DEM uncertainty modelling using spatial data mining

Felix Hebeler[1], Thorsten Behrens[2]

[1]Dept. of Geography / University of Zurich / Switzerland
felix.hebeler@geo.uzh.ch

[2]Institute of Geography / University of Tübingen / Germany
thorsten.behrens@uni-tuebingen.de

DEM UNCERTAINTY MODELLING

As all models are an approximation of reality, they are by definition subject to uncertainty. In order to test the reliability and robustness of a particular model, uncertainty associated with this model has to be assessed and simulated to quantify its impact on results (Plewe, 2000).

The existence of uncertainty in topographic models that use digital elevation models (DEM) has long been recognised. Consequently, a number of approaches have been devised to model DEM data uncertainty, which have been applied to test its influence on, among others, the calculation of aspect and slope (Oksanen & Sarjakoski, 2005), viewsheds (Fisher, 1998), or the topographic index (Wechsler, 2007).

Where higher accuracy reference data is available, error can be deduced and information on the spatial autocorrelation as well as the dependency on topographic parameters can be included in the uncertainty model (e.g. Kyriakidis et al., 1999). Where no reference data is available and uncertainty modelling is solely based on global DEM accuracy measures such as RMSE, assumptions have to be made about the statistical and spatial distribution of uncertainty, that are often disputable (e.g. Oksanen & Sarjakoski, 2005; Fisher, 1999). Modelling uncertainty originating from DEM accuracy and data processing, Hebeler and Purves (2004) implicitly include dependency of DEM uncertainty on topographic features, while spatial and statistical distribution of uncertainty remains assumption based.

In order to overcome this limitation, Hebeler and Purves (2008) analysed error of the Global Land One-km Base Elevation (GLOBE) DEM for areas where higher accuracy reference data was available. Using regression modelling, Hebeler and Purves (2008) then used these dependencies to model GLOBE DEM uncertainty for areas without reference data.

This approach delivers physically viable, realistic topographic surfaces and is suitable as input for Monte Carlo Simulations, as it contains stochastic elements. However, a number of shortcomings include the laborious transfer of the approach to different regions and data sources, and the strong dependency on the selection of topographic parameters used within the regression.

THE DATA MINING APPROACH

In this paper, we propose an uncertainty modelling framework based on a spatial data mining approach, which has several advantages over the simple regression model approach by Hebeler and Purves (2008).

The data mining approach applied was originally developed as a terrain based spatial data mining and pattern recognition framework for Digital Soil Mapping (Behrens et al., 2008). To predict errors on the basis of the GLOBE dataset, we use Random Forests (Breiman, 2001), a powerful ensemble prediction method based on multiple randomized regression trees.

Typically, a range of terrain attributes are derived and machine learning approaches applied to derive a regression between these terrain attributes with the derived error surfaces. In contrast, we do not use terrain attributes, but simply the difference in elevation of the center pixel to each point within a local neighborhood. Thus, we shortcut the process of terrain analysis and rely on the machine-learning algorithm to extract relevant information directly from the differences in elevation. As there is no necessity to choose a set of terrain attributes the method is very flexible, non-linear and can include potentially unknown surface functions. Additionally, large trends and subregions in the data can be detected and accounted for in the prognosis. The moving window size is derived by stepwise optimization, by minimizing the RMSE between the the prognosis and the training data.

For first experiments, SRTM3 research grade DEM data was used as ground truth to derive GLOBE DEM error (Jarvis et al., 2004) for southern Scandinavia (Fig. 1C). Applying the data mining approach delivered much higher correlations of the derived error (Fig. 1B) with the underlying GLOBE DEM (R^2>0.7) than those achieved using simple and compound topographic indices within the regression approach (R^2=0.45). Consequently, the explained variation of error is improved (>70%) and residual error (noise) is minimised. The predicted error (Fig. 1A) captures the spatial autocorrelation as well as the dependencies on the underlying terrain of the DEM error (Fig. 1B) well.

The approach presented thus provides a method for modelling uncertainy in lower accuracy DEMs such as GLOBE where reference data is only available for regions outside the area of interest. At the same time, by using the prognosis model error in combination with accuracy information from the reference data, spatially explicit uncertainty information is available for DEMs which are improved using the presented method, which allows for the testing of associated models using Monte Carlo Simulations.

When the presented data mining model is combined with an automated selection of representative training areas, it promises to be easily applicable for varying data of different type and location, e.g. GLOBE, SRTM or LIDaR.

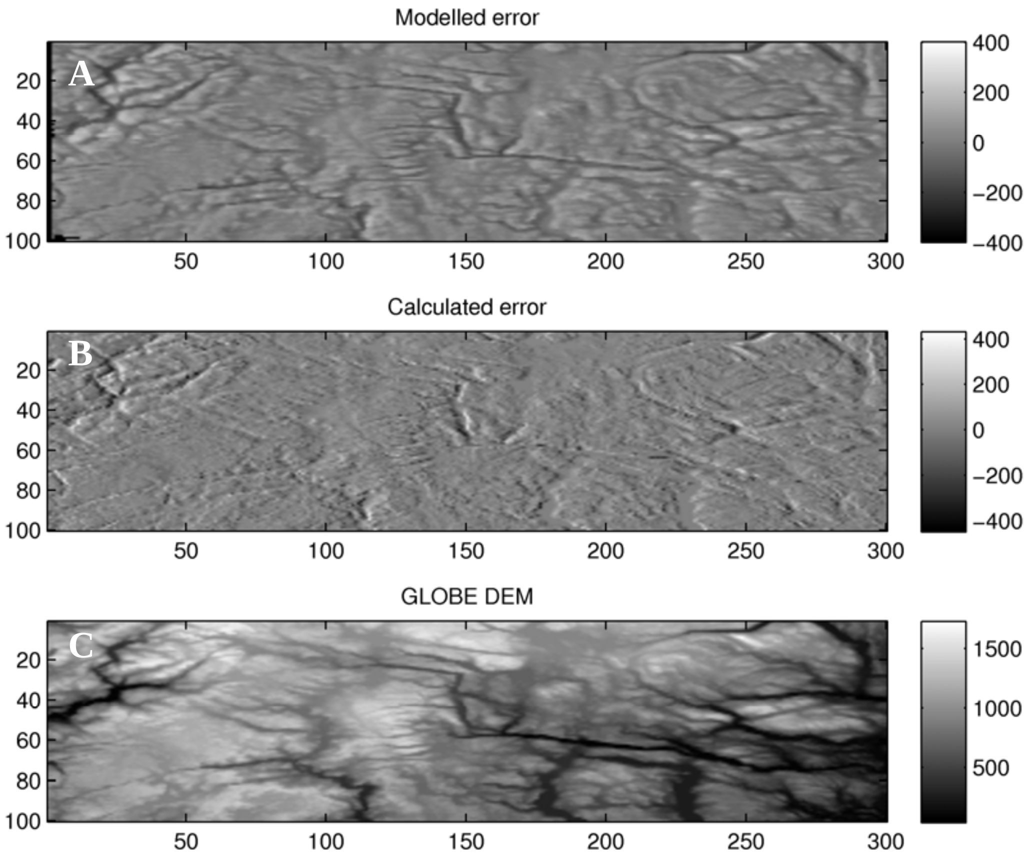

Fig. 1: Error modelled using the data mining approach (A), GLOBE DEM error derived using SRTM (B), and associated GLOBE DEM (C). The data mining approach (A) reproduces the spatial structure of the error (B) well. Effective range of modelled error is (-220 to 210m). All values in meters [m].

REFERENCES

Behrens, T., K. Schmidt, A.X. Zhu and T. Scholten (2008). A contextual terrain analysis approach for spatial pattern recognition and surface function approximation in soil landscape modelling. (submitted)

Breiman, L. 2001. Random Forests. Machine Learing 45, p.5-32

Fisher, P.F. (1998). Improved Modeling of Elevation Error with Geostatistics. GeoInformatica 2:3, 215-233.

Hebeler, F. and R.S. Purves (2008). Modelling DEM data uncertainties for Monte Carlo Simulations of Ice Sheet Models. In: A. Stein, J. Shi & W. Bijker: Quality Aspects in Spatial Data Mining, CRC Press, Boca Raton, p.175-196.

Hebeler, F. and R.S. Purves (2004). Representation of topography and its role in uncertainty: a case study in ice sheet modelling. Proceedings of the GIScience 2004 conference, College Park, Maryland, p.118-121.

Jarvis, A., J. Rubiano, A. Nelson, A. Farrow and M. Mulligan (2004). Practical use of SRTM data in the tropics: Comparisons with digital elevation models generated from cartographic data. Working Document No. 198. International Centre for Tropical Agriculture (CIAT), Cali, Columbia

Kyriakidis, P.C., A. Shortridge and M. Goodchild (1999). Geostatistics for conflation and accuracy assessment of digital elevation models. International Journal of Geographical Information Science 13, p.677-707.

Oksanen, J. & T. Sarjakoski (2005). Error propagation of DEM-based surface derivatives. Computers & Geoscience 31, 1015-1027.

Plewe, B. (2002). The Nature of Uncertainty in Historical Geographic Information. Transactions in GIS 6 (4), p.431-456.

Wechsler, S.P. (2007). Uncertainties associated with digital elevation models for hydrologic applications: a review. Hydrology and Earth System Sciences 11, p.1481-1500.