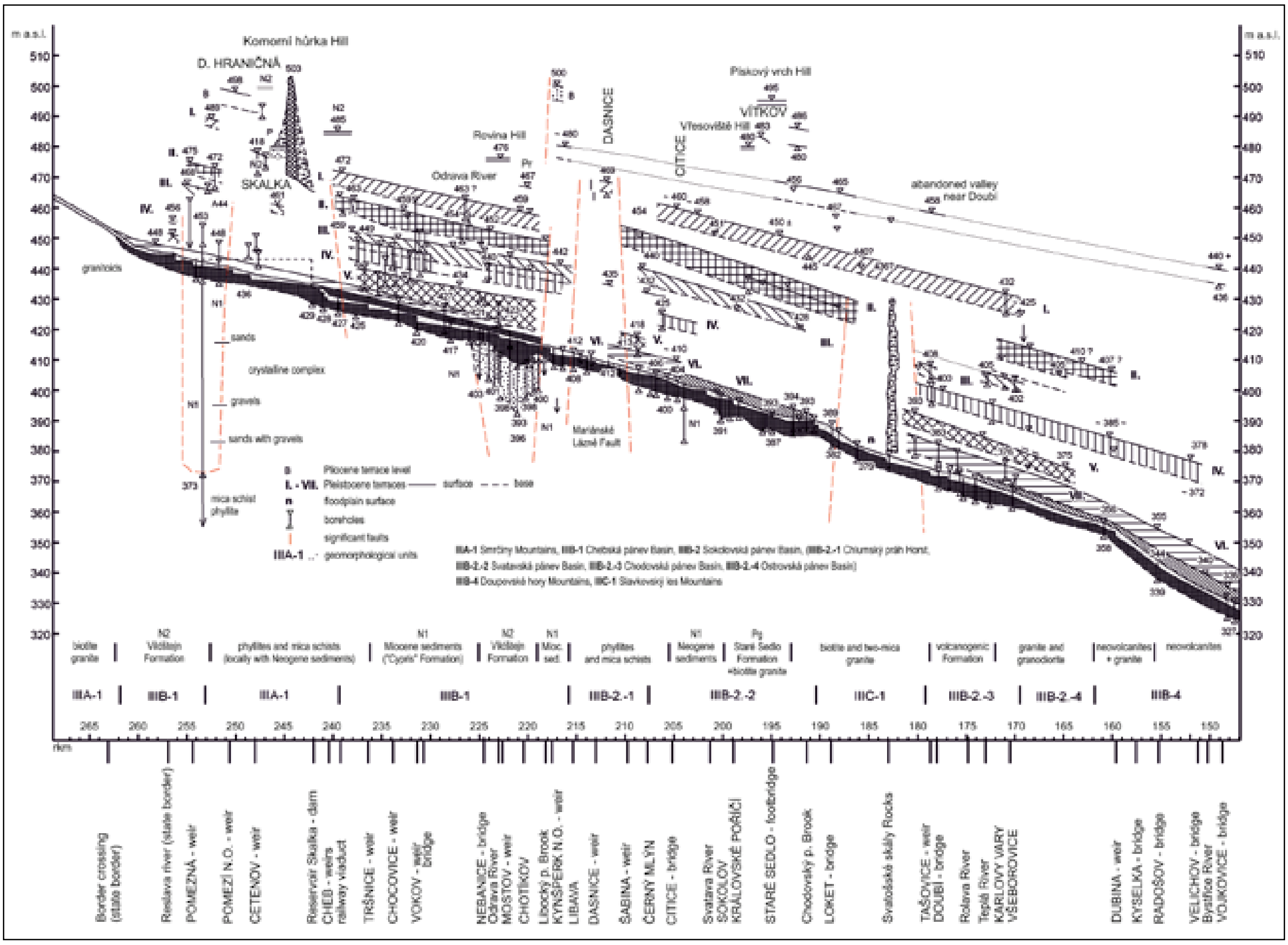

Fig. 7 A longitudinal profile of the Eger River terraces between the Smrčiny Mountains and the Doupovské hory Mountains.

4.1 The Pliocene terrace niveau B

Several small localities of the oldest and the locally highest terrace niveau B were identified. Its name and stratigraphical classification in the Pliocene correspond with the terrace niveau B in the Mostecká pánev Basin at the foot of the Doupovské hory Mountains (Balatka and Sládek 1976). The two highest localities of the terrace niveau B (in the direction against the Eger River flow) are situated near Dolní Hraničná. The first one was found on the right valley slope in Dolní Hraničná and the second one is 500 m east of the same village. The relic of the 2 m high terrace gravel covers the moderately angled slope at 492–498 m a.s.l. (56–61 m above the former Eger River level, Fig. 8). Mlčoch et al. (1993) and Müller et al. (1998) classified these gravel localities to the Günz. The bedrock of the fluvial sediments is formed by quartz-mica schists and by the sediments of the Vildštejn Formation.

Another locality of the niveau B was found in Chlumský práh Horst, north-east of the road from Kynšperk nad Ohří to the Libava River valley. This small moderately angled platform consists of quartz-mica schist bedrock, which is covered by coarse sub-angular gravel and sharp-edged fragments of quartz, quartzite, phthanite and brownstone (15–30 cm) with sporadic appearance of well rolled boulders. A larger block of quartzite (2.0 × 1.2 × 0.3 m) was also found, as well as the crystalline bedrock that was uncovered in the 3m deep exposure. This locality is situated at 495–500 m a.s.l., which corresponds with the relative height of 83–88 m above the Eger River. The excavation described by Kvaček (1987, 1989) consist of a 1.5m thick layer of sandy (loess?) loam covering a 1.5 m thick layer of clay gravel, which was situated on the top of the clay regolith of the crystalline complex. The sediments discovered in this location are alluvial rather than terrace sediments and were probably accumulated by the Libava River. Šantrůček et al. (1994) presented these sediments as "the fluvial sand gravel of the Upper Pliocene age", while Kvaček (1987, 1989) was of the opinion that the same sediments (with a question mark) originated in the Donau glacial stage, so that they are of Pleistocene age. However, if it really was the oldest Pleistocene terrace (I), it would have been uplifted by ca 50 m during the Quaternary. According to Šantrůček et al. (1994), the small accumulation of gravel located above the left slope of the Libava River valley at 480 m a.s.l. (relative height of ca 70 m) and an even smaller gravel accumulation ca 1 km south-west of Šabina at 475–480 m a.s.l. (relative height of 70–75 m) are also of Upper Pliocene age.

Table 1 Vertical distribution of Neogene fluvial sediments and Quaternary terraces (surface/base in metres above river bed) in the Eger River valley, Czechia.

Neogene sediments and river terraces	Chebská pánev Basin Balatka et al. (this paper)	Chlumský práh Horst Balatka et al. (this paper)	Doupovské hory Mountains Balatka (1993)	Mostecká pánev Basin (western part) Balatka (1993)	Mostecká pánev Basin (Žatec area) Balatka and Sládek (1976)	Bílina River (Hostomice area) Balatka (1995)	Elbe River (Ústí nad Labem area) Balatka (1995)
A	–	–		116/110	197/190		
B	–	88/~83		122/125	–		
I	43/~38	61/58	67/63	71/65	I_1 124/118 I_2 116/107 I_3 103/96	I_3 79/77	I_1 138/128 I_2 119/115
II	35/30	48/~45	II_1 58/53 II_2 50/47	II_1 64/58 II_2 62/58	II_1 96/90 II_2 89/84 II_3 83/79	II_1 72/65	II_2 93/77
III	25/19	30/~24	III_1 41/– III_2 36/33	III 47/43	III_1 70/63 III_2 66/59	III_1 55/53 III_2 43/36	III_1 78/58 III_2 70/58
IV	21/14	26/~20	IV_1 29/24 IV_2 22/–	IV_1 40/36 IV_2 39/34 IV_5 20/15	IV_1 60/56 IV_2 55/51 IV_3 51/45 IV_4 48/45 IV_5 45/41	IV_1 27/23 IV_2 21/18 IV_5 17/10	IV_1 51/42
V	13/6	13/6	15/–	V_1 17/9 V_2 14/9	V_1 37/30 V_2 33/30	V_1 (~10/– V_2 5/−2	V_1 38/23 V_2 33/23
VI	6/~0	7/3	7/0	VI_1 12/6 VI_2 8/1 VI_3 5/1	VI_1 26/20 VI_2 23/17 VI_3 20/15 VI_4 17/12	VI_1 3/−6	VI_1 20/12 VI_2 16/11
VII	3.5/−3	~4/−4	–	–	VII_1 14/10 VII_2 8/2 VII_3 6/2	–	VII_1 8/−4 VII_2 4/−14
N flood plain	1.5/−3	1.5/−4	2/−5	2–3/−4	3/−3		

Therefore, the sediment relics of the oldest terrace niveau B were preserved only in the morphostructural zones of the Smrčiny Mountains and the crystalline Chlumský práh Horst offsets, which have been uplifted during the Upper Cenozoic. The locations of the terrace niveau B in the Smrčiny Mountains with their surface at 61 m above the river level are almost at the same level as the highest location of the Vildštejn Formation in the surrounding area. The surface of the terrace niveau B is ca 10 m higher than the surface of the oldest Quaternary terrace (Cheb terrace, I) and also ca 10 m higher than the highest planation surface of the Vildštejn Formation in the Chebská pánev Basin around the Eger River valley. The top of the Pleistocene volcano Komorní hůrka Hill is situated slightly higher than the surface of the terrace niveau B (503 m a.s.l., Fig. 9). The Chlumský práh Horst is another significant area of the terrace niveau B. Besides the already described location of the highest level with its surface at 88 m above the river level (Fig. 10), many other small accumulations of these fluvial sediments were preserved here. The surface of these relics is up to 20 m lower than the highest levels and their body consists of a coarse clay gravel with imperfectly rolled clasts. The highest situated accumulation of the terrace niveau B in the area of the Chlumský práh Horst is located ca 25–30 m above the highest levels of the terrace niveau B that was formed on the sedimentary rocks of the Vildštejn Formation.

4.2 The Quaternary terraces

The Eger River has formed only sporadic and small terrace sediments in the Smrčiny Mountains. On the other hand, it has created large and continuously developed terraces in the Chebská pánev Basin. In the tectonically uplifted morphostructure of the Chlumský práh Horst, the rare relics of the higher (and therefore older) terrace levels exhibit a significant increase of its relative heights in comparison to those of the Chebská pánev Basin. Geomorphological analysis of the Eger River valley fluvial landscape, using the reconstruction method in the valley profiles, along with the previously published data, were used to assess the terrace system with these levels (Figures 4, 7 and Table 1): Terrace I (Cheb), Terrace II (Hradiště), Terrace III (Chvoječná), Terrace IV (Jindřichov), Terrace V (Nebanice), Terrace VI (Chocovice), Terrace VII (Chotíkov) and N (recent flood plain).

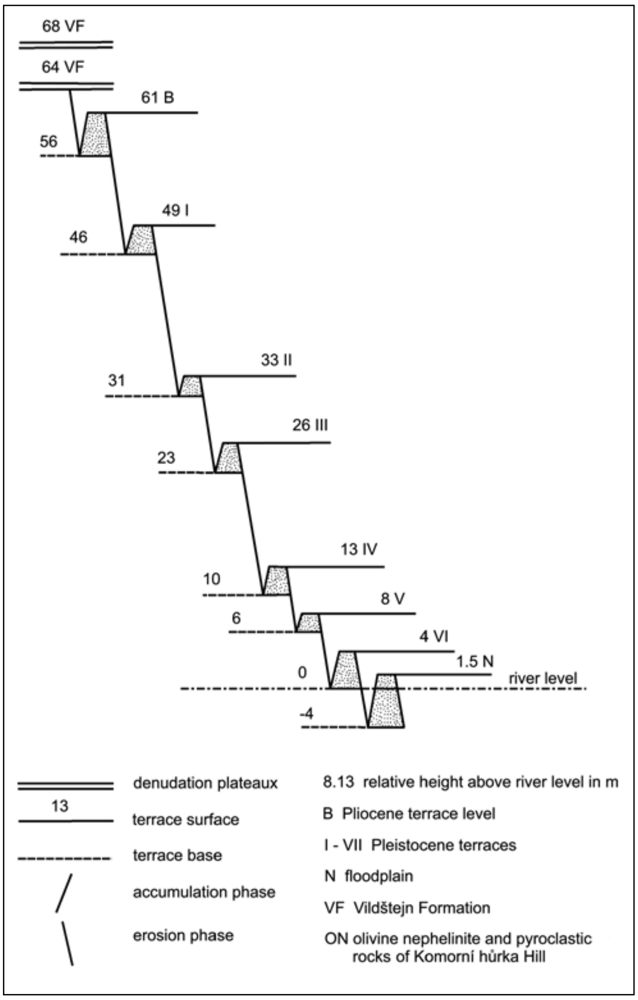

Fig. 8 Graphical scheme of successive incision of the river during the Eger valley evolution in the Smrčiny Mountains and the Františkolázeňská kotlina Basin (the section between river km 257–251).

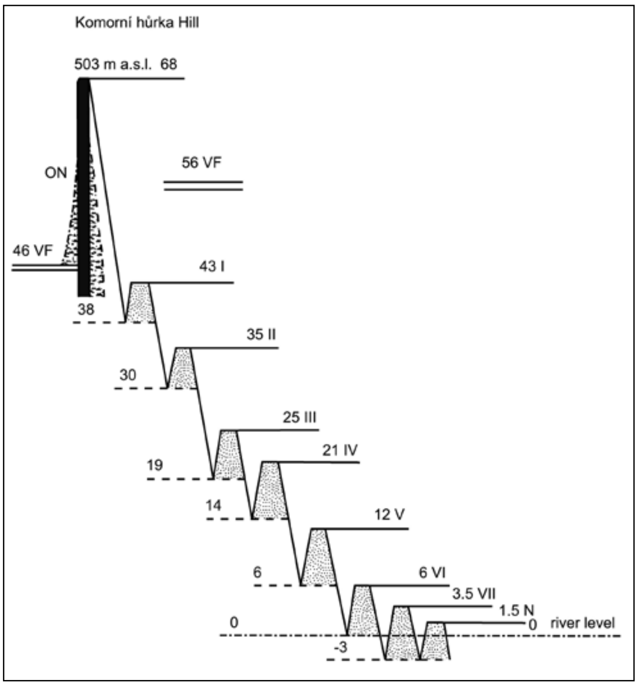

Fig. 9 Graphical scheme of successive incision of the river during the Eger valley evolution in the western part of the Chebská pánev Basin (river km 245–237). Explanations see in Fig. 8.

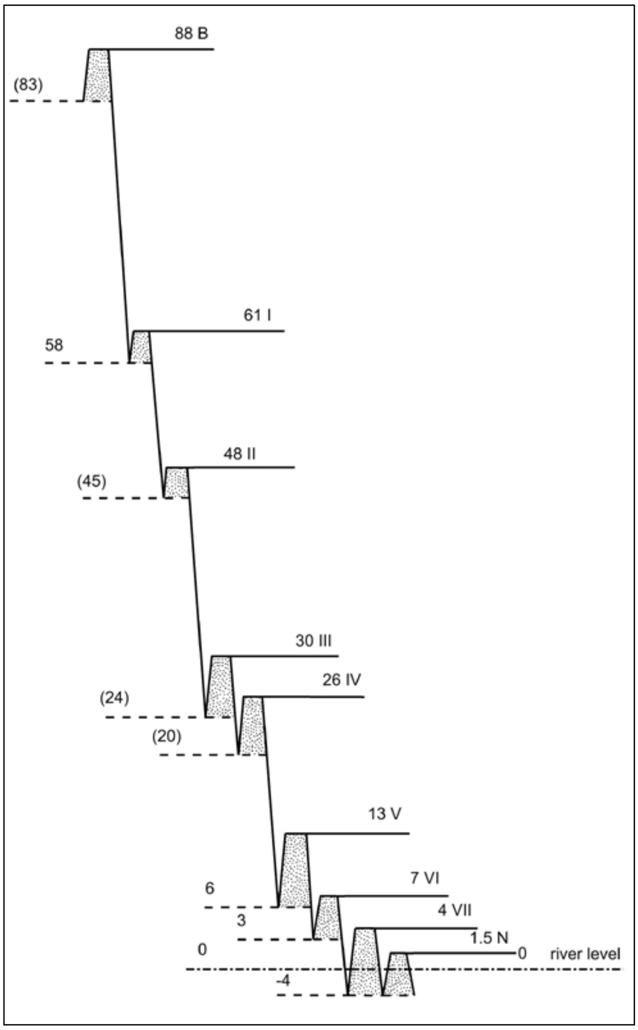

Fig. 10 Graphical scheme of successive incision of the river during the evolution of Eger valley in the Chlumský práh Horst (river km 218.5–209.7). Explanations see in Fig. 8.

Terrace I (Cheb)

In the Chebská pahorkatina Hills (which is a part of the Smrčiny Mountains), a small slightly angled platform west of Hraničná with its surface close to the 490 m contour line (thus at 49 m above the Eger River level) was classified as Terrace I (Fig. 8). This relic of the gravel is 2 m thick (Ambrož et al. 1958) and it covers the sediments of the Vildštejn Formation. Mlčoch et al. (1993) classified these sediments to the Günz. In the western part of the Chebská pánev Basin, another location of Terrace I was found near the railway station, on the platform located at 470–472 m a.s.l. (relative height of 41–42 m), which was also classified to the günz by Šantrůček et al. (1994). This terrace level also includes 5 m high gravel on the significant platform north-west of Potočiště at 460–463 m a.s.l., thus at 40–43 m above the river. Furthermore, two small accumulations of gravel are situated at the Terrace I level near the Hlínová and Dobříš at 455–459 m a.s.l., whose bedrock consists of the Vildštejn Formation sediments.

In the Chlumský práh Horst area, a small dorsal location 1 km south-east of Dasnice with its surface at 469 m a.s.l. is classified as Terrace I. Coarse clay sand with sub-angular boulders and fragments of quartz, strongly weathered phylites, mica schists and gneiss, quartz and ferric sandstone and conglomerate were found it the 3 m deep sand pit. Assuming the continuous flight of the Eger River terraces in its equilibrium longitudinal profile, this accumulation of gravel is located 15–20 m above the Terrace I level.

Presented locations of Terrace I including the alluvial gravel of the same level are situated almost at the same altitude as the planation surface of the Vildštejn Formation sediments (Upper Pliocene – Lower Pleistocene). The accumulation of Terrace I sediments probably occurred after the retreat of the Pliocene-Lower Pleistocene paleo-lake. Therefore, this Cheb Terrace is classified to the Tiglian stage of the Pleistocene (Table 2). According to Škvor and Satran et al. (1974), these accumulations outside of the Chebská pánev Basin (using the then stratigraphical classification) are of Günz – Donau age. A. Kopecký (in Šantrůček et al. 1969) classified the highest terrace accumulations east of Cheb to the old Donau – oldest Günz. In the Sokolovská pánev Basin, the sand gravel accumulations on the left bank of the Eger River between Citice and Svatava with their surface at 455–460 m a.s.l., thus in a slightly lower position (ca 5 m) than the accumulation near Dasnice, are

Table 2 Morphostratigraphy of terrace deposits of the Eger River (the Chebská pánev Basin and neighbouring regions) and their correlation with fluvial sediments in the middle and lower located sections of the Eger valley up to the Elbe River confluence area.

European stratigraphical stage divisions of the Quaternary (Gibbard and Cohen 2008; Gibbard et al. 2010)	EGER Chebská pánev Basin and neighbouring regions (Balatka and Kalvoda 2018; Balatka et al., this paper)	EGER Doupovské hory Mountains (Balatka, 1993) and Mostecká pánev Basin (Balatka and Sládek 1976; Balatka 1995)	EGER Mostecká pánev Basin (Tyráček 1995, 2001; Tyráček et al. 2004; Tyráček and Havlíček 2009)	ELBE Ústí nad Labem area (Balatka 1995; Balatka and Kalvoda 1995; Kalvoda and Balatka 1995)
Holocene Late Pleistocene Weichselian 0.12 Ma	Alluvial deposits Chotíkov Terrace (VII)	Rock bottom and alluvial deposits VII_3–VII_1	Thalweg gravel VII_3–VII_1	Rock bottom and alluvial deposits VII_2–VII_1
Middle Pleistocene Saalian (Warthe, Drenthe) 0.20 Ma	Chocovice Terrace (VI)	VI_4–VI_1	VI_4–VI_1	VI_2–VI_1
Middle Pleistocene Saalian (Wacken, Fuhne) 0.38 Ma	Nebanice Terrace (V)	V_2–V_1	V_2–V_1	V_2–V_1
Middle Pleistocene Elsterian Cromerian 0.78 Ma	Jindřichov Terrace (IV)	IV_5–IV_1	IV_5–IV_3 IV_2–IV_1	IV_1
Early Pleistocene (Bavelian, Menapian, Waalian, Eburonian) 1.78 Ma	Chvoječná Terrace (III)	III_2–III_1	III_2– III_1 II_3	III_2–III_1
Early Pleistocene Tiglian 2.58 Ma	Hradiště Terrace (II) Cheb Terrace (I)	II_2–II_1 I_4–I_1	II_2–II_1 I_4–I_1	II_1 I_2–I_1
Neogene Pliocene 5.30 Ma Miocene	Niveau B	Niveau B Niveau A		

presented by Škvor and Satran et al. (1974) also as Terrace I. Nevertheless, they are currently covered by the spoil heaps of the coal mines.

Terrace II (Hradiště)

The highest location of Terrace II (in the direction against the Eger River flow) was discovered on the small residual platform at 469–472 m a.s.l. (up to relative height of 32 m), which is covered by 1–4 m thick sand gravel. According to the borehole data, these deposits are situated on the weathered phyllites (Rousek 1960). The Terrace II has developed continuously between Cheb – Hradiště and the Odrava River valley south-east of Loužek in the Chebská pánev Basin. The surface of Terrace II is situated at 463 m and its base at 458 m a.s.l. near Hradiště, while the surface of this terrace east of Dolní Dvory is at 460 m and its base at 455 m a.s.l. and the surface of that south of Loužek is at 455 m and its base at 450 m a.s.l. South and south-west of Chvoječná, the surface of Terrace II is located at 455–460 m a.s.l. The thickness of its gravel varies mostly between 2.5–6 m. The terrace surface descends slightly to the north and gradually blends with the surface of Terrace III.

Terrace II surface was found at 458–453 m a.s.l. near Potočiště. According to Kvaček (1987), these sediments belong to the Vildštejn Formation. On the other hand, Šantrůček et al. (1994) classified sand gravel found in these locations as the older Mindel. The locality near Loužek, whose surface is situated at 450–455 m a.s.l., is the last occurrence of Terrace II in the eastern part of the Chebská pánev Basin. Alluvial clayey sand gravel (2–5 m thick) south of Hartoušov with the surface at 448–452 m a.s.l. corresponds to the level of the Eger Terrace II. The localities of the Hradiště Terrace in the Chebská pánev Basin were presented by Kvaček (1987) mostly as of Günz age. While according to A. Kopecký (in Šantrůček et al. 1969), these locations of Terrace II are of Lower Pleistocene age.

Sand gravel 0.5 km north-west of Šabina at 454–450 m a.s.l. (relative height of 48–44 m), which is presented as of Günz age by Šantrůček et al. (1994), is assigned to Terrace II in the area of the Chlumský práh Horst. Quartz and crystalline gravel (diameter of 15–40 cm) 2–3 m in thickness are located on the narrow ridge on the right bank of the Eger River. The geological map in Šantrůček et al. (1994) also reported that the small relic of Terrace II north of Hlavnov at 450 m a.s.l. is of Günz age. In the eastern part of the Chlumský práh Horst, these locations of Terrace II are situated at 15–20 m higher altitude (Fig. 10) than that of the (suspected) flight of the equilibrium longitudinal profile of the Eger River terraces.

Terrace III (Chvoječná)

Terrace III has been exceptionally preserved in the area of the Smrčiny Mountains. The small locality at 468 m a.s.l. (relative height of 26 m) containing fragments and sub-angular boulders of quartz and crystalline complex was identified south-east of the Reslava River and the Eger River confluence. The small residual platform near Skalka (461 m) is covered by a 1–2 m thick gravel, which is situated on the top of the phyllite eluvium (Rousek, 1960). In the Chebská pánev Basin, Terrace III has evolved between the northern part of the Hradiště area and the Odrava River valley south-east of Loužek (Fig. 9). To the west, the surface of the terrace is located at 453 m a.s.l., at 450–452 m a.s.l. near Chvoječná and its base is at 438–442 m a.s.l. The maximum thickness of the sand gravel (11–14 m) represents the recessed furrow of the terrace base. The surface of the small and narrow locality with a relic of sand gravel near Loužek is situated at 446–448 m a.s.l.

North of Obilná, the fluvial sediments on the left slope of the most downstream part of the Odrava River valley were excavated in the sand pit, which uncovered 2–3.5 m thick terrace gravel sands and sand gravel. These accumulations cover finer-grained sand gravel, which contain layers of course sand and clay lenses of the Vildštejn Formation. This sand pit revealed 5–7 m thick terrace sediments. The highest part of the currently abandoned sand pit reaches the level of Terrace III (max. 448 m), whereas the predominant parts belong to Terrace IV (max. 440 m). According to Kvaček (1987), the main part of these sediments is of Mindel age, while the lower parts belong to the Riss.

A few small locations of Terrace III in the form of alluvial gravel were recorded on the left bank of the Eger River. These accumulations were presumably deposited by the tributaries of the Eger River. Approximately 3 m thick gravel sand covers a small residual platform at 447.5 m a.s.l. near Tršnice. Šantrůček et al. (1994) sets their origin, as well as that of the close locations of the gravel sands south-west of Doubí (449 m) and between Třebeň and Lesina (450–447 m), to the younger Mindel. The same authors consider the gravel north-west of Chotíkov to be the alluvial accumulations of Günz age.

In the Chlumský práh Horst (Figures 7 and 10), the gravel of the Libocký potok Brook near Horní Pochlovice (max. 453 m) also has an alluvial origin. Clayey quartz, slate, phyllite and gneiss gravel (7–10 cm in diameter) were found in the 6m deep sand pit south of the previously mentioned location. The narrow ridge above the right bank of the Eger River north-east of the Kynšperk nad Ohří is covered by the 3 m thick gravel of Terrace III with its surface at 442 m a.s.l. A small locality of the gravel accumulations at 435–440 m a.s.l. near the Černý Mlýn was also categorised into this level.

Terrace IV (Jindřichov)

The highest upstream appearance of the Eger (Jindřichov) Terrace IV was found south-east of the Reslava River mouth. This denudation relic is situated on the small ridge at 456 m a.s.l. (relative height of 13 m) and it includes dispersed boulders and fragments of the quartz and crystalline complex. A small locality of this terrace level (with the erosionaly lowered surface at 456 m a.s.l.) was also found on the right bank of the Skalka dam 1 km west-north-west of Skalka. In the Chebská pánev Basin, Terrace IV occupies a large area between Jindřichov and the northwestern vicinity of Chvoječná. The surface of this terrace descends from 445–448 m a.s.l. (max. relative height of 40 m) to 440–445 m west of Chvoječná, where its base was located at 430–435 m a.s.l. Moderately coarse sand gravel (5–12 cm in diameter) with layers of pea gravel and coarse sand was uncovered in the 4–5 m deep sand pit south-east of Jindřichov. The upper layers of these fluvial sediments were significantly churned by frost patterns (palsas).

The southern part of the location near Obilná at the lower valley of the Odrava River (surface max. 440 m), which extends to the rock exposure of the mentioned sand pit, was also classified as Terrace IV. On the left bank of the Eger River, the localities of Terrace IV were found north-west of Lesinka (440–442 m), north of Nebanice (442 m) and Chotíkov (434 m) and north-west of Dolní Pochlovice (440 m) above the left bank of the Libocký potok Brook. In the eastern part of the Chlumský práh Horst, a denudation relic (up to 1 m) of small and partly sub-angular gravel near Černý Mlýn with its surface at around 430 m a.s.l. belongs presumably to Terrace IV as well. In the longitudinal profile of the Eger River, Terrace IV does not show any significant anomalies in the area of the Chlumský práh Horst, although its relics form (probably) two altitude levels at up to 431 m and 425 m a.s.l.

Terrace V (Nebanice)

In the area of the Smrčiny Mountains, Terrace V is represented by the gravel accumulation on the left bank of the Eger River valley above the Reslava River mouth. Mlčoch et al. (1993) is of the opinion that this accumulation is of Riss age. On the right bank of the Eger River ca 0.5 km east of the Reslava River mouth, the location of sub-angular boulders and mostly quartz fragments at 451 m a.s.l. (relative height of 6–8 m) were classified as Terrace V. The same applies to the denudation relic of the gravel accumulation on the coastal abrasion cliff of the Skalka dam, which was found ca 1 km west-north-west of Podhoří (the surface is at 446–447 m a.s.l., the base discovered by boreholes is at 442–444 m). On the abraded bank of the Skalka dam, sand gravel (quartz, quartzite, gneiss, basalt) was discovered under the deluvial sediments at ca 3–4 m above the back water level (443 m a.s.l.).

In the Chebská pánev Basin, Terrace V has preserved itself on the large area inside the bend of the Eger River valley, which is located east and north-east of Jindřichov. The surface of this terrace is at 440 m a.s.l. near Jindřichov and at 436–437 m a.s.l. north-west of Chvoječná. The base of this 5–8 m thick sand gravel was discovered by numerous boreholes on the west at 335–330 m a.s.l. Similarly to the surface, the base of this terrace continuously descends to the north and develops into lower Terrace VI. The moderately coarse sand gravel (5–12 cm in diameter, max. 25 cm) mostly of quartz and quartzite was excavated in the former sand pit (4–5 m deep) south-east of Jindřichov. The sand gravel was found in the sand pit at the north-east edge of Jindřichov. These accumulations are significantly frost-churned on the surface and they also contain frost wedges filled with coarser gravel (15–20 cm). In addition to the frequent quartz, weathered phyllites, gneiss and sporadic granites and basalts were also represented. The 80–100 cm thick layer of pellodite, leafy jointed silt and clayey sand was also found in this profile.

The surface of the narrow Terrace V north and north-east of Loužek is situated at 430 m a.s.l., while its decreased north-eastern part reaches the level of Terrace IV. On the left bank of the Eger River, the large gravel-sand plateaux, between the vicinity of Třídvoří and Vrbová (surface at 435–433 m), was also identified as Terrace V. Several currently abandoned sand pits, 2.5–4 m in depth, were established in Terrace V near the north edge of Nebanice (above the railway). These gravel sands were significantly churned by cryogenic processes (Ambrož 1958). According to the borehole data, the thickness of these terrace sediments varies between 1.7–3.8 m. The geological map 11-14 Cheb (Šantrůček et al. 1994) classifies most of the discovered localities of Terrace V to the young Mindel, while some of them to the older Riss.

Terrace VI (Chocovice)

In the Smrčiny Mountains, Terrace VI (the surface at 5–7 m above the Eger River level) was rarely identified. A few accumulations, which do not have the form of a terrace, were located by the boreholes under the deluvial sediments. These deposits are situated, for example, at the river km 258 at 448 m a.s.l. and on the right bank of the Skalka dam ca 1.3 km west of the village Skalka. Sand gravel, which was located by the boreholes in the vicinity of Cheb (for example with the surface at 437.5 m and the base at 432.6 m a.s.l., Šantrůček et al. 1994) could not be verified due to the current built-up area.

In the Chebská pánev Basin, Terrace VI was preserved at the northern edge of the large fluvial terrace accumulation near Chocovice, thus inside the extensive bow of the Eger River valley. Its surface is located at 428–431 m and its base at 426 m a.s.l. An even larger locality of the Chocovice terrace was found north of Vokov with its surface at 423–427 m a.s.l. North of Nebanice, Terrace VI is represented by the southern part of the gravel sand accumulation located above the railway. Its surface is situated at 425–426 m a.s.l. The surface of the rather extensive Terrace VI locality near Chotíkov (below the railway, 3 m above the flood plain) is at 420–423 m a.s.l. The sand pit (only 3 m deep) near the eastern edge of the village, was founded in the medium coarse to coarse sand gravel (5–10 m in diameter), which is built mostly by the quartz and quartzite.

In the Chlumský práh Horst, the denudation relics of the gravel are located at the level of Terrace VI. The surface is at 415–420 m a.s.l. (base at 409–411 m a.s.l.) south-west and west of Chlumek. These accumulations are, according to Šantrůček et al. (1994), of young Mindel age. Scattered gravel at the higher level (up to ca 430 m a.s.l.) corresponds to the already lowered Terrace VI. Note that the geological map by Šantrůček et al. (1994) marks this gravel locality over an unreasonably large area and, moreover, at the altitude of 450–433 m. Thus, it should also contain the levels of Terrace IV and V.

Terrace VII (Chotíkov)

The occurrences of Terrace VII, with its surface not more than 3.5–4 m above the river level, are closely connected to the sediments of the valley floor, with which this terrace usually shares its base. In numerous places, the surface of Terrace VII continuously evolves into the surface of the flood plain. The longitudinal profile of the Eger River valley demonstrates (Fig. 7) that Terrace VII follows the river (and flood plain) only in the Chebská pánev Basin, i.e. Terrace VII appears below the slope change near Cheb. Its surface is situated at 431–433 m, while its base is at 425–427 m a.s.l. Particularly, this terrace was located near Chocovice with its surface at 428 m and its base at 423 m a.s.l., near Vlkov and Třídvoří (424 m) and near Nebanice and Chrtíkov with its surface at 418–419 m and its base at 414 m a.s.l. In the Chlumský práh Horst, the extensive locality of Terrace VII (surface at 415–417 m) is situated at the entry of the Eger River to the horst near Dolní Pochlovice. The lowered parts of Terrace VII surface are locally covered by the relics of the Holocene flood clay.

Recent flood plain and valley floor

The sedimentary filling of the current valley floor of the Eger River is formed by two layers: the lower mostly sandy gravel and gravel sands of Upper Pleistocene age and the overburden mostly sandy loam and loamy sands of Holocene age. The course of the valley floor bedrock is often characterised by a significantly uneven slope ratio. The surface of the flood plain demonstrates a change of its gradient below the Špitálský vrch Hill (between river km 242–241) in the north-western part of Cheb, i.e. approximately at the entrance of the river to the basin. This gradient step of the river, which is being shifted in the upstream direction by back erosion, is lithologically conditioned by the more resistant quartzite mica shists that form the valley slopes up to the north-eastern vicinity of Cheb (river km 238–239). Two fault systems with a NW–SE direction, which cut the Eger River valley at the river km 241 and 239, were also important for the formation of this gradient step of the river.

The bedrock of the valley floor sediments does not have (according to boreholes) an equilibrium longitudinal profile. Locally, significant differences in sediment thickness were discovered. Similarly to the surface of the flood plain, the gradient step of the valley floor bedrock was found below the dam of the Skalka Reservoir, i.e. at the entry to the Chebská pánev Basin from the crystalline complex of the Smrčiny Mountains. Another slight increase of the valley floor sediment thickness (about 2 m) appears between the river km 233.5–232, i.e. on the lithological boundary of the Vildštejn Formation (Upper Pliocene – Lower Pleistocene) and the Miocene Sokolov Formation.

A significant change of the valley base gradient was found in the foreground of the Mariánské Lázně Fault (Vrba 1959), where the thickness of the valley floor sediments rises up to about 5 m above the adjacent sections. The sand gravel of this location covers the sediments of the Vildštejn Formation, which fill the tectonic depression in the Miocene accumulations of the Cypris Formation. The lowest point of the Vildštejn Formation base is situated at 393 m a.s.l., i.e. 26 m under the flood plain surface. Right before the Mariánské Lázně fault, the base of the valley floor sediments creates a distinct elevation, which causes the reduced thickness of the sediments (4 m), i.e. it is situated ca 3 m under the river level. Further below the Mariánské Lázně fault, the course of the Eger River valley floor returns to the level, which correspond to its gradient above the mentioned anomaly. The smaller thickness of the valley floor sediments (between river km 220–211) was found on the most elevated part of the Chlumský práh Horst.

In the foreground of the Mariánské Lázně fault, the tectonic trench, whose floor lies 26 m under the flood plain surface, creates a depression in the overburden sediments, which reaches up to 8 m (Vrba 1959). Usually, the thickness of these overburden Holocene sandy loam and clay sands vary between 1–2 m. The underlying sediments are 3–5 m thick (Vrba, 1959) and they contain granularly heterogeneous sands with quartz boulders (3–8 cm in diameter) and of the basal coarse gravel (5–12 cm in diameter, up to 25 cm). In between the two strands of the Mariánské Lázně fault (river km 219.5–217), both Miocene and Vildštejn sediments were uplifted up to 450 m a.s.l. (relative height of 35 m). Considering that Terrace III (the surface at 442 m a.s.l. in the area of the Chlumský práh Horst) does not show marks of the tectonic disruption, it can be stated that the uplift took place in the Lower Pleistocene.

The longitudinal profile of the Eger valley floor (Fig. 7) is substantially connected to the lithological, tectonic and geomorphological conditions of the studied area. The more distinct anomaly of the gradient occurs in the incised part of the Eger valley in the Hazlovská vrchovina Hills (Smrčiny Mountains). Here, between river km 266–260, the average gradient of the valley floor reaches 3.33‰. In the area of the Františkolázeňská kotlina Basin, between river km 258.6–252.6, the average gradient of the valley floor reaches 1.00‰. In the Chebská pahorkatina Hills, between the vicinity of Bříza and the north-western edge of Cheb (river km ~252.6–240), the average gradient of the Eger valley floor reaches only 0.833‰. It is therefore smaller than that of the Františkolázeňská kotlina Basin. This low gradient value corresponds to the conditions of the significantly meandering channel and the wide flood plain, which evolved here before the construction of the Skalka Reservoir. Between the Reslava River mouth and Skalka, the mean gradient of the floodplain axis is 1.571‰, while the channel gradient reaches 0.846‰. A markedly increased gradient of 5.00‰ occurs in the ca 1km long section of the valley under the dam of the reservoir and in several weir locations. In the Chebská pánev Basin (river km 240–217), the gradient of the meandering river level is 19 m to 23 km, which corresponds to the mean gradient of 0.826‰. However, the average gradient of the valley (flood plain) axis is 1.58‰. The significant change of the river level gradient (1.418‰) occurs in the area of the Chlumský práh Horst (river km 216.8–209.2). Subsequently, this gradient decreases in the western part of the Sokolovská pánev Basin (river km 209.2–206.6) to 1.05‰.

Morphostructural characteristics of the valley and the analysis of the gradient conditions of the Eger River level implies, that the increased river level gradient in the Hazlovská vrchovina Hills was caused by the smaller discharge of the river and also by the more resistant rocks of the Smrčiny Mountains massif, which are located in the tectonically uplifted area in the proximity of the Chebská pánev Basin. The balanced gradient conditions of the Františkolázeňská kotlina Basin and of the subsequent section of the Chebská pahorkatina Hills correspond to the presence of the less erosion resistant rocks. The lithologically conditioned section of the increased river level gradient (ca 5‰ to 1 km) occurs before the entrance to the Chebská pánev Basin above Cheb. The balanced gradient of the Eger river level in the Chebská pánev Basin, based on the resistant and homogenous basin sediments of the Cypris and Vildštejn Formation, is not substantially influenced by the river meandering. The increase of the river level gradient in the Chlumský práh Horst corresponds to the Quaternary uplift of the crystalline block between the Chebská pánev Basin and Sokolovská pánev Basin at the Mariánské Lázně fault. The average gradient of the Eger River meandering channel in the Chlumský práh Horst (1.418‰) is similar to the gradient of its flood plain axis in (1.58‰) and above (1.571‰) the Chebská pánev Basin.

5. Discussion

The correlation of the Eger River terraces between the Smrčiny Mountains and the Sokolovská pánev Basin and the terrace system of the middle course of the Eger River valley is based on the research made in the Doupovské hory Mountains (Balatka 1993) and in the Mostecká pánev Basin (Balatka and Sládek 1976; Tyráček 1995). The evaluation of these findings, together with that of the terrace system of the lower Eger River course (for example Balatka and Sládek 1976; Tyráček 1995, 2001; Tyráček et al., 2004), the Elbe between Děčín and Hřensko (Balatka and Kalvoda 1995; Kalvoda and Balatka 1995) and the central part of the Bohemian Massif (Balatka and Kalvoda 2008, 2010; Balatka et al. 2010a,b, 2015) are used to suggest the classification of the Eger River terraces in the studied area into the stratigraphical system of the Quaternary (Tab. 2).

Because of the significant river channel meandering, only 6 Eger River accumulation terraces have evolved in the Chebská pánev Basin (Fig. 7, Balatka and Kalvoda 2018). The relative height of their surfaces reaches up to 25–30 m, while their thickness is mostly 4–6 m, rarely 8–10 m. The surface of the highest preserved level (Terrace II) is embedded 10–25 m into the platform surface of the Vildštejn Formation, which originated in the Lower Pleistocene. Presumably, the stratigraphical classification of the Vildštejn Formation corresponds to that of Terrace I in the lower course or the Eger River valley. In comparison with the terrace system of the Sokolovská pánev Basin, the equivalent terraces of the Chebská pánev Basin are situated at a significantly lower level, namely there is a 25 m difference in the case of Terrace II and a 8 m difference in that of Terrace V. This was caused by the neotectonic uplift of the asymmetric Chlumský práh Horst (Figures 7 and 10).

In the Sokolovská pánev Basin, through which the Eger River flows in the length of 55 km including the 10 km long fault gap in the granodiorite of the Slavkovský les Mountains, the 10 levels of the river accumulation terraces were discontinuously preserved. The highest terrace levels (up to relative height of 101 m) have marks of tectonic deformation. The fold deformations of the Tertiary sediments and of the overburden terrace sand gravel were found in the exposures of the brown coal mines. Their formation was most likely caused by the Quaternary tectonic movements (e.g. L. Kopecký 1978, 1985). Due to the greater uplift of the northern part of the Sokolovská pánev Basin along the Krušné hory fault, which includes the lacustrine sediments covering the edge of the granites of the Slavkovský les Mountains, the Eger River channel took a position along the southern edge of the basin in the Neogene. Only small relics of the terraces have been preserved in the deep epigenetic and fault gap valley of the Eger River at 50–58 m and 40–44 m above the river (Čtyroký 1996). In its middle course, the Eger River flows through the Mostecká panev Basin in a 39 km long section, which is characterised by the deeply incised valley (over 400 m) cutting the Doupovské hory Mountains. The river level gradient has a higher value in this area (2.03‰ on average, maximum 6.9‰). Locally, the river also erodes the crystalline bedrock of the neovolcanites.

The Doupovské hory Mountains represent a tectonically active volcanic block of the Eger rift, which has been gradually uplifted since the Lower Miocene up to the present (L. Kopecký 1985). The rate of this uplift is estimated at 200–300 m. In the Doupovské hory Mountains, the Eger River valley was created during the Neogene in the zone of the tectonic bend between the lava flows, which descends from the centre of the volcanic massif in a northward direction and the volcanic bodies, which face away from the fault slope of the Krušné hory Mountains in a southward direction (Balatka 1993). In this part of the Eger River valley, the incomplete river terrace system of the 9 levels shows a significant convergence in the direction against the river flow (ca 30 m at the highest levels). Therefore, the river flows approximately at the level of Terrace VI at its entry to the Doupovské hory Mountains below Karlovy Vary. The surface relics of the oldest Quaternary Terrace I are situated at 71–54 m above the river. However, its lowest situated surface (near the mouth of the Bystřice River) was found at 106 m above the river.

In the western part of the Mostecká pánev Basin, the course of the oldest terraces shows a tectonic uplift of ca 15–20 m in the longitudinal profile. The highest channel gradient of the Czech section of the Eger river valley is connected to the uplifted block (horst) of the crystalline complex along the Střezov fault. This gradient reaches up to 10‰ (Balatka 1993). The distinctive fold deformations with a brachyanticline shape, which affected the Miocene clays and coarse sandy gravel, were found in the sediments of Terrace III and IV near the Střezov fault. These deformation structures were created mainly by frost and by the extrusion of the underlying clays into the terrace sediments during a period of periglacial climate in the Pleistocene, although the neotectonic movements of the basin bedrock could also play their role (L. Kopecký 1978, 1985).

The stratigraphical system of the Eger River terraces in the Mostecká pánev Basin was created by Tyráček (1995, 2001). Its correlation with the Elbe terrace system was made using the terrace system of the lower course of the Bílina River (the older levels, Balatka 1995) and that of the lower course of the Eger River (the younger levels, Balatka and Sládek 1976). The richly segmented structure of the terrace system in the Mostecká pánev Basin has 24 levels. Such a system has not been found in any other Czech river valley. Low-resistant incoherent lacustrine Tertiary sediments of the Severočeská pánev Basin (sands, silts and clays) together with the tectonic uplift of this area enabled the long-term gradual evolution of the Eger River valley. All of the Eger River terrace levels were preserved in the Mostecká pánev Basin due to the movements and meandering of its channel. These terraces are characterised by a relatively low thickness (3–7 m, locally around 10 m) and by a small vertical difference. In the main accumulation area of the Mostecká pánev Basin, 24 river terraces were identified (Balatka and Sládek 1976; Tyráček 1995; Tyráček et al. 2004). These terraces were stratigraphically classified into 7 groups (I–VII), which correspond to the main river terraces of the Elbe. In the Mostecká pánev Basin, the surface of the oldest terrace level (I_1) is situated at 125 m above the Eger River level. According to the paleomagnetic research in the Mostecká pánev Basin, the oldest Terrace I group and presumably even Terrace II_1 and II_2 are older than 1.64 Ma (Tyráček et al. 2004).

In the Mostecká pánev Basin, the river terraces appear in two separate areas: the older levels (group I–IV) have evolved on the extensive platforms outside the valley cut, while the younger levels (group V–VII), mostly smaller-scale meander terraces, are located inside the valley cut. In the downriver direction, the terrace levels show a distinct divergence (10–40 m) in the longitudinal profile. This divergence is more prominent in the case of the older terraces. In the upstream direction, both of the younger terrace groups (VI, VII) were unified with the current valley floor, while some of the accumulations of the older levels have been joined together as well. Therefore, the number of terrace levels decreases to 17 in the western part of the Mostecká pánev Basin. In the central part of the basin, the terrace groups I–V are directed north-east to the current Bílina River valley and the older terraces have distinctively shorter course than the younger levels (Balatka and Sládek, 1976). The last terrace of this paleogeographical period (V_1) has the longest course, which follows the eastern margin of the Mostecká pánev Basin and enters the České středohoří Mountains. These relics of the fluvial sediments (the surface at the relative height of 37 m) of Middle Pleistocene age have been preserved in the abandoned river valley connecting (in the hanging position) the current Eger River valley with that of the Bílina River.

The valley meanders of the various evolutional stages are typical for the Eger River valley in the Mostecká pánev Basin, such as the entrenched or abandoned meanders, which have been formed since the Middle Pleistocene to the Holocene. In the western part of the Mostecká pánev Basin, the meander evolution is estimated to have begun at the same period as the formation of Terrace V_1. This period is characterised by intense river meandering, which was caused by the backward erosion progressing through the neovolcanites in the České středohoří Mountains. The impulse to the incision of the originally free meanders and to the evolution of the entrenched meanders was given by the relocation of the Eger River channel from the eastern margin of the Mostecká pánev Basin to its current position in the subsidence area of the Eger fault zone, containing the less resistant sediments of Upper Cretaceous age, which took place after the sedimentation of the sandy gravel of Terrace V_1. In the eastern part of the Mostecká pánev Basin, the erosion processes were slowed-down during the Saal due to the relocation of the Eger River into the Dolnooharská pánev Basin. Therefore, the surface of the coal beds was not substantially denuded.

All of the older Eger River terraces, including the level $V_{1,}$ are directed across the Mostecká pánev Basin to the north and northeast towards the Bílina River valley, which is also occupied by these terraces in its 45 km long section heading towards the Elbe valley in Ústí nad Labem. In the area of the former Eger River and the Bílina River confluence, 15 terraces have evolved. Their relative height reaches the level of the oldest Terrace I_1 at ca 95 m, while the terrace groups VI and VII are entirely missing (Balatka 1995). All 7 terrace groups were identified in the lower Bílina River valley, i.e. 15 levels. The surface of the highest Terrace I_1 is situated at 118 m. Only 12 levels of the corresponding Elbe terraces have evolved in the vicinity of Ústí nad Labem. The surface of the oldest Terrace I_1 is situated at 138 m above the river (Tab. 1).

The morphostructural analysis of the tectonically disrupted terrain of the western headland of the Cretaceous Dolnooharská tabule Plateau, thus in the fault zone along the south-eastern margin of the Mostecká pánev Basin, proved the existence of the relatively young (presumably Middle Pleistocene) tectonic movements. This subsidence of the narrow rift depression floor contributed to the relocation of the Eger River to the east. Due to this relocation, which took place during the Middle Pleistocene, only terrace groups VI and VII evolved in the lower Eger River valley in the Cretaceous Dolnooharská tabule Plateau. Their surfaces are situated at 25 m above the river level. Balatka and Sládek (1976) proved the existence of the tectonically conditioned depression, which affected the sediment basis of the Eger River valley floor as well as that of the lowest terrace in the wider area of the Eger River and Elbe confluence. The thickness of its fluvial deposits reaches up to 20 m. The base of the Eger River valley floor sediments is situated in a hanging position in relation to the previously mentioned depression. There, the difference in altitude reaches more than 10 m.

6. Conclusions

Geomorphological research of the Eger valley between the Smrčiny Mountains and the western part of the Sokolovská pánev Basin was aimed at the investigation and morphogenetic evaluation of localities with river terraces and further fluvial sediments. The whole Eger valley developed during the Neogene in morphotectonic depressions of extensive fault zones. The Chebská pánev Basin originated at the intersection of the Eger rift and the Cheb-Domažlice fault zone and its river network is incised ca 40 m into planation surfaces of the sedimentary basin. Sedimentary material in the Chebská pánev Basin reached a maximum thickness up to 400 m and these accumulations, including the Vildštejn layers of Pliocene-Pleistocene age, are disturbed by numerous faults. Both volcanic processes and frequent seismic activity in the region are associated with Late Cenozoic tectonic movements.

In the area between the Smrčiny Mountains and the Sokolovská pánev Basin the terrace system of the Eger River with following levels was identified (Table 1): the Pliocene terrace niveau B, the Cheb Terrace (I), the Hradiště Terrace (II), the Chvoječná Terrace (III), Jindřichov Terrace (IV), Nebanice Terrace (V), Chocovice Terrace (VI), Chotíkov Terrace (VII) and the recent flood plain (N). Geomorphological analysis and reconstruction of terraces flights of the Eger River enabled their inclusion in the current stratigraphical scheme of the Quaternary. It was determined to be a morphostratigraphical system of 7 river terraces of Quaternary age (Table 2). Older levels of fluvial sediments, occupying a still higher morphological position in the area between the Smrčiny Mountains and the Sokolovská pánev Basin, have been classified to the Pliocene.

A comparison of terrace flights in the longitudinal profile of the Eger River (Fig. 7) allowed the range of the Quaternary tectonic processes in the region to be specified. The Quaternary tectonics in the Smrčiny Mountains may be proven by the uplifted relics of the first terrace (Fig. 8), namely about 10 m in comparison with its level in the Chebská pánev Basin (Fig. 9). In the Chlumský práh Horst area, the oldest Pleistocene terraces (I and II), which originated during the Tiglian stage, were uplifted approximately about 15 m (Figures 7 and 10). The Terrace III, which is younger than 1.78 Ma, is uplifted about 2–4 m in this crystalline horst. In the western foreland of the Mariánské Lázně fault, an at least 8 m deep tectonic depression of the valley bottom sediments was found. This depression is also incised into the Vildštejn Formation.

The ascertained morphostratigraphy of the river accumulation terraces of Quaternary age between the Smrčiny Mountains and the western part of the Sokolovská pánev Basin is also compared to the current knowledge about the evolution of the valley and river terraces at the middle and lower stream of the Eger (Tables 1 and 2), namely in the Doupovské hory Mountains and the Mostecká pánev Basin. According to the current stratigraphical scheme of the Quaternary, the Eger River terrace system was formed mostly by the Pleistocene during the Tiglian to the Weichselian stages. The proposed morphostratigraphy of

Eger terraces River River (Table 2) can be used as a basis for their systematic radiometric dating. This procedure then allows to complete the chronostratigraphy of the formation of these fluvial landforms in the Quaternary.

Acknowledgements

The paper was completed in the framework of Charles University project PROGRES Q44 "Geography".

References

Ambrož, V. (1960): Zpráva o výzkumu a mapování čtvrtohorních pokryvných útvarů Sokolovské pánve. Antropozoikum 10, NČSAV, Praha, 221–228.

Ambrož, V., Šantrůček, P., Mrázek, P. (1958): Závěrečná zpráva základního výzkumu Chebské pánve za léta 1956–1957. Ústřední ústav geologický, Praha, MS Geofond, P 10418, Praha.

Babuška, V., Plomerová, J., Vecsey, L. (2010): Links between the structure of the mantle lithosphere and morphology of the Cheb Basin (Eger Rift, Central Europe). Geologische Rundschau 99, 1535–1544.

Balatka, B. (1993): K vývoji údolí Ohře v Doupovských horách. Sborník České geografické společnosti 98(2), 107–122.

Balatka, B. (1995): Terasy střední a dolní Bíliny – spojovací článek terasových systémů Ohře a Labe. Sborník České geografické společnosti 100 (4), 249–267.

Balatka, B., Kalvoda, J. (1995): Vývoj údolí Labe v Děčínské vrchovině. Sborník České geografické společnosti 100(3), 173–192.

Balatka, B., Kalvoda, J. (2006): Geomorfologické členění reliéfu Čech. Kartografie, a.s., Praha.

Balatka, B., Kalvoda, J. (2008): Evolution of Quaternary river terraces related to the uplift of the central part of the Bohemian Massif. Geografie 113(3), 205–222.

Balatka, B., Kalvoda, J. (2010): Vývoj údolí Sázavy v mladším kenozoiku. Evolution of the Sázava valley in the late Cenozoic. Edice Geografie, Česká geografická společnost, Praha.

Balatka, B., Kalvoda, J. (2018): Terasový systém řeky Ohře mezi Smrčinami a západní částí Sokolovské pánve. Informace České geografické společnosti 37(2), 1–15.

Balatka, B., Sládek, J. (1962): Říční terasy v českých zemích. Geofond v NČSAV, Praha.

Balatka, B., Sládek, J. (1976): Terasový systém střední a dolní Ohře. AUC Geographica 11(2), 3–26.

Balatka, B., Gibbard, P. L., Kalvoda, J. (2010a): Evolution of the Sázava Valley in the Bohemian Massif. Geomorphologia Slovaca et Bohemica 10(1), 55–76.

Balatka, B., Gibbard, P. L., Kalvoda, J. (2010b): Morphostratigraphy of the SázavaRiver terraces in the Bohemian Massif. AUC Geographica 45 (1–2), 3–34.

Balatka, B., Kalvoda, J., Gibbard, P. (2015): Morphostratigraphical correlation of river terraces in the central part of the Bohemian Massif with the European stratigraphical classification of the Quaternary. AUC Geographica 50(1), 63–73, https://doi.org/10.14712/23361980.2015.87.

Bůžek, Č., Kvaček, Z., Holý, F. (1985): Late Pliocene palaeoenviroment and correlation of the Vildštejn floristic complex within Central Europe. Rozpravy ČSAV, ř. MPV, 95, 7, Praha, 72 p.

Čtyroký, J. (1996): Vývoj údolí Ohře ve Slavkovském lese. Diplomová práce, PřF UK, Praha.

Danzer, E. (1922): Morphologische Studien im mittleren Egergebiete zwischen dem Karlsbad-Falkenauer und dem Komotau-Teplitzer Tertiärbecken. Arbeiten des geographisches Instituts der deutschen Universität in Prag. Neue Folge, 3, Prag, 13–48.

Demek, J., Mackovčin, P., eds. et al. (2006): Zeměpisný lexikon ČR: Hory a nížiny. Vydání II. Agentura ochrany přírody a krajiny ČR, Praha.

Dobeš, M., Hercog, F., Mazáč, O. (1986): Die geophysikalische Untersuchung der hydrogeologischen Strukturen im Cheb-Becken. Sborník geologických věd, Užitá geofyzika, 21, ÚÚG Praha, 117–158.

Engelmann, R. (1922): Die Entstehung des Egertales. Abhandlungen der Geographischen Gesellschaft in Wien, 12, Praha.

Gibbard, P. L., Cohen, K. M. (2008): Global chronostratigraphical correlation table for the last 2.7 million years. Episodes 31, 243–247, https://doi.org/10.18814/epiiugs/2008/v31i2/011.

Gibbard, P. L., Head, M. J., Walker, M. J. C. and the Subcomission on Quaternary Stratigraphy (2010): Formal ratification of the Quaternary System/Period and the Pleistocene Series / Epoch with a base at 2.58 Ma. Journal of Quaternary Science 25, 96–102, https://doi.org/10.1002/jqs.1338.

Gottsmann, J. (1999): Tephra Characteristics and Eruption Mechanics of the Komorní Hůrka Hill Scoria Cone, Cheb Basin, Czech Republic. Geolines, 9, Praha, 35–40.

Chlupáč, I. et al. (2002): Geologická minulost České republiky. Academia, Praha.

Kalvoda, J., Balatka, B. (1995): Chronodynamics of the Labe River Antecendence in the Děčínská vrchovina Highland. Acta Montana, Ser. A, 8 (97), 43–60.

Kolářová, M. (1965): Hydrogeologie chebské pánve. Sborník geologických věd, řada HIG, 3, ÚÚG v NČSAV, 7–101.

Kopecký, A. (1960): Předběžná zpráva o základním geologickém výzkumu a mapování čtvrtohorních pokryvných útvarů Sokolovské pánve za léta 1958–1959. Antropozoikum 10, NČSAV, Praha, 229–236.

Kopecký, A. (1966): Zpráva o mapování kvartéru na území listu Cheb v roce 1965 (M-33-61-D). Zprávy o geologických výzkumech v roce 1965, ÚÚG v Academii, Praha.

Kopecký, L. (1978): Neoidic taphrogenic evolution and young alkaline volcanism of the Bohemian Massif. Sborník geologických věd, Geologie 31, 91–108.

Kopecký, L. (1985): Neovulkanity (paleogén/neogén). In: M. Malkovský et al.: Geologie severočeské hnědouhelné pánve a jejího okolí. Oblastní regionální geologie ČR, ÚÚG v Academii, Praha, 138–147.

Krejčí, J. (1939): Profil rovnováhy jakožto základ studia říčních teras. Spisy Odboru české společnosti zeměpisné v Brně, řada A, 5, Brno.

Kvaček, J. (1987): Terasy řeky Ohře mezi Kynšperkem nad Ohří a Tršnicemi u Chebu. Diplomová práce. PřF UK, Praha.

Kvaček, J. (1989): Terasy Ohře v chebské pánvi mezi Tršnicemi a Kynšperkem nad Ohří. Zprávy o geologických výzkumech v roce 1986, ÚÚG v Academii, Praha.
Maheľ, M., Kodym, O., Malkovský, M. (1984): Tektonická mapa ČSSR 1 : 500 000. Vysvětlivky k Tektonickej mape ČSSR 1 : 500 00.
Malkovský, M. (1976): Saxonische Tektonik der Böhmischen Masse. Paleogeography of the Miocene. Geologische Rundschau 65, 127–143, https://doi.org/10.1007/BF01808459.
Malkovský, M. (1979): Tektogeneze platformního pokryvu Českého masívu. Knihovna ÚÚG, 53, ÚÚG v Academii, Praha.
Malkovský, M. (1980): Model of the origin of the Tertiary basin at the foot of the Krušné hory Mts.: volcanotectonic subsidence. Věstník Ústředního ústavu geologického 55(3), 141–150.
Mazáč, O., Pokorný, O. (1961): Geofyzikální výzkum chebské pánve. Sborník geologických věd, Užitá geofyzika 1, 81–126.
Mlčoch, B., Hradecký, P., Kovanda, J. (1993): Geologická mapa ČR, 11-13 Hazlov, 1 : 50 000. Soubor geologických a účelových map, Český geologický ústav, Praha.
Müller, V. ed. et al. (1998): Vysvětlivky k souboru geologických a účelových map v měřítku 1 : 50 000. Listy 11-13 Hazlov, 11-14 Cheb. Edice ekologických map České republiky. Český geologický ústav, Praha.
Peter, J. (1923): Geologische-morphologische Studien über das Falkenauer Tertiär-Becken. Lotos 71, 379–420.
Peterek, A., Reuther, C. D., Schunk, R. (2011): Neotectonic evolution of the Cheb Basin (Northwestern Bohemia, Czech Republic) and its implications for the late Pliocene to Recent crustal deformation in the western part of the Eger Rift. Zeitschrift für Geologie und Wissenschaft 39(5/6), 335–365, 24 Abb.
Rousek, V. (1960): Zpráva o geologickém mapování zátopné oblasti přehradního místa u Skalky na Ohři. Ústřední ústav geologický, Praha, MS Geofond P 11391.
Shrbený, O. (1982): Geochemistry of the West Bohemian neovolcanics. Časopis pro mineralogii a geologii 24(1), 9–21.
Šantrůček, P., Kopecký, A., Kopecký, L., Škvor, V. (1969): Vysvětlující text ke geologickým mapám 1 : 25 000, listy M-33-61-D-a Františkovy Lázně, M-33-61-D-b Tršnice, M-33-61-D-c Cheb jz. část, M-33-61-D-d Cheb. Komplexní úkol T-1-20: Regionální geologický výzkum ČSSR. Dílčí úkol ÚÚG 4/04: Základní geologický výzkum a mapování severočeských pánví. Ústřední ústav geologický Praha, MS Geofond P 21770.
Šantrůček, P., Králík, F., Kvičinský, Z. (1994): Geologická mapa ČR 1 : 50 000. List 11-14 Cheb. Soubor geologických a účelových map, Český geologický ústav, Praha.
Šibrava, V., Havlíček, P. (1980): Radiometric age of Plio-Pleistocene volcanics rocks of the Bohemian Massif. Věstník Ústředního ústavu geologického 55(2), 129–139.
Škvor, B. V., Satran, V. eds. et al. (1974): Krušné hory – západní část. Soubor oblastních geologických map s textovými vysvětlivkami. 6 listů map v měřítku 1 : 50 000. ÚÚG, Praha.
Špičáková, L., Uličný, D., Koudelková, G. (2000): Tectonosedimentary Evolution of the Cheb basin (NW Bohemia, Czech Republic) between Labe Oligocene and Pliocene: a preliminary note. Studia geophysica et geodetica 44, 556–580, https://doi.org/10.1023/A:1021819802569.
Tyráček, J. (1995): Stratigraphy of the Ohře River terraces in the Most Basin. Sborník geologických věd, Antropozoikum 22, 141–157.
Tyráček, J. (2001): Upper Cenozoic fluvial history in the Bohemian Massif. Quaternary International 79, 37–53, https://doi.org/10.1016/S1040-6182(00)00121-X.
Tyráček, J., Havlíček, P. (2009): The fluvial record in the Czech Republic. A review in the context of IGCP 518. Global and Planetary Change 68(4), 311–325 https://doi.org/10.1016/j.gloplacha.2009.03.007.
Tyráček, J., Minaříková, D., Kočí, A. (1985): Stáří vysočanské terasy Ohře. Věstník Ústředního ústavu geologického 60(1), 77–86.
Tyráček, J., Westaway, R., Bridgland, D. (2004): River terraces of the Vltava and Labe (Elbe) system, Czech Republic, and their implications for the uplift history of the Bohemian Massif. Proceedings of the Geologists' Association 115, 101–124, https://doi.org/10.1016/S0016-7878(04)80022-1.
Václ, J. (1979): Geologická stavba chebské pánve a jejího okolí. Geologický průzkum 21, 133–135, https://doi.org/10.1080/00076797900000016.
Vodohospodářská kancelář ministerstva techniky v Praze (1950): Podélný profil řeky Ohře od ústí až po státní hranici. Praha.
Vrba, J. (1959): Hydrogeologické zhodnocení území mezi Ávou, Nebanicemi, Mostovem a Kynšperkem na pravém břehu Ohře. Závěrečná zpráva. Vodní zdroje, Praha, MS Geofond P 11203 Praha.
Vrba, J. (1981): Nové poznatky o režimu mělkých podzemních vod v oblasti soutoku Ohře a Odravy. Věstník Ústředního ústavu geologického 36, 471–474.
Wagner, G. A., Gogen, K., Jonckheere, R., Kampf, H., Wagner, J., Woda, C. (1998): The age of Quaternary volcanoes Železná hůrka and Komorní hůrka (Western Eger Rift), Czech Republic, alpha-recoil track, TL, ESR and fission track chronometry. Magmatism and Rift Basin Evolution, Excusion Guide, Abstracts. IGCP No 369 Workshop, Liblice, 95–96.
Wilschowitz, H. (1917): Zur Morphologie des Kaiserwald – Egertales, ein Beitrag zur Heimatkunde des Elbogener Kreises. Lotos 65, 89–102.
Záruba-Pfeffermann, Q. (1943): Podélný profil vltavskými terasami mezi Kamýkem a Veltrusy. Rozpravy II. tř. České akademie 52(9), Praha.
Záruba, Q., Bucha, V., Ložek, V. (1977): Significance of the Vltava Terraces System for Quaternary Chronostratigraphy. Rozpravy ČSAV, MPV, 87, 4, Academia, Praha.