Spatial Interpolation of Airborne Laser Scanning Data with Variable Data Density

Jaroslav Hofierka – Michal Gallay – Ján Kanuk

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

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 covered with occasional meadows and woods comprising a plateau dissected by a deep canyon and a few occasional small hills. The altitude ranges between 540 to 704 meters a.s.l.

The data were acquired in leaf-on conditions in 2003 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 217,894 points classified as bare ground of spacing varying between 1.6 to 60 m with the average data density around 0.054 point/m².

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

Regularized Spline with Tension

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 smooth). The minimum standard 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 database.

Evaluating the interpolation quality

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 parameterisations 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 were achieved by the setting with dm1=6 m (RMSE=0.1432 m). However, Table 2 shows that these best results were achieved for dm1=0 m with additional points with the highest interpolation errors identified during the computation with dm1=2 m. In contrast, the best result identified in Table 1 is the worst in Table 2. The overall interpolation results of the RST method and s-RBF 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 dm1=6 m while still preserving the mapped geometric features such as sinkholes (doline). 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.

Orthophotomaps 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

References


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 dm1=6 m while still preserving the mapped geometric features such as sinkholes (doline). 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.