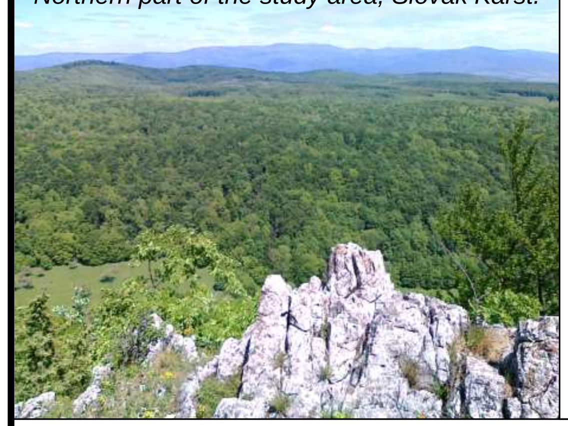
Spatial Interpolation of Airborne Laser Scanning Data with Variable Data Density

Jaroslav Hofierka – Michal Gallay – Ján Kaňuk

Institute of Geography, University of Pavol Jozef Šafárik in Košice, Slovakia

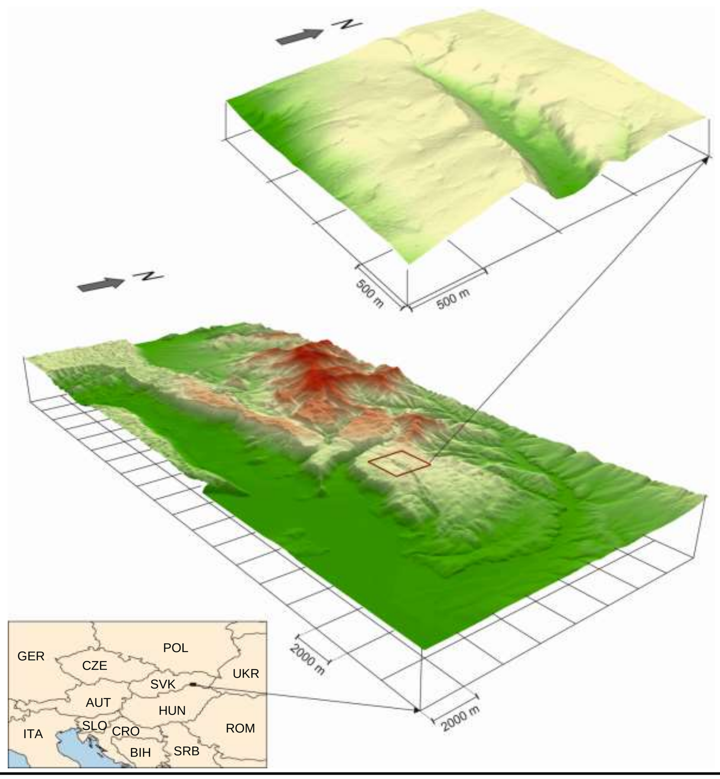
Northern part of the study area, Slovak Karst.



Airborne laser scanning (ALS) data are increasingly available for various applications including digital elevation modeling. One of the biggest challenges for its successful use is the varying data density caused by land cover properties. Areas with dense canopy cover have much lower data density than the open areas. This varying data density may pose a problem for some spatial interpolation methods, such as Regularized Spline with Tension.

In this study, we propose a methodology to eliminate the interpolation artifacts caused by the varying data density by a proper selection of data points entering the interpolation process. The number of data points is controlled based on the minimal distances between the points as well as the selection of the most important points derived from the analysis of interpolation errors. The proposed methodology has been demonstrated using the application example from the Slovak karst area.

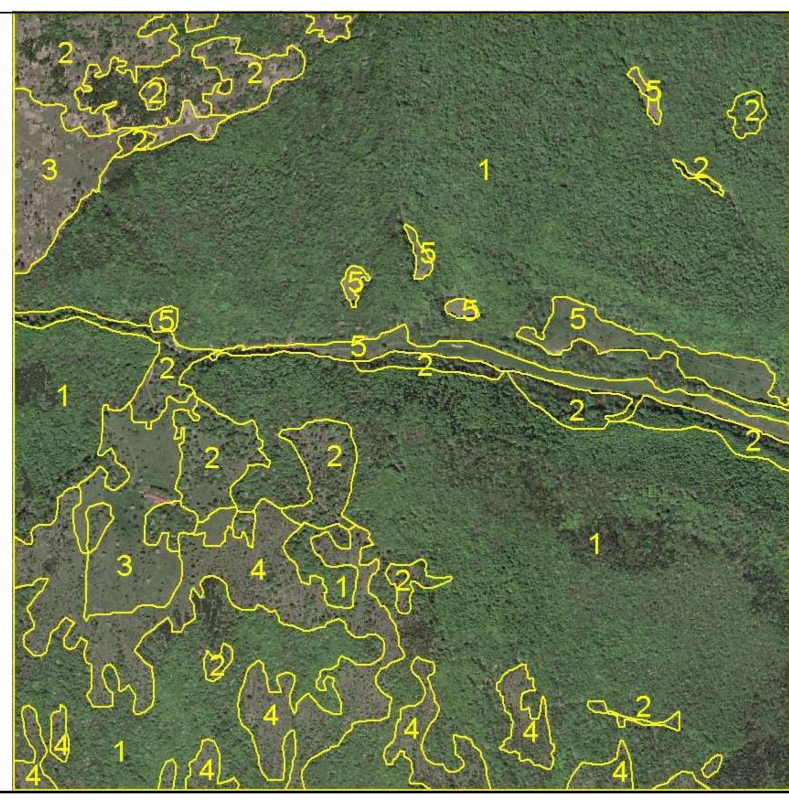
Study area and input data



The ALS data used in this study represent a 2 by 2 km portion of the Slovak Karst, East Slovakia. The area is mostly wooded with occasional meadows and scrubs comprising a plateau dissected by a deep canyon and a few occasional dolines. The altitude ranges between 540 to 704 meters a.s.l.

The data were acquired in leaf-on conditions in 2009 and supplied as a filtered point cloud (LAS). The supplier claims the vertical root mean square error (RMSE) of 23 cm. The sample area contains 217 984 points classified as bare ground of spacing varying between 1–80 m with the average data density around 0.054 point/m^2.

Spatial distribution of points is very uneven with a higher data density associated with open grass land with the average data density around 0.16 point/m^2 and the lowest data density in closed forest areas with dense canopy cover (0.03 point/m2).



Orthophotomap of the study area with land cover classes:

1 – closed forest
2 – open forest
3 – grass with sparse trees
4 – grass and shrubs
5 – grass

Regularized Spline with Tension

$$f(\mathbf{x}) = a_1 + \sum_{j=1}^{N} \lambda_j \{-[E_1(\rho) + \ln\rho + C_E]\}$$

This interpolation function is implemented in ***GRASS GIS*** as the ***v.surf.rst*** command (Neteler and Mitasova, 2008). The interpolation process is controlled by the following set of parameters: tension, smoothing, anisotropy, minimum and maximum distances between points. The number of points used for interpolation is controlled by four parameters: ***dmin*** – minimum distance between points (to remove almost identical points), ***dmax*** – maximum distance between the points on isoline (to insert additional points), ***segmax*** - defining the maximum number points in an interpolation segment, and ***npmin*** – minimum number of points used for interpolation in a segment (Neteler & Mitasova, 2008). The segmentation procedure of the *v.surf.rst* module divides the whole area into a set of overlapping segments to ensure a smooth connection of the segments to form the final surface. These parameters can be selected empirically, based on the knowledge of the modeled phenomenon, or automatically, by minimization of the predictive error estimated, for example, by a cross-validation procedure (Hofierka, 2005, Hofierka et al., 2007).

Evaluating the interpolation quality

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|z_i^* - z_i\right|$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(z_i^* - z_i\right)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i^* - z_i)^2}$$

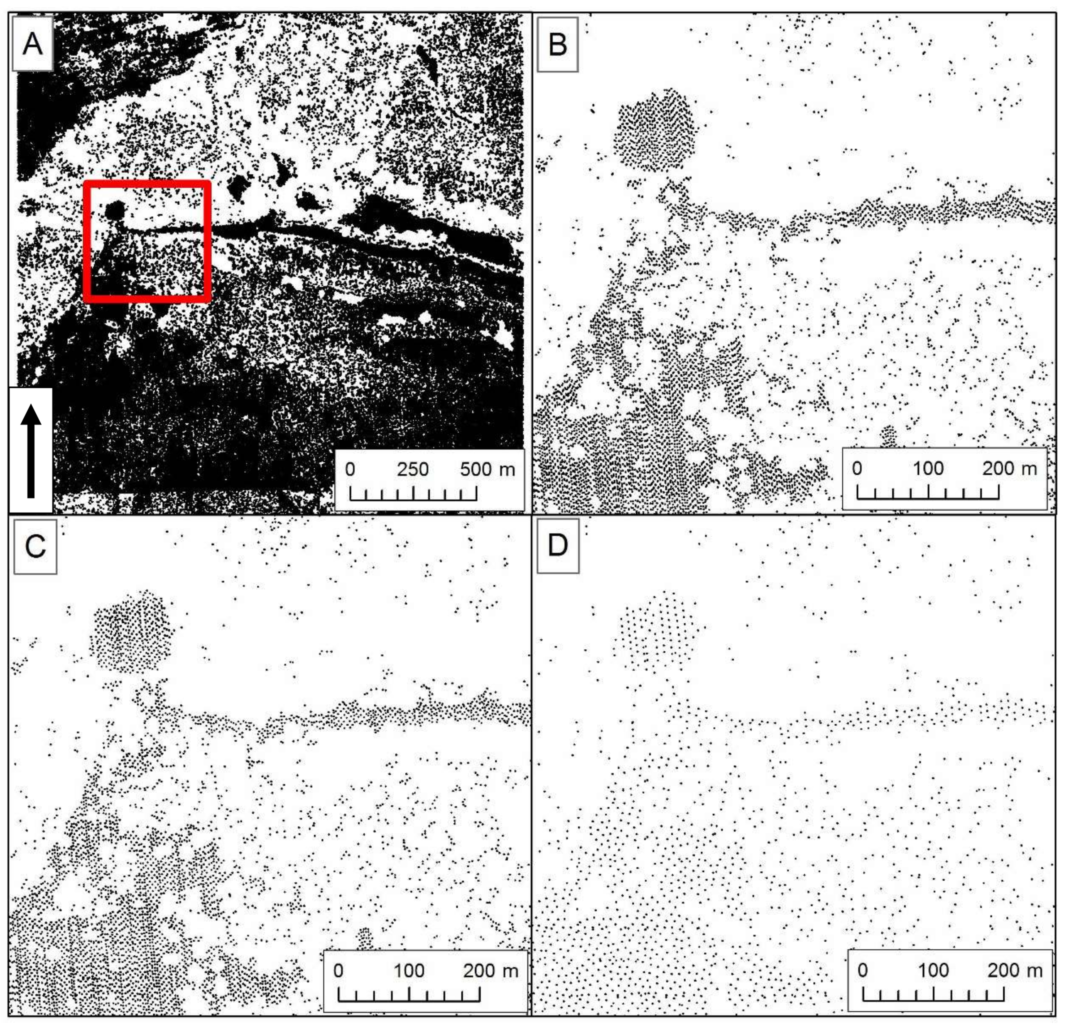
z^*_i - interpolated (estimated) value at location *i*

z_i - actual (measured) value.at location *i*

The **cross-validation (CV)** procedure is based on removing one input data point at a time, performing the interpolation for the location of the removed point using the remaining samples and calculating the residual between the actual value of the removed data point and its estimate. The procedure is repeated until every sample has been, in turn, removed. This form of CV is also known as the ''leave-one-out'' method (Hofierka et al., 2002).

CV is especially suitable for relatively dense data sets, as removing points from already under-sampled areas can lead to misrepresentation of the surface to be interpolated (the surface is smoothed). The minimum statistical errors calculated by CV can be used to find the optimum interpolation parameters (Mitasova et al., 1995; Hofierka et al., 2002). Hofierka et al. (2007) have suggested that the evaluation of interpolation accuracy can be also assessed using an evaluation dataset containing data not used in interpolation. The error between actual and interpolated value is calculated for each evaluation point and the overall accuracy is tested. This evaluation dataset can be taken from independent measurement of points or from the original dataset using points selected by a random generator. The selected points are then removed from the interpolation dataset.

Results



Input ALS points. A – entire study area. B – detail of the area in the red square on A. C – reduced dataset with a minimal spacing of input data set to 2 m, D – reduced dataset with a minimal spacing of input data set to 6 m.

[DTM] v.surf.rst parameter setting	NP	RE	ME	MAE	RMSE	CT
[1] tension=20, dmin=1.0, npmin=300	183 418	17.23	-0.000028	0.1068	0.1596	65
[2] tension=20, dmin=2.0, npmin=400	136 502	16.05	-0.000022	0.1191	0.1797	120
[3] tension*=185, dmin=6.0, npmin=400	44 682	7.29	0.000011	0.0959	0.1432	23
[4] tension*=185, dmin*=6.0, npmin=400	62 640	13.98	0.000015	0.1456	0.2196	44

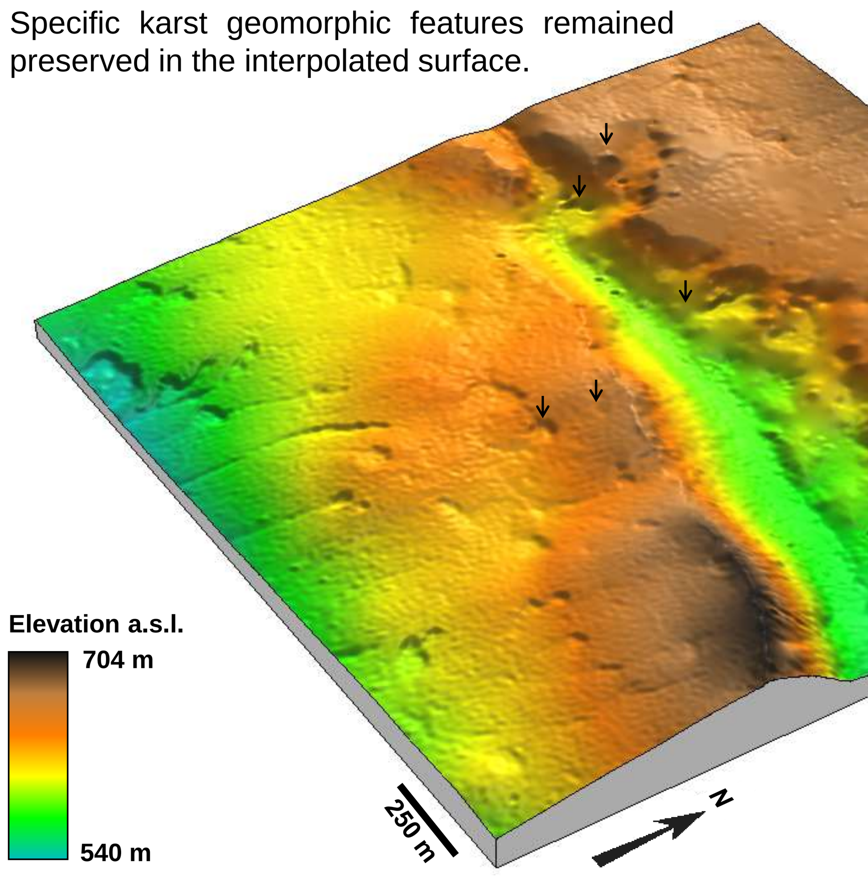
Table 1. Interpolation accuracy at given points (tension* - unnormalized tension, dmin* - selective dmin, NP - number of points, RE - range of errors, ME - mean error, MAE - mean absolute error, RMSE - root mean square error, CT - computational time in minutes).

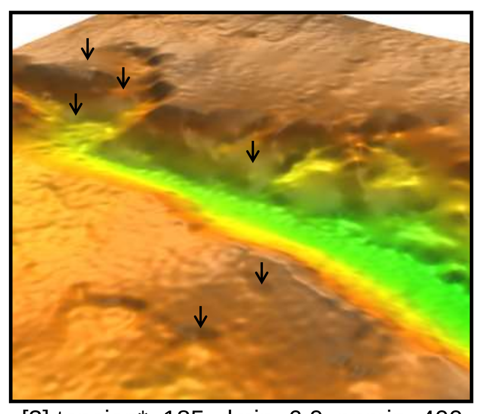
[DTM] v.surf.rst parameter setting	ME	MAE	RMSE
[5] tension*=185, dmin=2.0, npmin=400	-0.02125	0.1676	0.2447
[3] tension*=185, dmin=6.0, npmin=400	-0.02163	0.1782	0.2592
[4] tension*=185, dmin*=6.0, npmin=400	-0.01137	0.1570	0.2195

Table 2. Interpolation accuracy using an evaluation dataset of 1000 randomly selected points.
NB: The unnormalized tension of DTM [5] is equivalent to the normalized tension of DTM [2] in Table 1.

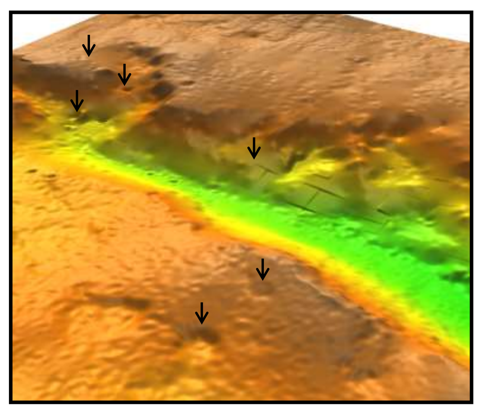
Resulting DTM [4]

Specific karst geomorphic features remained preserved in the interpolated surface.





[3] tension*=185, dmin=6.0, npmin=400



[1] tension*=185, dmin=6.0, npmin=400

Comparison of interpolation accuracies achieved by the suggested set of parameters is presented in Table 1 and Table 2. While Table 1 presents the interpolation accuracy at given points, Table 2 presents the predictive error of the interpolation method with various parameterisation settings using an evaluation set of 1000 randomly selected points withheld from further interpolation. The comparison clearly shows the drawbacks of evaluation of interpolation accuracies at given points. The best results by RMSE were achieved for the setting with dmin=6 meters (RMSE=0.1432 m). However, Table 2 shows that the best results were achieved for dmin*=6 meters with additional points with the highest interpolation errors identified during the computation with dmin=2 meters. In contrast, the best result identified in Table 1 is the worst in Table 2. The overall interpolation results of the RST method and v.surf.rst module are very good because in all cases the RMSE is very close to the declared overall accuracy of the ALS data (RMSE=0.23 m).

However, the interpolation artifacts were clearly visible in the parameterisation using npmin=300. Almost complete elimination of interpolation artifacts can be seen using npmin=400 and further reduction of point using dmin=6 meters while still preserving the mapped geomorphic features such as sinkholes (dolines).

The lowest RMSE was achieved using the suggested approach of selective data points reduction (RMSE=0.2195 m). Another benefit of the proposed method is also in the substantial increase in the speed of computation (Table 1).

We demonstrated that the interpolation artifacts can be minimized by controlling the number of data points used in the interpolation process based on the minimal distances between the points as well as the selection of the most important points derived from the analysis of interpolation errors.

References

Hofierka J (2005) Interpolation of Radioactivity Data Using Regularized Spline with Tension. Applied GIS 1: 16-01 - 16-13. DOI: 10.2104/ag050016

Hofierka J, Cebecauer T (2007) Spatial Interpolation of Elevation Data With Variable Density: A New Methodology to Derive Quality DEMs. IEEE Geoscience and Remote Sensing Letters 4: 117–121.

Hofierka J, Cebecauer T, Šúri M (2007) Optimisation of Interpolation Parameters Using a Cross-validation. In: Peckham R.J., Jordan G. (eds.) Digital Terrain Modelling, Development and Applications in a Policy Support Environment, Se-ries: Lecture Notes in Geoinformation and Cartography, Springer: 67-82

Hofierka J, Parajka J, Mitasova H, Mitas L (2002) Multivariate Interpolation of Precipitation Using Regularized Spline with Tension. Transactions in GIS 6: 135-150.

Hutchinson M F (1998) Interpolation of rainfall data with thin plate smoothing splines: I, Two dimensional smoothing of data with short range correlation. Journal of Geographic Information and Decision Analysis 2: 152-167

Liu X (2008) Airborne LiDAR for DEM generation: some critical issues. Progress in Physical Geography 32: 31-49.

Mitas L, Mitasova H (1999) Spatial Interpolation. In: P.Longley, M.F. Goodchild, D.J. Maguire, D.W.Rhind (Eds.), Geographical Information Systems: Principles, Techniques, Management and Applications, GeoInformation International, Wiley: 481-492.

Mitasova H, Mitas L (1993) Interpolation by Regularized Spline with Tension: I. Theory and implementation. Mathematical Geology 25: 641-655.

Mitasova H, Mitas L, Brown B M, Gerdes D P, Kosinovsky I (1995) Modeling spa-tially and temporally distributed phenomena: New methods and tools for GRASS GIS. International Journal of GIS, 9 (4), special issue on integration of Environmental modeling and GIS: 443-446.

Neteler M, Mitasova H (2008) Open Source GIS: A GRASS GIS Approach. Springer Verlag, New York.

Wehr A, Lohr U (1999) Airborne laser scanning - an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing 54: 68-82.

Weng Q (2006) An evaluation of spatial interpolation accuracy of elevation data, In Riedl, A., Kainz, W. and G.A. Elmes (eds.): Progress in Spatial Data Handling. Springer-Verlag, Berlin: 805-824.

Acknowledgements

European Union
European Regional Development Fund
Investing in your future

OPVaV-2008/2.1/01-SORO ITMS:26220120007, VVGS 63/12-13, VVGS-PF-2012-62 UPJŠ, VEGA 1/1251/12, VEGA 1/0272/12.