sciendo

Slovak Journal of Civil Engineering

Vol. 30, 2022, No. 4, 1 – 10
DOI: 10.2478/sjce-2022-0022

THE POTENTIAL AND IMPLICATIONS OF AUTOMATED PRE-PROCESSING OF LIDAR-BASED DIGITAL ELEVATION MODELS FOR LARGE-SCALE ARCHAEOLOGICAL LANDSCAPE ANALYSIS

David, NOVÁK[1], Filip, PRUŽINEC[2*], Tibor, LIESKOVSKÝ[2]

Abstract

LiDAR-derived digital elevation models (DEMs) have transformed the archaeological study of landscape features, broadened our technical capabilities, and enhanced the accuracy with which terrain relief is described. These models also place demands on how researchers and analysts interpret DEM content in the context of the modern landscape. LiDAR-based DEMs contain modern man-made structures that can significantly influence model properties. Although data are usually filtered and some of these artificial features are removed during bare-earth classification, many terrain interventions remain visible. This large-scale case study applies established methods to a freely available DEM of the Czech Republic in an attempt to evaluate differences between original and filtered DEMs. It applies a fully automated filtering procedure using vector topographic maps to avoid manual corrections that would make the procedure problematic when used on a macro scale. The results of our archaeological GIS analysis demonstrate that this procedure, despite its relative simplicity, can achieve a significantly better representation of a landscape compared to that offered by an unfiltered DEM. Finally, we propose a series of future steps with a view to developing a more comprehensive and accurate model and overcoming its limitations.

Address

[1] Institute of Archaeology of the Czech Academy of Sciences, Prague, Czech Republic

[2] Dept. of Theoretical Geodesy and Geoinformatics, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Slovakia

* ***Corresponding author:*** filip.pruzinec@stuba.sk

Key words

- Landscape archaeology,
- DEM filtering,
- GIS analysis,
- Visibility,
- Land surface curvature,
- Drainage networks
- Czech Republic.

1 INTRODUCTION

During recent decades, light detection and ranging (LiDAR) data has become an integral element of archaeological landscape research (Crutchley, Crow, 2009; Gojda, John, 2013; Posluschny, 2015; Chase et al., 2017; Garstki, 2020). As well as serving the purposes of surveying and documentation, LiDAR data offers significant promise for contemporary landscape analysis and the modeling of terrain attributes in relation to the past. The use of LiDAR-derived digital elevation models (DEM) for the analysis of large landscapes on a national scale has long been hampered by the inaccessibility of data. The situation has changed in recent years (Hobič, 2020). In the Czech Republic, it is now possible to freely obtain a relatively precise LiDAR-derived DEM raster grid through the portal of the Czech State Administration of Land Surveying and Cadastre (http://geoportal.cuzk.cz/). Researchers can now access a simplified Digital Terrain Model of the Czech Republic of the 5th Generation (DMR 5G), which significantly broadens the possibilities for landscape interpretation by facilitating automated algorithmic methods.

Taking advantage of this favorable situation, the Institute of Archaeology of the Czech Academy of Sciences in Prague recently partnered with the Slovak University of Technology to create a set of models for description and analysis of the landscape of the territory of Bohemia. These include a total visibility model (Llobera, 2003; Gillings, 2017; Brughmans et al., 2018), a stream network model (Kuna, Danielisová, 2009; Hengl, Reuter, 2009; Tarboton et al., 2009) and a terrain classification model (Hengl, Reuter, 2009; De Reu et al. 2013; Novák, 2018). In most cases, preparing these types of models on a large scale is time-consuming. The total visibility model in particular presents considerable computational demands (Rášová, 2017; Kuna et al. 2022).

© 2022 The Author(s). This is an open access article licensed under the Creative Commons Attribution 4 .0 License (https://creativecommons.org/licenses/by/4.0/)

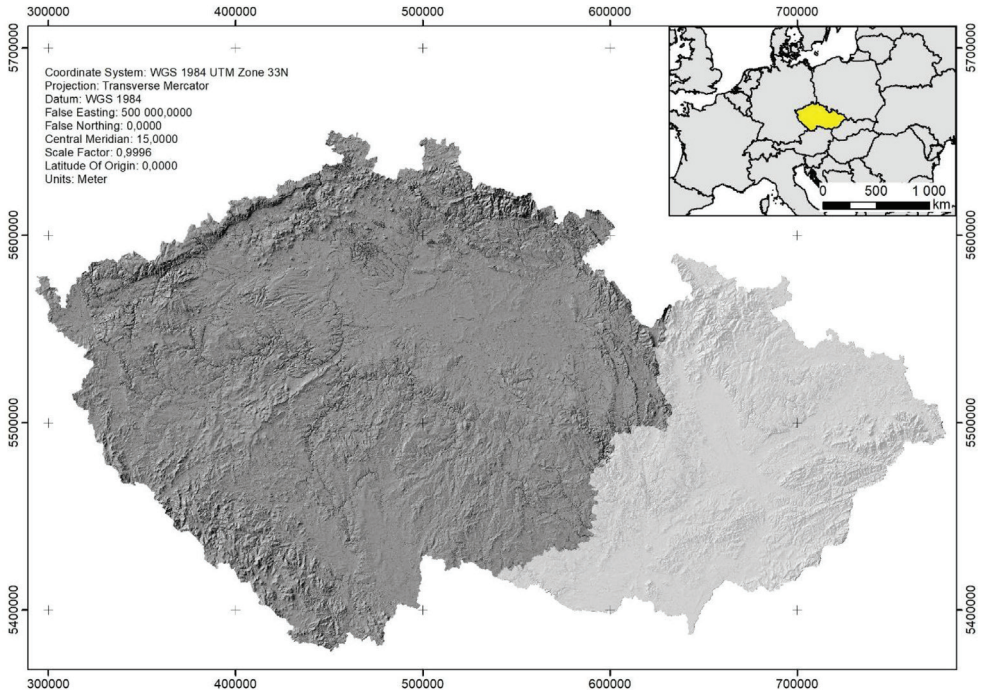


Fig. 1 *Hillshade DMR 5G of the region analyzed (data provided by ČÚZK and EuroGeographics).*

Therefore, the availability of centrally available reference models is of paramount importance for researchers tasked with analyzing datasets to ensure that results can be compared across different projects. The need for this kind of resource corresponds with the mission of the Archaeological Information System of the Czech Republic (https://www.aiscr.cz/), a central infrastructure that supports research in this domain.

The use of LiDAR-derived DEMs, however, also gives rise to a number of complications. An obvious issue that can introduce distortions is the presence of modern artificial features, which can be difficult to distinguish from the natural landscape relief. Bodies of infrastructure constructions, terrains destroyed by mining, built areas, and water management structures are some of the features that negatively impact results. One solution to this problem is to significantly reduce DEM resolution and thus suppress minor anthropogenic features. However, this approach is not recommended due to its tendency to distort the overall digital representation and deform natural landforms. It is also inadequate for handling large terrain interventions. The ideal method is to identify undesirable features and then to remove and replace them by interpolation or using legacy data (e.g. historical maps; see Zwertvaegher et al., 2010; 2013; Werbrouck et al., 2011; Schmidt et al., 2018). This paper presents a procedure used to create a filtered DEM of Bohemia (supplemented by the Vysočina region; Fig. 1) suitable for further spatial analysis. It also compares the extent and nature of the changes to archaeologically relevant properties brought about by the resulting DEM compared to the original unfiltered version. The total area under examination is relatively large at 57,019 km^2, which represents approximately two-thirds of the Czech Republic.

2 THEORY: WHAT TO FILTER OUT AND WHY?

DEMs derived from laser scanning represent a given landscape at the time of data collection and contain features produced by human activity. This is entirely consistent with the idea that – especially in the context of Central Europe – landscapes are a cultural product consisting of a set of artifacts and ecofacts (Kuna 2004; David, Thomas, 2008). Although it is difficult to find a location in the Czech Republic that has not undergone human intervention, from the point of view of analyzing digital landscape models, it is especially important to consider those interventions that have substantially altered the visibility, drainage conditions, and morphology of a landscape. Such interventions can typically be dated to periods close to the present, enabling them to be better distinguished from the natural terrain. Ideally, these features should be identified and removed to bring the resulting DEM surface representation more in line with pre-industrial topography and to reduce the risk of bias in future analyses (Werbrouck et al., 2011).

In addition to human intervention, landscapes have been permanently influenced by natural erosion and accumulation processes, leading to characteristic changes. However, capturing these processes is not the aim of this paper for several reasons. As demonstrated by previous studies (De Smedt et al., 2013; Habicht et al., 2017), the patterns and extent of erosion are strongly dependent on local conditions and should therefore be identified by targeted field surveying such as core sampling or geophysical prospecting. Moreover, erosion and accumulation are dynamic processes, the effects of which vary over time. Therefore, any such landscape reconstruction must take into account the specific time period on which it is modeled. It is also worth noting that in the case of Czech Republic, the effects of erosion processes on the overall landscape are particularly evident in river floodplains with meandering watercourses. Yet, even in these areas, recent interventions fundamentally changing the characteristics of the otherwise relatively well-preserved Holocene relief (see section 3.1) have had a strong impact.

The aim of this study was to identify a means of removing the maximum number of recent intrusions with the highest possible efficiency and the least possible manual interventions. Although previous studies have applied similar approaches, they have been typically limited to much smaller regions with uniform conditions involving the manual processing of input data or other interventions (Werbrouck et al., 2011; Zwertvaegher et al., 2010; 2013; Schmidt et al., 2018). This study evaluates the extent to which commonly available topographic data are adept at correcting DEMs for archaeological purposes. We outline a model that retains most of the qualitative parameters associated with accurate LiDAR images of natural landforms, which are typically represented with significant inaccuracies in DEMs derived from other sources (cf. Novák, 2018).

3 METHODS

Depending on the availability of data and the type of DEM analyzed, approaches adopted by various authors concerning the removal of more recent DEM features vary (Werbrouck et al., 2011; Höfler et al., 2015; Schmidt et al., 2018). The three most common steps involve the use of vector layers as a feature identifier, buffer masking of the affected surface, and interpolation. Other supplementary procedures used to retrieve missing data or to more clearly identify affected surfaces generally require manual work or terrain analysis, measures that can only be accomplished within a reasonable timeframe when analyzing small areas. Given that our area of interest was on a national scale, we selected a straightforward buffering approach based on inverse distance weighting (IDW) interpolation provided by the GDAL *FillNodata* tool. QGIS software (https://qgis.org/) was used for most of the data analysis, with data processing scripted in Python using the GDAL (https://gdal.org/) application programming interface (API).

3.1 Source data

The DEM (DMR 5G) source is a digital representation of a bare earth surface with a vertical accuracy of 0.18 m in an open landscape and 0.3 m in forested areas as well as local deviation extremes of up to 1 m. The model was created from data obtained by airborne laser scanning of the Czech Republic between the years 2010 and 2013, with an average density of points of approx. 1 p/m^2. The DMR 5G model itself is a result of automated filtering, manual corrections, TIN triangulation, and subsequent raw data smoothing. Therefore, the distribution of data points tends to be uneven, and they are only derivatives of the originally measured values (Brázdil et al. 2016). In its basic form, the DMR 5G is distributed as point-based data (LAS) and subject to significant financial charges. As a result, it is practically impossible for this data to be deployed in country-wide analysis. The public version of DMR 5G used in this paper has been downscaled and is distributed as a regular raster grid with a spatial resolution of 2 x 2 m. Considering the complex process of creating the DMR 5G and the possible vertical deviations involved, simplifications present in the public raster grid compared to the full DMR 5G were deemed insignificant, with negligible effects on properties such as visibility, drainage and local curvature.

To determine the modern features of the model, we selected a set of layers (Tab. 1) from the planimetry section of the Fundamental Base of Geographic Data of the Czech Republic (ZABAGED), available online (http://geoportal.cuzk.cz/). Maintained as a seamless database, ZABAGED contains two-dimensional spatial and descriptive information on settlements, communications, utility networks, pipelines, hydrology, administrative units, protected areas, vegetation, surfaces, and terrain relief within the territory of the Czech Republic. The version we used was acquired in 2014 and closely matches the terrain captured by DMR 5G laser scanning.

3.2 Preparation of the clipping mask

We first identified ZABAGED layers that would be most suitable for use as an extraction mask for modern terrain features. Once selected, these ZABAGED vector layers were then grouped according to buffer sizes large enough to cover the affected surface. The layers were divided into 10 groups with buffer sizes of 2, 4, 6, 12, 15, 16, 20, 25, 30 and 50 meters (Tab. 1). To determine the size of the buffer, we analyzed terrain profiles on the vector layers computed from the DEM across different landscape types (Fig. 2). The surfaces affected were generally defined based on an optical assessment of the DEM hillshade visualization. Areas of the natural terrain displaying the highest level of deformation were selected fo a profile analysis. For each ZABAGED layer type, dozens of profiles were monitored in order to determine maximum ranges within the altered terrain. Buffer sizes were then assigned according to the maximum range identified, with an addition of 5–10% based on anticipated deviations for the individual layer type. Due to the unique nature of the possible alterations across different landscape types and the need to estimate the original relief pattern, we primarily relied on qualified estimates based on the above procedure when assigning size groups. Finally, a single polygon mask layer was created as an input to compute the new DEM. The resulting buffer layer contained approx. 3.8 million polygons. Clipping buffers are available through the Zenodo repository (https://doi.org/10.5281/zenodo.6368001).

In the first iteration of the filtering of the model, all water streams and water bodies were included in the layers used for clipping. This approach proved problematic. Firstly, unlike infrastructure and other built features that are subject to targeted planning, the delineation of water streams in ZABAGED compared to DMR 5G is imprecise, mainly due to the dynamically changing stream channels in the natural landscape (Fig. 3). Secondly, although many watercourses pass

Tab. 1 *List of ZABAGED layers used for filtering, descriptions, geometric representations and clipping buffer sizes.*

Layer title	Description	Geometry	Buffer (m)
BudovaBlokBudov	blocks of buildings	polygon	4
Cesta	dirt roads	polyline	12
HaldaOdval	spoil heaps	polygon	2
HradbaVal	ramparts/defensive walls	polyline	12
ChladiciVez	cooling towers	polygon	25
Kolejiste	railway stations	polygon	4
KulnaSklenikFoliovnik	sheds/greenhouses	polygon	4
ParkovisteOdpocivadlo	parking areas	polygon	2
Pesina	footpaths	polyline	6
PovrchovaTezbaLom	open pit mines	polygon	4
PrehradniHrazJez	dams/weirs	polyline	6
RozvalinaZricenina	ruins	polygon	4
SilniceDalnice [D, R]	highways	polyline	50
SilniceDalnice [S1]	roads (first class)	polyline	30
SilniceDalnice [S2]	roads (second class)	polyline	20
SilniceDalnice [S3]	roads (third class)	polyline	15
SilniceNeevidovana	unlisted roads	polyline	12
SilniceVeVystavbe	roads under construction	polyline	20
Silo	silos	polygon	4
Skladka	dumps	polygon	25
TramvajovaDraha	tramways	polyline	12
Ulice	streets	polyline	12
UsazovaciNadrzOdkaliste	sedimentation tanks	polygon	4
ValcovaNadrzZasobnik	cylindrical tanks	polygon	2
VodniPlocha [see 3]	water bodies	polygon	20
Zed	walls	polyline	6
ZeleznicniTrat	railways	polyline	16
ZeleznicniVlecka	railway sidings	polyline	16

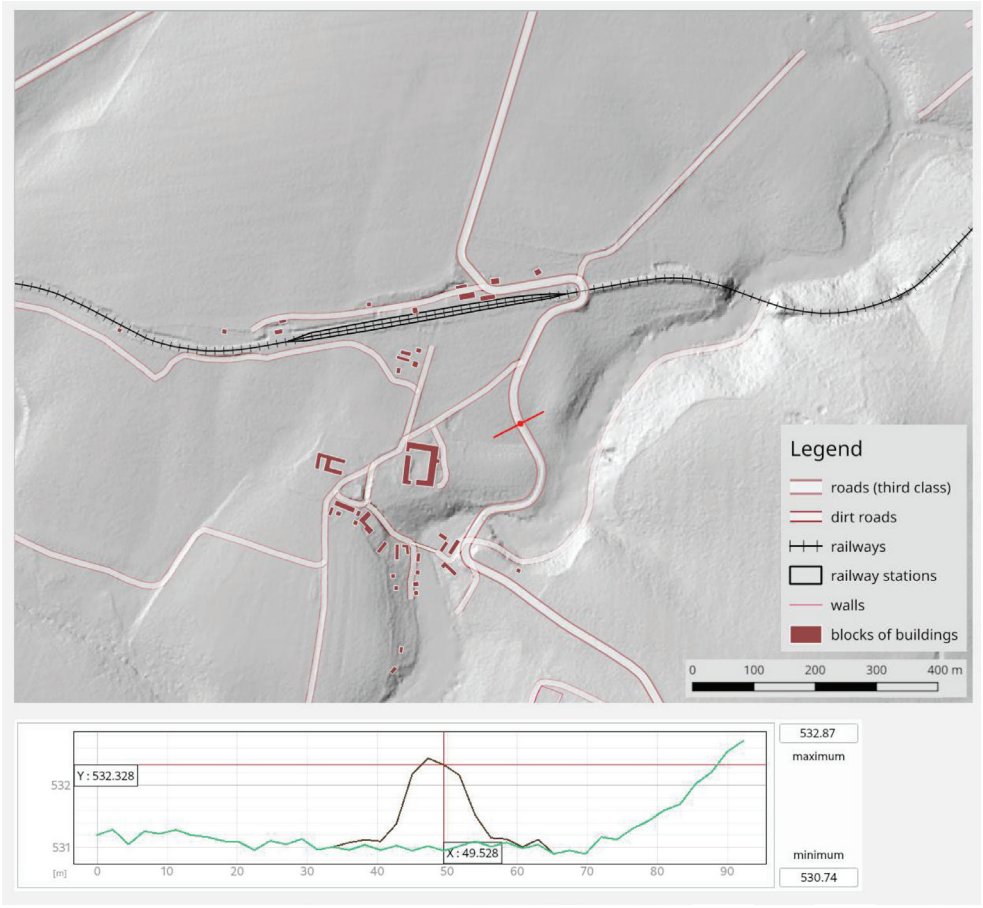


Fig. 2 *Example of a terrain profile (marked in red on map) computed from the input DEM (brown line on graph) and the filtered DEM (green line on graph). The terrain profile of the output DEM displays a continuous surface that most likely reflects the pre-construction terrain (data provided by ČÚZK).*

through artificial channels, a significant proportion of them flow through an open landscape and have thus avoided direct human modification. When these watercourses were used for clipping and subsequent interpolation, the ratio between the correctly removed artificial features and the newly introduced distortions proved unsatisfactory. Especially in the case of complex terrains with narrow, deeply incised valleys, the resulting surfaces were significantly reduced from the original well-defined landforms. Thirdly, in the case of large water bodies, interpolation proved unreliable when filling in missing data. Given these inaccuracies, we found it preferable to preserve the original flat-water surface for further modeling instead of creating an interpolated surface based on an insufficient number of known data points. Thus, water surfaces were only included for still waters with a total perimeter of up to 3000 m. In practical terms, this resulted in the exclusion of large reservoirs, where back-replacement produced the greatest deviations from the expected real terrain, especially reservoirs situated in deep river valleys with steep slopes such as the Vltava Cascade; fishponds and similar smaller features were masked out. Given that there are only six naturally occurring lakes in the country (out of 65,027 filtered water bodies), there was no need to find a solution for the undifferentiated artificial and natural water bodies in ZABAGED.

3.3 Interpolation process

To remove features from the DEM, we used a buffer layer to insert *nodata* values into a GeoTIFF raster (Fig. 4). New values were interpolated using the GDAL FillNodata tool, which uses IDW interpolation to compute unknown values from adjacent pixels within a given search distance specified in cells. For each cell, a four-direction conic search was performed to determine the values for interpolation. Once all the values were interpolated, zero or more smoothing iterations (3x3 average filters on the interpolated cells) were applied to smooth out the artifacts.

Small testing areas of the DEM, which were selected to represent most landscape types, were used to determine a suitable search distance and number of smoothing iterations. We chose a 200-meter (100-cell) search distance and 20 smoothing iterations. However, the buffer layer contained polygons that can generate *nodata* areas in the DEM beyond the scope of the 100-cell search distance specified by the *FillNodata* algorithm. Therefore, these polygons were identified and removed from the processing (Fig. 4). Furthermore, since interpolating values over such large areas can result in inaccuracies, the original values from the output DEM were retained. To identify these features, we converted polygons into closed lines and computed a buffer comprising a search distance radius for every line. Polygons (buffers) containing holes were selected and removed from the buffer layer. For this reason, 4.17% of the total buffer area originally included in the filtering went unresolved. As it was not possible to assess the significance of these individual exclusions, the exact impact of this step on the results could not be quantified. The affected areas, which consist mostly of large-scale mining zones, could be reliably reconstructed using legacy elevation data without the need for interpolation. In regions impacted by large-scale mining (e.g., the northwest Bohemia region), this shortcoming of the model is undoubtedly locally significant; on the other hand, it did not detract from the overall trajectory of the model or the degree of positive changes achieved (see section 4).

The new DEM was processed using a Python script that slices the

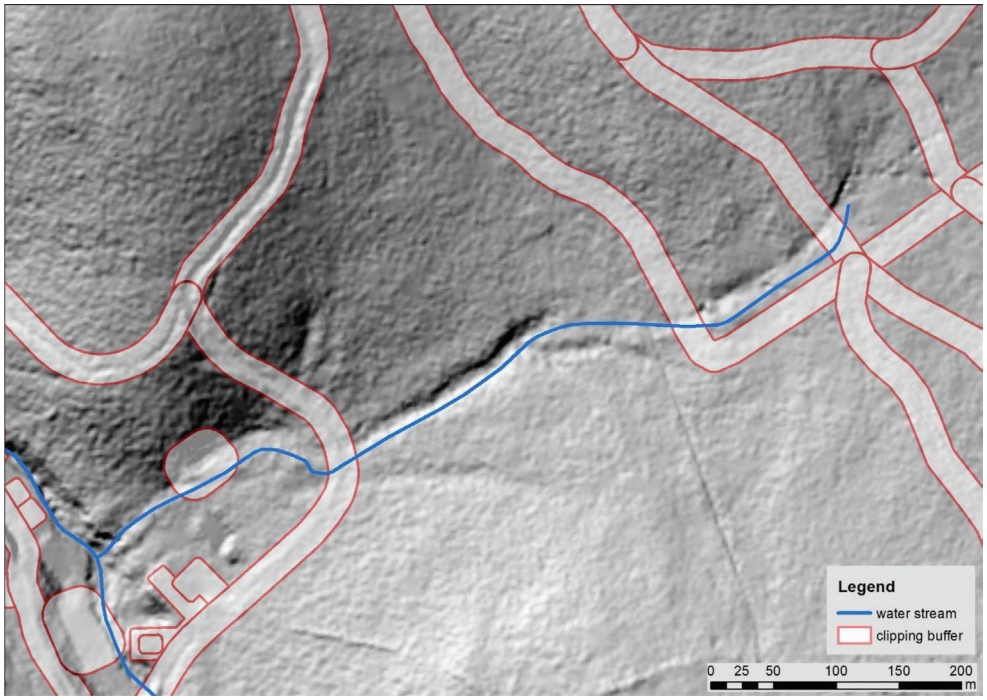


Fig. 3 *The difference in the real position of a watercourse (hillshade DMR 5G) and a stream delineated according to the ZABAGED vector map (blue line). The deviation visible in this section of the stream often exceeds 15 m. Buffers used for clipping are also shown (data provided by ČÚZK).*

DEM into 20 x 16 km raster tiles featuring 200-meter overlays. This ensures seamless merging of the tiles interpolated within a 100-cell search filter. Tiles with a specified extent were created using GDAL Python API's *Translate* function. The same extent was applied as a spatial filter to the mask layer. Where the tiles contained elevation data and polygons remained in the mask layer after applying the filter, the GDAL *RasterizeLayer* function was used to insert *nodata* values to the tile layer using polygons from the mask layer. Finally, the *FillNodata* function was used to interpolate missing values, with the tile cropped by the overlay distance using the *Translate* function. The other option involved merging tiles in this step, thereby saving some processing time; however, replacing an incorrectly processed tile is simpler than processing the whole DEM again in cases where the buffer filtering step did not finish properly. QGIS was used to build a virtual raster, which was then converted into a GeoTIFF. Since the *FillNodata* tool also interpolates values beyond the original border, the merged raster had to be cropped using the original raster value boundary (Fig. 4). To suppress the edge effect, additional data from MERIT DEM (http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/) was used to interpolate the area close to the Czech border.

The tiling step can be replaced with GDAL command-line tools used directly on the input DEM. However, this tiling approach takes 4-6 hours to process (without tile merging) on an average computer (Intel® Core™ i5-4690 CPU @ 3.50GHz × 4, 16GB RAM). Directly inserting *nodata* values to the whole DEM takes about 25 hours and interpolating new values takes an additional 8 hours. The duration of the tile merging varies, depending on the tools used and the read/write speed of the hard drive used.

3.4 Validation methodology

Some parts of the modern features fell outside the interpolated areas of the model, but it was unfortunately not possible to assess with any statistical precision the proportion of artificial elements removed or to reliably distinguish how many of them continued to produce distortions. Therefore, our evaluation was focused more on comparing the relevant properties of the input and output DEMs, while assessing changes and their possible influence on future archaeological analyses.

The main objective was to create a DEM with better analytical potential in terms of visibility, drainage networks, and terrain morphology. To reduce computational complexity with a view to testing these model properties, we used downscaled versions of the DEMs at a cell size of 8 x 8 m. The following models were created for both DEMs (original and interpolated): (1) cumulative visibility model; (2) stream network model; (3) land surface curvature (LSC) model. All the models were then used to assess the qualities of the filtered DEM compared to the input DEM. Validation-related data are available through the Zenodo repository (https://doi.org/10.5281/zenodo.6368001).

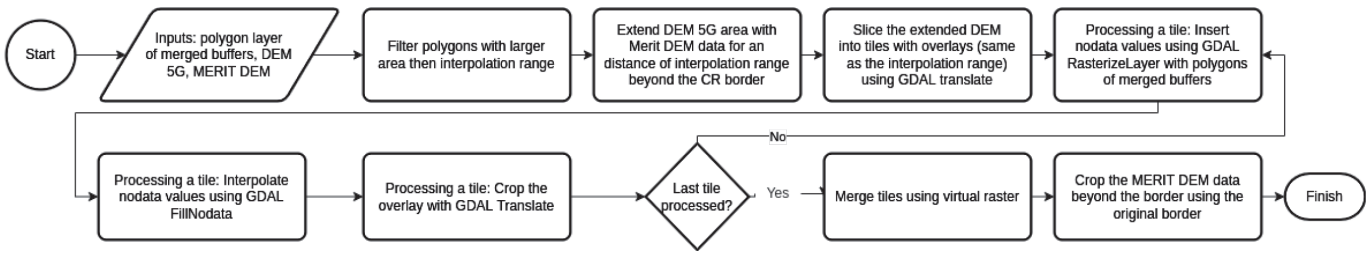


Fig. 4 *The interpolation process used to remove the more recent DEM features*

3.4.1 Testing the visibility properties

Visibility is a parameter with wide-ranging implications for the perception of historic human activity (Gibson 1986; Llobera 2007). Regardless of the specific application or theory, the principle of visibility analysis is to identify an uninterrupted visual link between two points in a terrain or DEM. However, the modern landscape is filled with objects that significantly alter visual properties and observability. When transcribed to LiDAR-derived DEMs, these features create potentially significant biases, which can compromise the analysis and study of pre-industrial landscapes. To evaluate the effect of these features on the visual properties of the original and filtered DEMs, two cumulative visibility models were calculated for regularly spaced points on a 10 x 10 km grid with a visibility radius of 5 km and an observer height of 2 m. In total, 574 viewsheds (without overlaps) were combined within a single model.

3.4.2 Testing drainage properties

As with visibility, runoff conditions are strongly influenced by the specific shape of the micro-relief. In particular, linear infrastructure features have a significant effect on the reconstruction of drainages, as they often act as channels that alter the natural gradient. To test the effect of filtration on the drainage modeling, a stream network for the original and filtered DEMs was calculated using standard ESRI ArcGIS tools and the TauDEM toolbox. Each input model was extended to the entire catchment using EU-DEM (https://www.eea.europa .eu/data-and-maps/data/copernicus-land-monitoring-service-eu-dem) as a data source for areas outside the Czech Republic (in particular, Lower Austria). Drainage-free zones were identified and filled in automatically. Drainages with a Strahler order greater than 4 were defined as potential watercourses (despite the simplifications, which had no bearing on our analysis).

3.4.3 Testing curvature properties

The degree of curvature is considered the basis for morphometric descriptions and analyses of the relief (Hengl and Reuter, 2009; Krebs et al. 2015; Minár et al., 2020). Landforms identified by curvature, especially terrain edges and ridges, are often the sites of building activity carried out by pre-modern societies, e.g., burial mounds (Kuna 2006) or feudal seats (Novák 2019). LiDAR-derived DEMs allow researchers to identify such shapes and to examine relationships between settlements and landforms in much greater detail. Accordingly, definitions of such landforms must be based on a model that is at least tentatively stripped of elements of the industrial landscape. These features are often characterized by linear or stepped cuts in the terrain, which tend to appear as unnatural structures, and which in turn affect the analytical properties of the DEM.

To evaluate the qualitative differences between the original and interpolated DEM, profile and plan LSC models were calculated using the *Curvature* function in ESRI ArcGIS. From these models, cells with a negative value exceeding the standard deviation were extracted in the case of profile curvatures (terrain edge peaks), while cells with a positive value exceeding the standard deviation were extracted in the case of plan curvatures (ridge peaks). In principle, the average of the values was close to zero for both models, as each concave feature has its own convex counterpart. The results were combined as a single model, indicating the location of significant relief features and then compared for both DEMs.

4 RESULTS

The DEM output was first validated with the same profile method used to determine the buffer sizes (Fig. 3). The total interpolated area of the DEM was 15% of the original model, i.e., 8,729 km^2. Changes in the DEM elevation within the interpolated areas were typically ± 5 m, with a minimum value of 64.5 m and a maximum value of 101.4 m (Fig. 5).

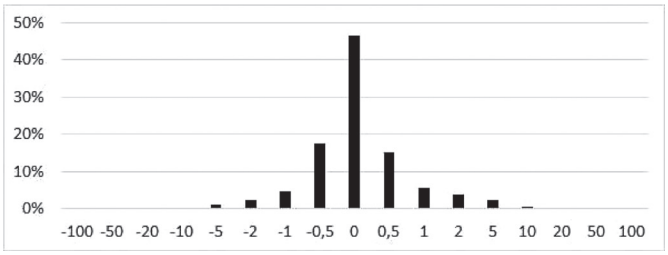


Fig. 5 *The difference in the distribution of the altitude values for the original and interpolated DEMs. The negative values indicate an increase in the terrain, while the positive values indicate a decrease. The measurements are in meters.*

4.1 Changes in the visibility properties

Based on the computations performed, it can be stated that even though the observer in the random sample was often placed in a highly unsuitable position and the overall visibility of the surroundings was therefore minimal (on average 4% of the defined radius), there was an improvement in overall visibility for the filtered DEM in 94% of the cases observed, with an average improvement of 11% (confidence interval +-5% for $p < 0.01$; standard deviation 51%; median 5%; Fig. 6). Statistically significant differences in observations at the individual sample points were validated by the paired two-sample t-test, which confirmed changes in the values measured at a significance level of $p < 0.01$ ($p = 7.36e\text{-}34$; $t = -12.96$; $t_{crit} = 2.58$).[1] The overall impact of the filtration on the visual properties of the resulting model was so great that even after allowing for the fact that the samples were independent and validated by the two-sample t-test, which assumes unequal variances (despite these values representing the same individuals), the difference between the two samples reached a significance level of $p < 0.1$ ($p = 0.099989$; $t = -1.64624$; $t_{crit} = 1.64619$).[2]

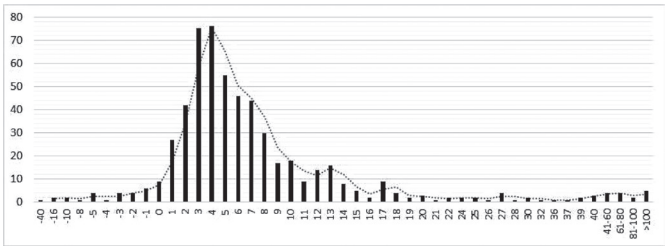


Fig. 6 *The difference in the distribution of the visibility values in the original and interpolated DEMs (based on a random sample of 574 viewsheds). There is a strong prevalence for positive values (improvement in visibility).*

These findings undoubtedly demonstrate that the use of the unfiltered DEM featuring modern formations had a significant impact on the results. In order to create the total visibility models mentioned in the introduction, our adjustment was a crucial step without which the models would have generated significantly different values compared to the previous state. Conditions around the early medieval hillfort

[1] Null hypothesis: *the number of cells visible from the selected point in the original model did not significantly differ from the number of cells visible from the same point in the filtered model.*

[2] Null hypothesis: *the set of visibility values derived from the original model was significantly different compared to the set of values acquired for the filtered DEM.*

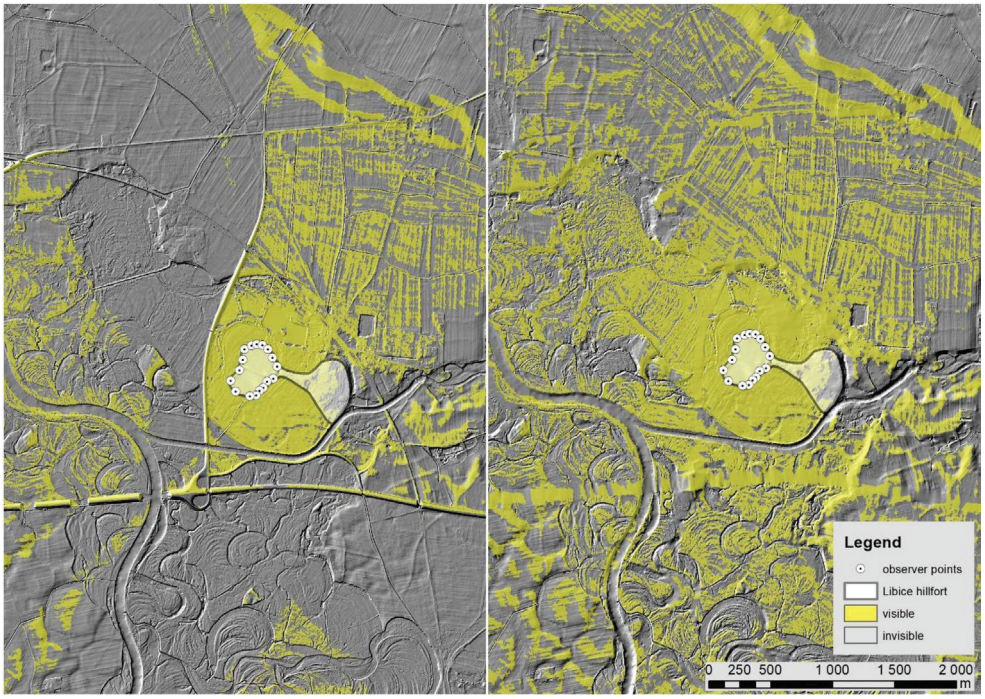


Fig. 7 *Viewshed comparison of the surroundings of the Libice nad Cidlinou hillfort before (left) and after DEM filtering (right). The visible sites are shaded in yellow (observer height 2 m; observation points located on the edge of the acropolis). In the flat terrain, the removal of road bodies had a significant impact, e.g., changing the visibility conditions considerably (data provided by ČÚZK).*

in Libice nad Cidlinou are illustrative in this respect, where a fundamental difference in the visibility of the surroundings from the hillfort acropolis before and after the filtering is clearly visible (Fig. 7).

4.2 Changes in the drainage properties

The results show that of the total 22,194,177 cells identified as part of the original drainage network, only half (52%) retained their location using the filtered DEM, while the total percentage of the interpolated terrain did not exceed 15% of the model area. Using a paired two-sample t-test on a sample area of one randomly selected sheet of the 1:50,000 Base Map (No. 22-11; 459 km^2; 205,911 drainage cells in the original model), at a significance level of $p < 0.01$ ($p = 0$; $t = 462$; $t_{crit} = 2.6$), we can completely reject the null hypothesis that the original drainage network has not changed.[3] These figures suggest large differences between the drainage patterns in the pre-industrial and modern landscapes, which are reflected by the substan-

[3] Null hypothesis: *If the cell in the original model is part of the drainage, it is also present in the filtered model and in the same Strahler order*.

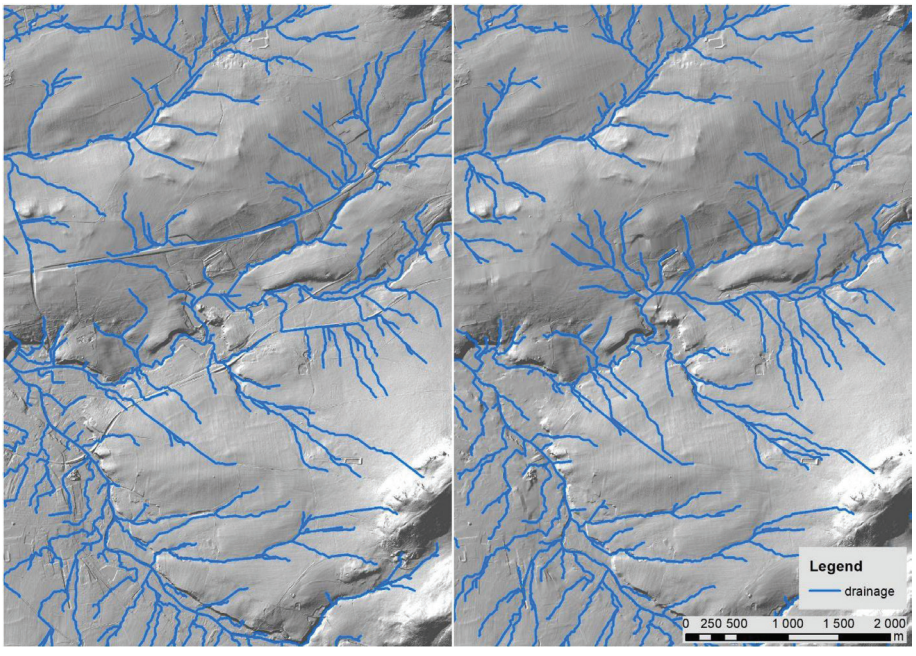


Fig. 8 *Comparison of the drainage network before (left) and after the DEM filtering (right). The drainage conditions more closely approximate the characteristics of the natural landscape, both in the open terrain and in the built-up areas (data provided by ČÚZK).*

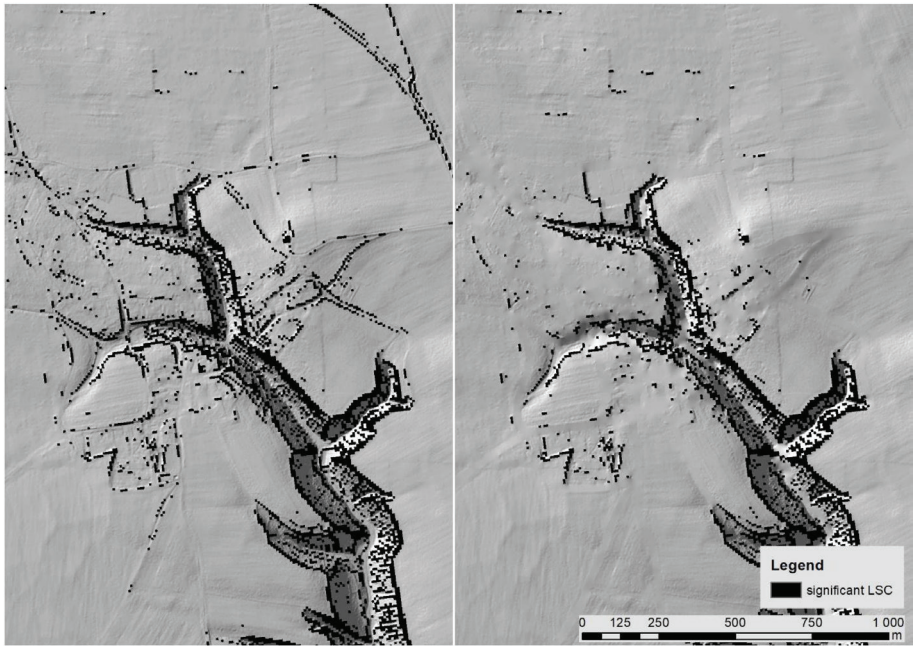


Fig. 9 *A comparison of LSC before (left) and after DEM filtering (right). Cells with a significant profile or plan curvature are highlighted in black. Only some of the more prominent artificial edges remain visible after cleaning, but most of the curvature follows the natural course of the terrain, and the noise fraction has been reduced (data provided by ČÚZK).*

tially different characteristics of the two models. An optical comparison of the models below shows that the filtering rectified the major shortcomings of the original DEM (Fig. 8).

4.3 Changes in the curvature properties

Based on our evaluation, the interpolated DEM contained 9% fewer significant edges and ridges than the original unrefined model. Using the paired two-sample t-test validation for the sample area once again (see 4.2; 330,590 cells with significant LSC in the original model), the difference between the models was also confirmed at a significance level of $p < 0.01$ ($p = 0$; $t = 422$; $t_{crit} = 1.96$).[4] Moreover, if we take into account that the clipping buffers contained 27% of the significant LSCs but occupied only 15% of the total DEM area, it is clear that a considerable part of the significant LSCs was due to modern man-made features captured by the laser scanner. It was therefore desirable that in the interpolated areas, the reduction of significant LSCs would reach 75% of the original occurrences. An optical comparison of the two results indicates that the interpolation removed substantial sections of the unnatural breaks, while preserving the basic characteristics and form of the original terrain (Fig. 9). Although many man-made features of the model were not removed, this type of interpolation substantially increases the opportunities for geomorphometric analysis of the DEM.

5 DISCUSSION

Our results convincingly demonstrate that DEM filtering, at least to the extent presented here, is a valuable technique for DMR 5G (or any similar LiDAR-derived DEM) and general archaeological geomorphometric applications. For the purposes of the surveying and documentation of archaeological sites, the application of the clipping mask can assist in the processing of data for automated classifications and also in removing disturbances before manual (visual) identification of burial sites (cf. Werbrouck et al., 2011). Based on a visual assessment of the outputs, the smooth interpolated segments of the DEM at first seem unnatural. To remedy this effect, a certain amount of noise can be applied to the interpolated values to make the shaded DEM appear more natural. On balance, however, retaining the smooth surface is recommended for future processing steps such as viewshed or stream network analysis.

As expected, and as can be seen from the resulting model, some of the modern elements in the interpolated DEM remain. In principle, these deficiencies are not caused by the chosen method, but by the quality of the input data (clipping mask). With a more complete record of modern interventions, such as marginal roads, sections of channelised streams, and other similar features, the credibility of the output model and the impact of the filtration would improve further. Interpolation is also problematic when used in highly urbanized environments, where achieving a plausible reconstruction of the original surface requires considerably more complex approaches (cf. Boháčová and Herichová, 2004; Bini et al. 2016; Herichová, 2019), which are extremely difficult to implement across such a general level.

The most extensive interventions, such as open-pit mines and water reservoirs, posed a persistent problem. In these instances, a reconstruction of the relief would likely require the use of historical mapping data containing elevation information. Therefore, in regions affected by large-scale mining, our approach would benefit from additional measures. There were local zones in the model where reconstruction was essentially impossible, resulting in the preservation of the original data from the output DEM. Some residues also remained from larger infrastructure constructions where the ground disturbance exceeded the normal extent evaluated in the initial analysis, as reflected by the size of the clipping buffers. These shortcomings may be overcome by optimizing the generation of clipping zones, so that the complexity (especially the local curvature) of the surrounding terrain is properly defined, e.g., by using the principle of mean curvature watersheds (Romstad and Etzelmüller, 2012).

Based on the properties of the input and output DEMs, our procedure, despite its limitations, significantly improves the applicability of the model for archaeological landscape research. Although it is not possible to precisely determine how closely the modifications reflect

[4] Null hypothesis: *if the cell in the original model has a significant LSC, the same cell will also have a significant LSC in the filtered model*.

the original terrain shape, and, despite the filtered DEM containing potential distortions (some caused by the filtering and interpolation itself), our results overwhelmingly point to a significant enhancement of the original model, thus increasing the potential for further analysis. At the same time, it should be noted that the objective of the modifications was not to completely reconstruct the pre-industrial landscape *per se*, but rather to eliminate the major deficiencies of DMR 5G when performing further calculations (see section 1). Creating a completely accurate reconstruction would require a much more complex approach, i.e., one that considers the many processes that led to these artificial changes in the landscape. That would also require the addition of historical data for large areas of the landscape that have been mined or otherwise deformed, which, in principle, cannot be achieved by interpolation alone.

6 CONCLUSIONS

LiDAR-derived DEMs have undoubtedly advanced technical capabilities in the field of archaeology, but their integration within landscape-archaeological analysis remains limited. One of the reasons for this may be an issue discussed in this paper, i.e., the presence of modern terrain interventions, which make the practical use of accurate elevation models for spatial analysis more difficult. Our study shows that the application of practical corrective methods to even large-scale DEMs, without high input costs (time or financial), produces a significant qualitative shift in the modeling of landscape properties.

Focusing on the territory of Bohemia, we extracted areas identified by vector topographic maps to clean the DEM of modern features. Due to the extent of the processed area, it was not possible to use procedures involving manual editing or cleaning. Although straightforward in principle, our approach is novel due to its practical applicability on the large scale. As our results show, this relatively simple method of DEM correction can fundamentally affect the properties of the model and its usefulness for subsequent analysis. Using filtering and interpolation, we managed to improve the model properties with respect to the visibility, drainage patterns, and geomorphological properties represented by the curvature.

Future improvements to the model should be aimed at bridging problems connected with removing certain typically modern large-scale landscape features (surface mining, large water bodies, urbanization zones, etc.). To that end, it will be necessary to create a yet-to-be-established databank of digitized historical elevation data based on legacy topographic maps to be used as a basis for the interpolation of accurate pre-industrial DEMs. We also recommend that the DEM raster grid be supplemented by a secondary layer with an estimate of the expected vertical resolution based on the filtering and interpolation methods, source data, and filtered features used. Considering the Czech context, enabling free access to the DMR 5G in the form of a point cloud would be another important step. This would allow researchers to select specific data points and subsequently triangulate missing data instead of making modifications to the raster grid. Notwithstanding these possible future measures, we consider that our filtered DMR 5G model, even in its present form, is the most suitable and best available centrally provided framework for archaeological use that can facilitate standardized derivations of total visibility, relief classification, and drainage networks on a national scale in the Czech Republic.

Acknowledgements

The authors would like to thank Alexandra Bucha Rášová, Martin Kuna and Peter Demján for their help, particularly their contributions to the theoretical discussions that informed the concept for this paper.

Funding: This work was supported by the European Union through the OP VVV programme administered by MEYS, within the project Archaeological Information System of the Czech Republic – Second Generation [Grant No. CZ.02.1.01/0.0/0.0/16_013/0001439] and by the Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences, VEGA [Grant Nos. 1/0300/19 and 1/0468/20].

REFERENCES

Bini, M. – Pappalardo, M. – Rossi, V. – Noti, V. – Amorosi, A. – Sarti, G. (2018) *Deciphering the effects of human activity on urban areas through morphostratigraphic analysis: The case of Pisa, Northwest Italy.* Geoarchaeology, Vol. 33, pp. 43–51.

Boháčová, I. – Herichová, I. (2004) *Georeliéf jako zdroj poznání historického vývoje lokality.* Forum urbes medii aevi, Vol. 1, pp. 10-17.

Brázdil, K. – Bělka, L. – Dušánek, P. – Fiala, R. – Gamrát, J. – Kafka, O. – Peichl, J. – Šíma**, J. (2016)** *Technická zpráva k Digitálnímu modelu reliéfu 5. generace: DMR 5G* Available at: https://geoportal.cuzk.cz/Dokumenty/TECHNICKA_ZPRAVA_DMR_5G .pdf, (accessed at 09/03/2022).

Brughmans, T. – Van Garderen, M. – Gillings, M. (2018) *Introducing visual neighbourhood configurations for total viewsheds.* Journal of Archaeological Science, Vol. 96, pp. 14–25.

Chase, A.S.Z. – Chase, D.Z. – Chase, A.F. (2017) *LiDAR for Archaeological Research and the Study of Historical Landscapes.* In: N. Masini, F. Soldovieri, eds. *Sensing the Past: From artifact to historical site.* Geotechnologies and the Environment, Vol. 16, Cham: Springer.

Crutchley, S. – Crow, P. (2009) *The light fantastic: using airborne laser scanning in archaeological survey.* Swindon: English Heritage.

David, B. – Thomas, J. (eds.) (2008) *Handbook of Landscape Archaeology*. London and NY: Routledge.

De Reu, J. – Bourgeois, J. – Bats, M. – Zwertvaegher, A. – Gelorini, V. – De Smedt, P. – Chu, W. – Antrop, M. – De Maeyer, P. – Finke, P. – Van Meirvenne, M. – Verniers, J. – Crombé, P. (2013) *Application of the topographic position index to heterogeneous landscapes.* Geomorphology, Vol. 186, pp. 39–49.

De Smedt, P. – Van Meirvenne, M. – Davies, N.S. – Bats, M. – Saey, T. – De Reu, J. – Meerschman, E. – Gelorini, V. – Zwertvaegher, A. – Antrop, M. – Bourgeois, J. – De Maeyer, P. – Finke, P.A. – Verniers, J. – Crombé, P. (2013) *A multidisciplinary approach to reconstructing Late Glacial and Early Holocene landscapes.* Journal of Archaeological Science, Vol. 40(2), pp. 1260–1267.

Garstki, K. (2020) *Digital Innovations in European Archaeology*. Elements in the Archaeology of Europe. Cambridge: Cambridge University Press.

Gibson, J.J. (1986) *The Ecological Approach to Visual Perception*. NJ: Psychology Press.

Gillings, M. (2017) *Mapping liminality: Critical frameworks for the GIS-based modelling of visibility.* Journal of Archaeological Science, Vol. 84, pp. 121–128.

Gojda, M. – John, J. (eds.) (2013) *Archeologie a letecké laserové skenování krajiny / Archaeology and airborne laser scanning of the landscape*. Plzeň: Katedra archeologie FF ZČU.

Habicht, H.L. – Rosentau, A. – Jöeleht, A. – Heinsalu, A. – Kriiska, A. – Kohv, M. – Hang, T. – Aunap, R. (2017) *GIS-based multiproxy coastline reconstruction of the eastern Gulf of Riga, Baltic Sea, during the Stone Age.* Boreas, Vol. 46, pp. 83–99.

Hengl, T. – Reuter – H.I. (eds.) (2009) *Geomorphometry: Concepts, Software, Applications*. Amsterdam: Elsevier.

Herichová, I. (2019) *Vrch hradní. Vývoj georeliéfu Pražského hradu v raném středověku.* Prague: Archeologický ústav AV ČR.

Hobič, J. (2020) *List of freely accessible LiDAR data and digital terrain models [online].* Archaeology of Slovenia. Available at: https://arheologijaslovenija.blogspot.com/p/ blog-page_81.html (accessed at 06/08/2021)

Höfler, V. – Wessollek, C. – Karrasch, P. (2015) *Modelling prehistoric terrain Models using LiDAR-data: a geomorphological approach.* I Proc. of SPIE, 9644, Earth Resources and Environmental Remote Sensing/GIS Applications VI, 96440B.

Krebs, P. – Stocker, M. – Pezzatti, G.B. – Conedera, M. (2015) *An alternative approach to transverse and profile terrain curvature.* International Journal of Geographical Information Science, Vol. 29(4), pp. 643–666.

Kuna, M. (2004) *Nedestruktivní archeologie: Teorie, metody a cíle*. Prague: Academia.

Kuna, M. (2006) *Burial mounds in the landscape.* In: L. Šmejda, ed. Archaeology of Burial Mounds. Plzeň: Dryada, pp. 83–97.

Kuna, M. – Danielisová, A. (2009) *Geomorfometrie a analýza reliéfu v archeologii.* Živá archeologie, Vol. 10, pp. 56–60.

Kuna, M. – Novák, D. – Bucha Rašová, A. – Bucha, B. – Machová, B. – Havlice, J. – John, J. – Chvojka, O. (2022) *Total viewshed in archaeology: problem solved, issues remain.* Journal of Archaeological Science, Vol. 142, pp. 105596.

Llobera, M. (2003) *Extending GIS-based visual analysis: the concept of visualscapes.* International Journal of Geographical Information Science, Vol. 17, pp. 25–48.

Llobera, M. (2007) *Reconstructing visual landscapes.* World Archaeology, Vol. 39, pp. 51–69.

Minár, J. – Evans, I.S. – Jenčo, M. (2020) *A comprehensive system of definitions of land surface (topographic) curvatures, with implications for their application in geoscience modelling and prediction.* Earth-Science Reviews, Vol. 211, pp. 103414.

Novák, D. (2018) *Wykorzystanie ALS do zautomatyzowanej analizy krajobrazu.* In: M. Gojda, Z. Kobyliński, eds. Lotnicze skanowanie laserowe jako narzędzie archeologii. Archaeologica Hereditas, 11. Warszawa: Instytut Archeologii Uniwersytetu Kardynała Stefana Wyszyńskiego w Warszawie, pp. 69–84 [in Polish].

Novák, D. (2019) *Funkční klasifikace a vývoj vrchnostenských sídel ve středověkých a raně novověkých* Čechách*.* Památky archeologické, Vol. 110, pp. 307–382.

Posluschny, A.G. (ed.) (2015) *Sensing the Past: Contributions from the ArcLand Conference on Remote Sensing for Archaeology.* Bonn: Habelt-Verlag.

Rášová, A. (2017) *Substanciálna analýza viditeľnosti v prostredí GIS*. Unpublished thesis (PhD), Slovak University of Technology in Bratislava [in Slovak].

Romstad, B. – Etzelmüller, B. (2012) *Mean-curvature watersheds: A simple method for segmentation of a digital elevation model into terrain units.* Geomorphology, Vol. 139-140, pp. 293–302.

Schmidt, J. – Werther, L. – Zielhofer C. (2018) *Shaping pre-modern digital terrain models: The former topography at Charlemagne's canal construction site.* PLoS ONE, Vol. 13(7), pp. e0200167.

Tarboton, D.G. – Schreuders, K.A.T. – Watson, D.W. – Baker, M.E. (2009) *Generalized terrain-based flow analysis of digital elevation models.* In: R.S. Anderssen, R.D. Braddock, and L.T.H. Newham, eds. *1*8th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*.* Christchurch: Modelling and Simulation Society of Australia and New Zealand, pp. 2000–2006.

Werbrouck, I. – Antrop, M. – Van Eetvelde, V. – Stal, C. – De Maeyer, Ph. – Bats, M. – Bourgeois, J. – Court-Picon, M. – Crombe, Ph. – De Reu, J. – De Smedt, Ph. – Finke, P.A. – Van Meirvenne, M. – Verniers, J. – Zwertvaegher, A. (2011) *Digital Elevation Model generation for historical landscape analysis based on LiDAR data, a case study in Flanders (Belgium).* Expert Systems with Applications, Vol. 38(7), pp. 8178–8185.

Zwertvaegher, A. – Finke, P. – De Reu, J. – Vandenbohede, A. – Lebbe, L. – Bats, M. – De Clercq, W. – De Smedt, P. – Gelorini, V. – Sergant, J. – Antrop, M. – Bourgeois, J. – De Maeyer, P. – Van Meirvenne, M. – Verniers, J. – Crombé, P. (2013) Reconstructing Phreatic Palaeogroundwater Levels in a Geoarchaeological Context: A Case Study in Flanders, Belgium. *Geoarchaeology*, 28, 170–189.

Zwertvaegher, A. – Werbrouck, I. – Finke, P.A. – De Reu, J. – Crombé, P. – Bats, M. – Antrop, M. – Bourgeois, J. – Court-Picon, M. – De Maeyer, P. – De Smedt, P. – Sergant, J. – Van Meirvenne, M. – Verniers, J. (2010) *On the use of integrated process models to reconstruct prehistoric occupation, with examples from Sandy Flanders, Belgium.* Geoarchaeology, Vol. 25, pp. 784–814.