USING THE AUTOMATIC VECTORISATION METHOD IN GENERATING THE VECTOR ALTIMETRY OF THE HISTORICAL VLTAVA RIVER VALLEY

Darina Kratochvílová*, Jiří Cajthaml

Czech Technical University in Prague, Faculty of Civil Engineering, Department of Geomatics, Thákurova 7/2077, 166 29 Prague, Czech Republic

* corresponding author: `darina.kratochvilova@fsv.cvut.cz`

Abstract. The article describes, in detail, the generation procedure of vector altimetry of an upper and middle course of the Vltava River historical valley. By the historical valley, the shape of the river valley before the construction of dams on the Vltava River in the second half of the 20th century is understood. The vector altimetry will serve as the base for creating a 3D model of this valley.

The initial input data were old maps, specifically scanned first-edition prints of the State Map 1 : 5 000 - Derived (SMO-5). By combining automatic and manual vectorisation, the altimetry component of these maps (contour lines and spot elevations) will be converted to a vector format. Individual processing steps, including the description of automatic vectorisation in the ArcScan extension to the ArcGIS system, together with examples of the results, are presented in the text below.

Keywords: Vltava, SMO-5, contour lines, automatic vectorisation, ArcScan.

1. Introduction

The Vltava River, the longest and, apparently, the most famous Czech river, stretches over a territory where significant changes occurred in the river landscape in the 20th century due to the construction of the so-called Vltava Cascade. The Vltava Cascade is a series of 9 dams built in the upper and middle course of the river. The construction of the first dam in Vrané started in 1930, and the whole cascade was completed in the early 1990s [1]. The construction of the dams resulted in the destruction of many settlements and historically renowned sites (e.g. Saint Johann's Rapids).

The objective of the Ministry of Culture NAKI DG18P02OVV037 project, "Vltava – transformation of historical landscape as a result of floods, dam's creation and land-use changes along with cultural and social activities in the river neighbourhood" is to present, in a complex way, the changes in the landscape occurring in the upper and middle course of the Vltava River. As there are numerous diverse archival materials available thanks to which we can gain a lot of information about the original Vltava River stream, the project will produce a self-contained information system containing processed archival materials, such as, above all, old maps, plans of historic structures, photographs, historical documents, etc. [2]. This system will serve both the professional and the general public as a source of information about the changes of the river and its surrounding territory, changes in its function and exploitation. One of the outputs will be a 3D model of the historical Vltava River valley, created on the basis of the vector altimetry obtained from old maps [3]. Apparently, old maps are one of the main sources of information for the reconstruction of the now extinct landscapes and for the creation of historical digital terrain models [4, 5].

The aim of this research is to find a suitable method for creating the vector altimetry of the Vltava River historical valley from old maps. The principal source of data for the creation of the historical Vltava River valley model are old maps, dating back to the period preceding the construction of dams. With a view to the size of the territory – encompassing the Vltava River course from the source to the confluence with the Berounka River – such a series of maps must be chosen that will cover the whole area of interest and contain the altimetry component, necessary for the creation of a territorial model. These conditions are met by a large-scale state map series, the State Map 1 : 5 000 - Derived (SMO-5) in its first edition [6].

2. Methods

2.1. State Map 1 : 5 000 - Derived

SMO-5 was published gradually from 1950. At first, it was to serve as a temporary replacement of the State Map of Czechoslovakia 1 : 5 000, whose creation was not sufficiently fast due to the fact that it was elaborated as an original map. SMO-5, on the contrary, originated exclusively by rewriting the existing topographic base, not by direct measurement. Thus, it complied with the requirement for covering the whole territory of the state of that time in as little time as possible. The planimetric map component in the first edition was derived from cadastral maps, while the altimetry component from topographic maps in the S-1952 system, or, in some cases, also from topographic profiles of the 3rd military mapping. The altimetry,

Figure 1. Representation of processed map sheets in the SMO-5 sheet line system.

Contour interval	Representation
1 m	2.7 % of territory
2 m	8.2 % of territory
5 m	5.7 % of territory
10 m	59.8 % of territory
20 m	23.6 % of territory

Table 1. Contour intervals on processed SMO-5 map sheets.

which will be the subject of the successive processing, was marked in brown comprising contour lines, spot elevations and, potentially, technical or topographic hachures and descriptions. The elevations are in the Baltic Vertical Datum after Adjustment. The map sheets have dimensions of 2.5×2 km and are in the 90-degree angle sheet line system projected in the S-JTSK coordinate system. SMO-5 was issued in several other editions, and until 1990, it was only available for the state purposes. These days, the first edition of the SMO-5 map can be bought as an archival document on the Geoportal of the Czech Office for Surveying, Mapping and Cadastre [7], with individual map sheets scanned in a JPEG format.

The area of interest, in our case, consists of 334 map sheets of the SMO-5 first edition, copying the upper and middle Vltava River course, as illustrated in Figure 1.

According to the off-frame map data, the altimetry in the processed map sheets was, in most cases, derived from topographic maps in a scale of 1 : 10 000 and 1 : 25 000. Unfortunately, the contour intervals that appear on the selected map sheets vary from 1 m to 20 m. Their proportion in the overall surface area of the territory are shown in Table 1. Different contour intervals could also be found within the individual map sheets.

The selected map sheets of the SMO-5 first edition represent the initial input data for the modelling of the Vltava River historical valley. The map sheets were digitized to the raster format by scanning [8].

2.2. Automatic vectorisation

To create a model, however, it is necessary to convert the altimetry content of the map sheets to the vector format. Raster data may be, in general, vectorised manually, semi-automatically or automatically [9]. The suitable method must be selected with a view to the type, amount and quality of input data.

Manual vectorisation runs completely under the user's control, it may be carried out in standard GIS software without special knowledge, the raster base may be of poorer quality, nevertheless, this method is rather time consuming for large amounts of data. The semi-automatic vectorisation also runs under the user's control, even in this case, the user must control the vectorisation process by the mouse, nevertheless, capturing raster images or automatic tracing of raster cells with successive vectorising is possible by using special software tools. Hence, semi-automatic vectorisation, again, seems to be more convenient for the vectorisation of small amounts of data. Unlike the previous two methods, automatic vectorisation consists in the setting of input parameters and a successive automated vector generation. The result of automatic vectorisation depends on the suitability and the quality of the input raster. Automatic vectorisation requires a more advanced setting by the user, which must mostly be preceded by the modification of the input raster, nevertheless, with a correct setting of the parameters and a suitable base material, this method allows a vectorisation of even larger volumes of data.

Examples of the usage of automatic vectorisation of scanned base maps can be found, for example, in [10–14]. In [10], the automatic vectorisation of raster historical maps with a use of open source software, such as the Geospatial Data Abstraction Library and QGIS Desktop, is described. In conclusion, there is mentioned that it is necessary to adapt the parameters used in each step of the procedure to the specifics of different maps and that the most important step of this procedure is the image processing before the automatic vectorisation. Automatic vectorisation of contour lines from scanned topographical maps is presented in [11]. The authors describe colour classification, thinning of binary rasters and the vectorisation using a Delaunay Triangulation, they also discuss problems like poor condition of maps or the situation when the same colour is used to represent contours and elevation numbers. The procedure in [12] combines automatic and interactive techniques of vectorisation of the colour map. It allows to correct data immediately if required and to minimise the quantity of digitization errors. The article [13] describes the whole process of extracting the contour lines and further creation of a digital elevation model from scanned topographic maps. The authors use HSV colour model for contour segmentation from the scanned image and WinTopo software for automatic vectorisation of the binary and thinned raster. In [14], an automatic vectorising sys-

tem with respect to various maps characteristics is proposed. Used input data are only binary rasters.

Due to the overall volume of vectorised data, the automatic vectorisation method with a successive manual adding of attributes and potential manual editing was chosen in our case. There are various programmes for automatic vectorisation of raster base e.g. Win-Topo [15], Scan2CAD [16], R2V [17]. In our case, the automatic vectorisation was carried out in the ArcScan extension to the ArcGIS 10.5 programme. Fine examples of a usage of ArcScan for automatic vectorisation can be found in [18, 19]. In [18], ArcScan is used for creating the historical vectorised boundaries of African British Colonies from the maps, which were taken from the British Colonial Reports. The vectorised boundaries were conflated to the boundary data from the GADM website. In [19], it is described how the geologic maps for Geologic Hazard Mapping could be digitized with ArcScan and this method is compared with manual vectorisation. The paper also deals with the conversion of raster to two colours using an image processor.

The ArcScan extension is a part of the ArcGIS programme starting from the 8.3 version allowing the user to vectorise a raster base by an automatic or semi-automatic method, at the same time, simple editing raster modifications can be carried out to improve the quality of the resulting vector file [20]. The condition for the raster processing in ArcScan is a binary (two-colour) raster.

The paragraphs below describe individual steps leading from the raster input data to the resulting vector altimetry.

3. Results

Prior to starting the vectorisation process, the raster files had to be modified. The modification can be divided into three phases: raster placement in a co-ordinate system, raster modifications necessary for working in ArcScan (i.e. to enable raster processing in ArcScan at all – conversion to a binary raster) and raster modifications directly in ArcScan (raster pre-processing – using ArcScan tools for editing and final raster clean-up). The modification process of raster files is schematically shown in Figure 2.

First, the scanned map sheets had to be placed in a coordinate system. In our case, they were geo-referenced into S-JTSK (Czech national coordinate system). The transformation of all sheets was carried out by the Projective Transformation Method using the functions of the ArcGIS programme. This type of transformation is often used in the georeferencing of old maps with rectangular map sheets [21]. At least 4 identical points must be determined to solve the transformation parameters, in our case, these were the 4 corners of the map sheets. Root mean square errors of the affine transformations using these 4 points were tested for quality control. The error never exceeded 5 meters, which means a very good quality of our

transformation results. The georeferenced map sheets were successively connected into a seamless mosaic, which will further be used in the resulting visualization of the valley. The map sheets were processed individually for the vectorisation needs.

3.1. Conversion of raster base materials to binary rasters

The georeferenced map sheets had a colour depth of 16.7 million colours, which was the same as the colour depth of the input data. The batch conversion of all files to the TIF format was defined in the Model-Builder environment in ArcGIS, and, at the same time, their colour depth was reduced to 256 colours (8-bit raster), which was selected as the optimum solution for the successive processing. Together with the creation of new raster files, a void vector file was automatically generated for each of the map sheets into which the vectorisation result would be later saved from the ArcScan programme.

Due to the size of the processed territory of interest, individual map sheets differed by their colour, see Figure 3, which made their successive processing complicated, mainly in the phase of the conversion to a two-colour raster. Thus, each sheet had to be treated individually.

As mentioned above, the ArcScan extension requires a two-colour raster file as the input (one colour for the foreground – contour lines, the other for the background). This file was created by the settings of the raster file symbology. The colours of pixels (shades of brown) occurring in the contour lines were selected manually, and a group forming the foreground was created from this selection. The remaining colours were also joined in a group, they formed the background and were recoloured white. It was not necessary at all to select all the colours represented in the contour lines as the appearance of a contour line could be successively automatically cleaned up, or missing parts of a drawing completed by using the ArcScan tools for the raster modifications. However, the colours that occurred both in the contour lines and in the planimetry caused problems, it was convenient to incorporate these colours in the background, if possible, so that they would not add an unnecessary noise in the successive processing. Other problematic spots were the intersections of contour lines with planimetric features or the spots with minimum distances between contour lines (steep slopes) where the contour lines basically merged. The same shade as that of the contour lines (i.e. the foreground colour) also occurred in some other components of the altimetry, which were not desirable to be automatically vectorised, such as contour line descriptions, spot elevations, rocks and hachures, etc. These objects had been, to a greater part, removed during the raster processing in ArcScan or later, during the vector file manual editing. The result of this phase is a two-colour raster, which can
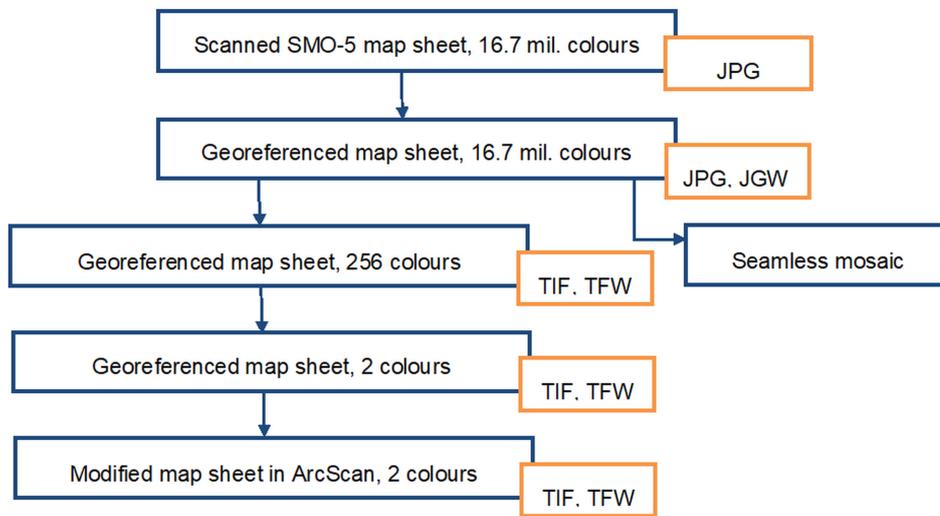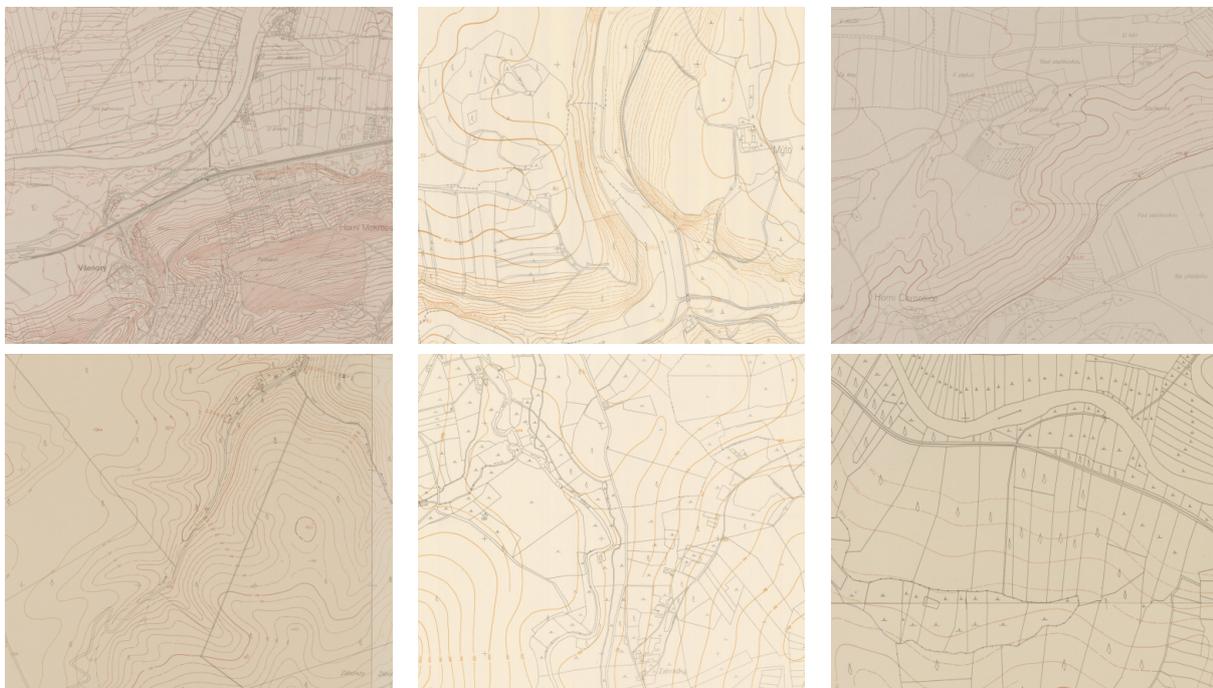
FIGURE 2. Modification process of raster files.

FIGURE 3. Different colour and quality of input files.

be opened in the ArcScan extension to be modified and successively vectorised there.

3.2. Modifications of rasters in ArcScan

First, the objects which would not be vectorised (e.g. descriptions, hachures, rocks) were deleted using automatic as well as manual tools. A tool selecting all connected cells containing smaller than user-defined numbers of pixels allowed an automatic selection and successive deletion of isolated descriptions and slight noises. The classic erase tool was available for manual deletion of objects (rocks, hachures, etc.). To improve the contour line drawing, two of the offered morphological operations were used, in particular Dilation and Closing. By these operations, the missing parts of contour lines were added, and potential holes closed. The resulting changes in the raster were saved.

The quantity of manual raster adjustments was individual for each map sheet. The views of individual phases of the raster processing are shown in Figure 4 below. This Figure 4 captures an ideal area without problematic spots. The vector file from ArcScan in this case will require only minimal manual editing.

As mentioned above, one of the problematic spots in automatic vectorisation was the contour line descriptions. Figure 5 shows the successful processing of contour line descriptions. In the case of sufficiently distant contour lines, it was possible to modify the raster file using morphological operations in ArcScan so that the resulting vector contour line was continuous. If it was not appropriate to use a morphological operation, it was possible to remove the descriptions either by editing the raster in ArcScan or by editing the vector file.

Figure 6 shows the problematic spot for the automatic vectorisation. Rocks had to be erased manually when editing the raster file in ArcScan. Dashed contour lines and a small interval of contour lines in a steep terrain cause a discontinuous vector line in many places. In such cases, further manual editing of the vector file was more time consuming.

3.3. Vectorisation in ArcScan

After the completion of the raster base modifications, the vectorisation itself could be performed. The Centreline Method where the resulting vector features are generated at the centre of the connected raster cells of the foreground was used. Before starting the vectorising process, ideal parameter settings had to be found that would be usable for the majority of the map sheets. The ideal values of the parameters were determined on the basis of testing a selected set of the map sheets. The map sheets from different parts of the territory, with different terrain types and contour intervals were tested. The values of the parameters illustrated in Figure 7 were chosen as the most appropriate with respect to the resulting quality of the vector output. These settings were saved as the

"project's parameter setting". For some map sheets, it was necessary to adjust the parameters individually because it was necessary to improve the resulting vector data.

Prior to the resulting generation of features in the selected file, the vectorisation view was always displayed according to which additional modifications of the parameters could be performed, or the raster base further modified if necessary. The resulting vector data were saved to the respective 2D file (the files had been prepared in a batch during the modifications of rasters) – Figure 4.

The quality of the vector output depended on the quality of the raster base and its successive processing, and also on the contour interval. At spots with steep terrain slopes and a small contour interval, individual contour lines tended to merge into each other even before, in the raster base. Thus, the vector contour lines at these spots were often discontinuous, with numerous imprecisions requiring a lot of interventions during the successive manual vectorisation. Problems in the vectorisation also resulted from contour lines drawn by a dashed line on some of the map sheets.

3.4. Modifications of vector data

All vector files generated in the ArcScan programme were successively manually modified using the ArcGIS standard editing tools. For each contour line, or all of its potential parts, the elevation according to the raster base had to be added in the attribute table. At the same time, the vectorisation of spot elevations from all map sheets was performed in the point shapefile. Errors in the descriptions of the contour lines in the base maps were revealed at several spots during the vectorised data treatment, see Figure [8].

Following the modifications of the map sheets, the vector files were connected into a geodatabase. Successively, transitions at the contact of individual map sheets where the contour drawing was not continuous due to the previous processing in single map sheets had to be solved. The "Clean Contour Gaps" tool in the ETGeoWizards programme [22] was used for interconnecting the contour lines between the map sheets. This tool can automatically fill in the gaps between linear features which are smaller than the present value. At the same time, the necessary condition that must be fulfilled is that the parts of the line that are to be connected have the same value of a selected attribute, in our case the elevation. When merging map sheets, only contour lines with same elevation were interconnected. Some contour lines remained disconnected in the case of different contour intervals at the contact edges of map sheets. To make the contour line pattern at the contact edges of map sheets smoother, or to avoid creating loops, the contour lines were first cut at a distance of 10 m from the original map frame on each map sheet. Successively, the above tool was run, which added the missing parts of contour lines, both at the contact edges of map sheets and at spots where
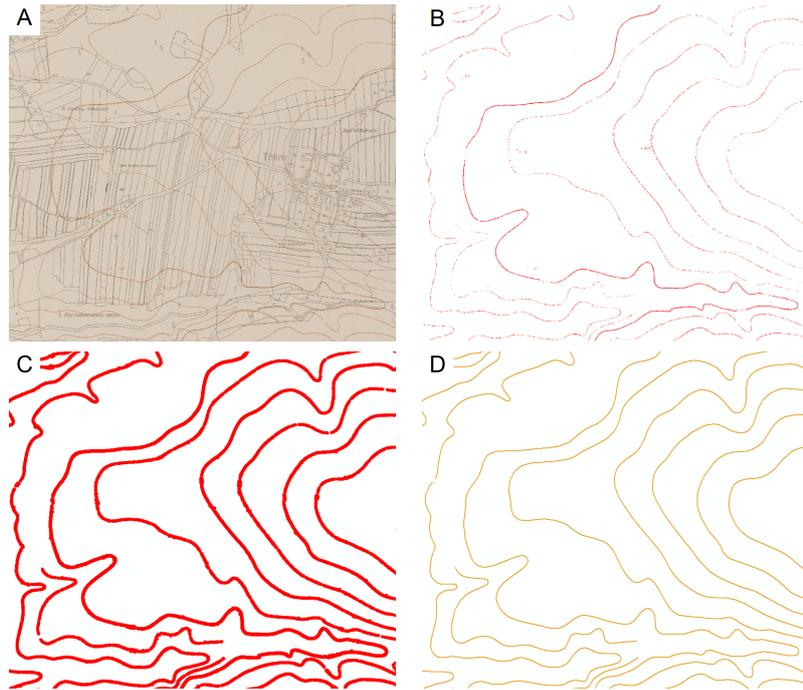
FIGURE 4. Illustration of raster processing. A - raster file, 256 colours, B - raster file, 2 colours, C - raster file after modifications in ArcScan, D - vector file from ArcScan.
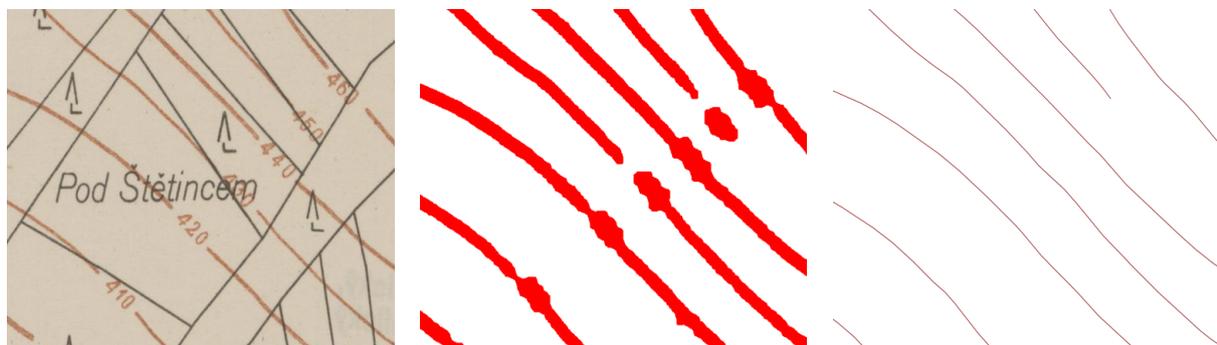

FIGURE 5. Example of processing contour line descriptions (raster file 256 colours, raster file after modifications in ArcScan and vector file from ArcScan).
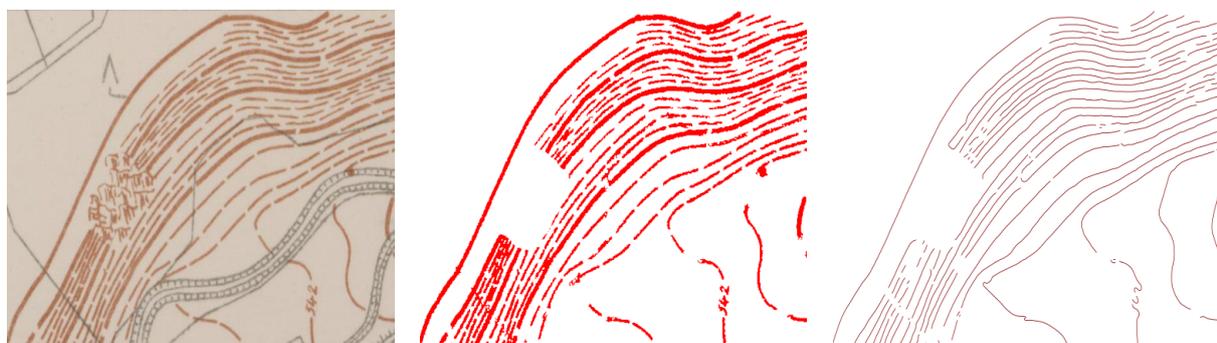

FIGURE 6. Example of processing some problematic spots, such as rocks and dashed contour lines (raster file 256 colours, raster file after modifications in ArcScan and vector file from ArcScan).

FIGURE 7. Parameters of automatic vectorisation in ArcScan.

gaps had remained after the automatic vectorisation. Using this tool, approximately 96 % of the gaps were filled. The treatment procedure at the contact edges of map sheets is shown in Figure 9. The application of this tool, at the same time, helped to reveal errors in wrongly assigned elevations to contour lines – these contour lines remained discontinuous. During further checks of the resulting data, the correct elevation was assigned to these contour lines and the remaining gaps were manually connected. To conclude, the "Clean Pseudo Nodes" tool was used, which interconnected the adjacent sections of the contour lines.

The above-mentioned tools and procedures allowed the vectorisation of contour lines and spot elevations on all selected SMO-5 map sheets, as is illustrated in Figure 10. These vector data will be subsequently used for the creation of a 3D model of the historical Vltava River valley. Alongside that, a polygon layer was generated dividing the whole area of interest into partial areas according to the contour interval on the base maps. This layer will serve to compare the created 3D model with the model created from the current data of the Digital Terrain Model of the Czech Republic of the fifth generation (DMR 5G) [23]. The accuracy statistics for individual contour intervals will be processed.

4. DISCUSSION

The resulting vector data were obtained by combining automatic and manual vectorisation. Due to the large amount of data, the original intention was to

use the automatic vectorisation. However, for such processing it was important to significantly adjust the input raster data. These adjustments were either applied together to every raster or individually for each map sheet. The rasters first had to be converted to a two-colour bitmap, which then enters the ArcScan processing. Due to the different colours in the input rasters, it was necessary to approach each raster individually. The conversion of the two-colour raster to vector data took place in the ArcScan extension of ArcGIS. Although optimally set conversion parameters can be used for most rasters, there are still bottlenecks that need to be addressed. The extent to which vectorisation had to be manually applied varied depending on the quality of the ArcScan output vector file. Also important was the state of the original raster file and how it was edited before being imported to ArcScan. When editing vector data, it was possible to use the tools ETGeoWizards, which helped especially when connecting parts of contour lines. It was necessary to manually assign altitude to each contour line. The most problematic places, especially the terrain with a high slope, had to be vectorised manually, because there was a graphic conflict of individual contour lines. The proposed vectorisation procedure is very well applicable to maps similar to SMO-5 maps.

When creating the 3D model, it is necessary to take into account that the altimetry on the map sheets was often derived from maps of smaller scales. The altimetry on these maps has, above all, an informative
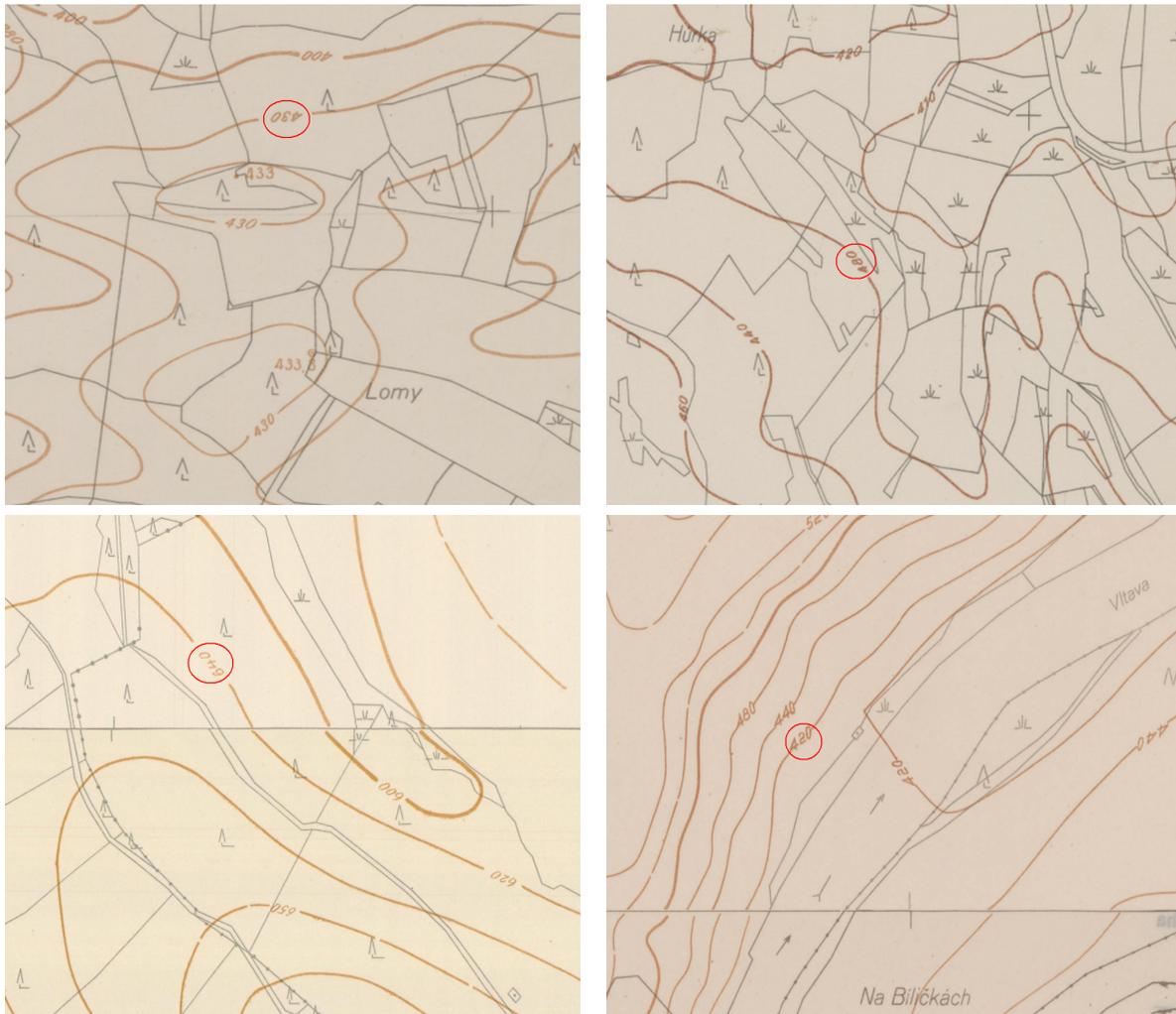
Figure 8. Examples of errors in base map data.

Figure 9. Modification of contour lines at the contact of map sheets.

FIGURE 10. SMO-5 base data and the resulting vector altimetry.

function and it cannot be relied on, due to not having the geometric accuracy of contours. Another complicating factor is the inconsistent interval of contour lines. In some localities, the interval of contour lines takes values of 20 m, which leads to an insufficient display of terrain. However, these old maps are, for us, the principal source of data for the creation of the historical Vltava River valley model. The processed accuracy statistics for individual contour intervals will serve for an assessment of the accuracy of the source data.

5. CONCLUSION

The ArcScan automatic vectorisation process itself is fast, the setting of the vectorisation input parameters is intuitive, and, in the case of similar data, the same settings for all such data can be advantageously used. If the input data are two-colour raster files (e.g. contour plans, cadastral maps, construction plans, etc), the automatic vectorisation in ArcScan is the most effective, because the adjustments of the rasters before the vectorisation are minimal or none. However, in most cases, we have colour raster files that must be converted to a binary raster for automatic vectorisation in ArcScan, which can sometimes be time consuming. It would be appropriate to try to automate raster modifications (particularly the conversion to a two-colour raster).

The method of automatic vectorisation should be chosen if we need to convert a large amount of scanned data (in our case old maps) into a vector form. This method can bring significant time savings, but the results of automatic vectorisation depend on many factors (primarily on the type and quality of the input raster data). Automatic vectorisation can always be complemented with manual or semi-automatic vectorisation, whereby we can have more control over the final form of the vector file.

REFERENCES

[1] Povodí Vltavy State Enterprise. Vltava cascade. http://www.pvl.cz/vodohospodarske-informace/vodni-dila/vltavska-kaskada.

[2] J. Cajthaml, D. Kratochvílová, T. Janata. 3D model of historical Vltava River valley: Combination of sources. In *Proceedings of the ICA*, vol. 2, pp. 1 – 5. 2019. DOI:10.5194/ica-proc-2-14-2019.

[3] J. Cajthaml, P. Tobiáš, D. Kratochvílová. Creating of 3D model of the historical landscape from contour lines displayed on old maps: Vltava River valley. In *ESaT 3rd International Conference on Engineering Sciences and Technologies*. Tatranské Matliare, 2018.

[4] A. de Boer. Processing old maps and drawings to create virtual historic landscapes. *E-Perimetron* **5**(2):49 – 57, 2010.

[5] J. Pacina, B. Bruna, J. Cajthaml. Complete 3D landscape reconstruction based on historical data sources. In *15th International Multidisciplinary Scientific GeoConference SGEM 2015*, pp. 547 – 554. Albena, 2015. DOI:10.5593/SGEM2015/B21/S8.068.

[6] J. Pták. Origin and purpose of State Map 1:5000 - derived. *Zeměměřictví* (2):146 – 148, 1952.

[7] Geoportál ČÚZK. State Map 1:5000 - derived. `https://geoportal.cuzk.cz/ (S(yvay04g542pszf0rd04nmfdm))/Default.aspx? mode=TextMeta&side=dSady_archiv&metadataID= CZ-CUZK-SMO5-R&menu=2905`.

[8] M. Talich, F. Antoš, O. Böhm. Automatic processing of the first release of derived state maps series for web publication. In *25th International Cartographic Conference (ICC2011) and the 15th General Assembly of the International Cartographic Association.* Paris, 2011.

[9] B. E. Davis. *GIS: A Visual Approach.* Cengage Learning, 2nd edn., 2001.

[10] I. Iosifescu, A. Tsorlini, L. Hurni. Towards a comprehensive methodology for automatic vectorization of raster historical maps. *E-Perimetron* **11**(2):57 – 76, 2016.

[11] S. Salvatore, P. Guitton. Contour line recognition from scanned topographic maps. *Journal of WSCG* **12**(1 - 3):419 – 426, 2004.

[12] S. Ablameyko, V. Bereishik, M. Homenko, et al. Automatic/interactive interpretation of color map images. *Object recognition supported by user interaction for service robots* **3**:69 – 72, 2002. DOI:10.1109/ICPR.2002.1047797.

[13] S. Oka, A. Garg, K. Varghese. Vectorization of contour lines from scanned topographic maps. *Automation in Construction* **22**:192 – 202, 2012. DOI:10.1016/j.autcon.2011.06.017.

[14] K.-H. Lee, S.-B. Cho, Y.-C. Choy. Automated vectorization of cartographic maps by a knowledge-based system. *Engineering Applications of Artificial Intelligence* **13**(2):165 – 178, 2000. DOI:10.1016/S0952-1976(99)00049-4.

[15] WinTopo. Raster to vector converter. `http://wintopo.com`.

[16] Scan2CAD. Raster to vector software. `https://www.scan2cad.com`.

[17] Able Software. R2v - raster to vector conversion, gis, mapping, cad. `http://www.ablesw.com/r2v`.

[18] T. W. Shawa. Guidelines for creating historical geospatial boundary data. *E-Perimetron* **12**(1):24 – 36, 2017.

[19] B. Erickson. Geologic hazard mapping & arcscan. In *Digital Mapping Techniques.* Salt Lake City, 2015.

[20] P. Sanchez. *Using ArcScan for ArcGIS: ArcGIS 9.* Redlands: Esri Press, 2004.

[21] J. Cajthaml. *Analysis of old maps in digital environment on thew example of Müller's map of Bohemia and Moravia.* CTU in Prague, 2012.

[22] ET GeoWizards for Windows. Gis tools from et spatialtechniques. `https://www.ian-ko.com/ETGeoWizards.html`.

[23] Geoportál ČÚZK. Digital terrain model of the Czech Republic of the fifth generation. `https://geoportal. cuzk.cz/(S(12bh0pl5lv0it05mgzdfcre3))/Default. aspx?mode=TextMeta&side=vyskopis&metadataID= CZ-CUZK-DMR5G-V&head_tab=sekce-02-gp&menu=302`.