Error Propagation Modeling for Terrain Analysis using Dynamic Simulation Tools in ArcGIS Modelbuilder

Paul A. Zandbergen
University of New Mexico
Department of Geography
Albuquerque, NM 87131 USA
zandberg@unm.edu

Abstract—**Error propagation modeling for terrain analysis can provide insights into the robustness of terrain derivatives. For unconstrained terrain derivatives, such as stream and watershed delineation, the most widely used technique for error propagation modeling employs a Monte Carlo simulation of a spatially autocorrelated error model. The current study seeks to make this methodology more accessible to a wider audience by developing an error propagation modeling for terrain analysis using the ArcGIS Modelbuilder environment. Results of the model are illustrated using stream delineation of a high resolution lidar-derived DEM.**

BACKGROUND

Digital Elevation Models (DEMs) contain errors which may have an adverse effect on the quality of the derivatives obtained from these datasets. In many analyses terrain datasets are used as error-free models of reality, even though the existence of elevation uncertainty is widely recognized [1-2]. The effects of errors in a DEM for terrain analysis can be investigated using analytical or numerical error propagation techniques. Numerical techniques usually employ a Monte Carlo simulation. Monte Carlo simulation of random errors in DEMs has been applied to feature extraction [3], flow-path direction [4], automatic drainage basin delineation [5], route optimization [6], and a number of other terrain derivatives, such as roughness, flow accumulation, curvature, and slope failure [7].

The effect of DEM accuracy is not well understood for all terrain derivatives, in particular the more complex ones. Calculating terrain derivatives is a procedure in which new variables describing the properties of the surface are computed from a DEM. These derivatives are commonly divided into primary topographic attributes (e.g., slope, aspect, curvature, and catchment area) and secondary topographic attributes (e.g., topographic wetness index and stream power index). Primary topographic attributes are calculated directly from the elevation data or from one of its derivatives, while secondary topographic attributes are calculated from two or more primary ones. While this distinction is useful, from the perspective of understanding the effect of DEM accuracy, a more useful classification of terrain derivatives is based on their spatial properties rather than their source of calculation. Derivatives based on a fixed neighborhood can be considered as *constrained*, while derivatives that are based on far-reaching spatial interactions can be considered as *unconstrained*. Derivatives such as slope and aspect would be considered constrained, while derivatives such as catchment area and the presence of depressions would be considered unconstrained. The behavior of constrained derivatives is fairly predictable inasmuch as they are commonly determined by analyzing a 3 × 3 cell window around the cell for which the derivative is calculated. This behavior can to some degree be described analytically. For unconstrained derivatives the behavior is much less predictable because it may vary across multiple scales. This behavior requires empirical characterization. Due to their complexity, the body of research on the effect of DEM accuracy on unconstrained terrain derivatives is not very extensive, but does include a characterization of watershed boundaries [5,8], stream networks [9-14] and depressions [15-16].

One recent study [5] determined that for unconstrained derivatives a numerical error propagation technique using Monte Carlo simulation is more appropriate than an analytical technique. The basis for a numerical error propagation technique using Monte Carlo simulation is that the original elevation data of the DEM is modified repeatedly by a modeled error and the analysis of the terrain derivatives is calculated from the modified data set. Statistical summaries are drawn from the stack of analysis results based on the modified data set. The number of repetitions or realizations is set either very high (e.g. 1,000) or is based on some type of convergence threshold. The modeled error in numerical error propagation is usually a random error based on the expected standard deviation of the vertical error in the DEM.

The error is modified using either an exponential [e.g. 7] or Gaussian [e.g. 17] spatial autocorrelation model. The difference between the two models as applied to DEMs has been found to be very small and the range (or window size) of the spatial autocorrelation model is of greater influence [5]. The range essentially defines the spatial extent of the spatial autocorrelation.

While several examples of error propagation modeling for terrain analysis have been developed, the topic remains somewhat out of reach from most DEM users. This is in part because error propagation models are mostly developed in research settings using custom software (for example, written in Java, C++ or Matlab). Commercial geospatial software platforms typically do not incorporate tools to develop error propagation models. Recent advances in the geoprocessing functions in the widely used ArcGIS software package have made it possible, however, to develop robust error propagation models that can be used and modified with limited programming or scripting skills.

The current study demonstrates the capabilities of the ArcGIS Modelbuilder environment to develop error propagation modeling for terrain analysis. The underlying motivation for this is to make error propagation modeling more accessible to a wider audience, including geospatial professionals and students. The Modelbuilder environment provides a very intuitive environment to demonstrate error propagation concepts. Most intermediate level users of the geoprocessing framework in ArcGIS are also already familiar with the Modelbuilder environment. And while Modelbuilder is part of a commercial GIS software package and does not offer some of the same advantages of open-source platforms, models can be shared with a broader user community.

ERROR PROPAGATION MODEL IN ARCGIS

ArcGIS Modelbuilder was used to create an error propagation model for terrain analysis. Stream delineation is used to illustrate the use of the model. The example employs a lidar-derived DEM of 20-feet resolution, but the model can be used to run on any DEM.

The first part of the model is the creation of a spatially autocorrelated error term to be added to the original DEM, shown in Figure 1. Starting at the beginning of the model a normal raster is created – this is a raster with random cell values which follow a normal distribution. Cell values will be both positive and negative, with a mean of zero and a standard deviation of 1. This normal raster is the starting point for the error model. However, errors in DEMs are typically not random, but reveal strong signs of spatial autocorrelation. Spatial autocorrelation is introduced with the Focal Statistics tool. Within a circular neighborhood of 5 cells the random error is averaged – the nature of this neighborhood can be modified based on the characteristics of the spatial autocorrelation. Determining these characteristics, however, is not addressed in the model itself.

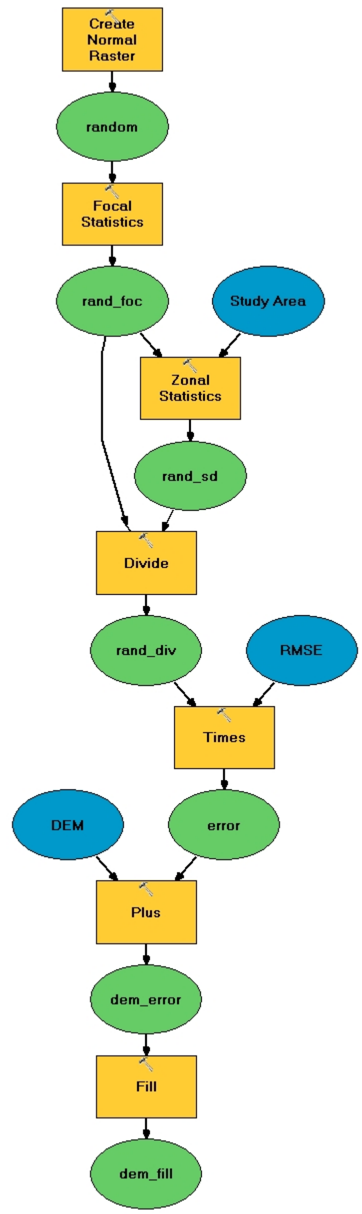

Figure 1. First part of the error propagation model to create a spatially autocorrelated error term for the input DEM.

Because of the averaging effect of focal statistics, the resulting spatially autocorrelated error raster has a much lower standard deviation than 1. The Zonal Statistics tool determines the standard deviation of the error raster and the Divide tool produces a new error raster with a standard deviation of 1.

The new error raster is multiplied with the user-defined Root Mean Square Error (RMSE) in vertical map units of the DEM. The RMSE value can be altered to simulate the effect of small and large errors. The final error raster is therefore a spatially autocorrelated error with a mean of zero and a standard deviation equal to the user-defined RMSE value. This error term is then added to the original DEM; this new DEM becomes the input into the stream delineation model.

The second part of the model, shown in Figure 2, creates the stream network: sinks are filled, flow direction and flow accumulation are derived from the DEM and streams are delineated based on a constant flow accumulation threshold.

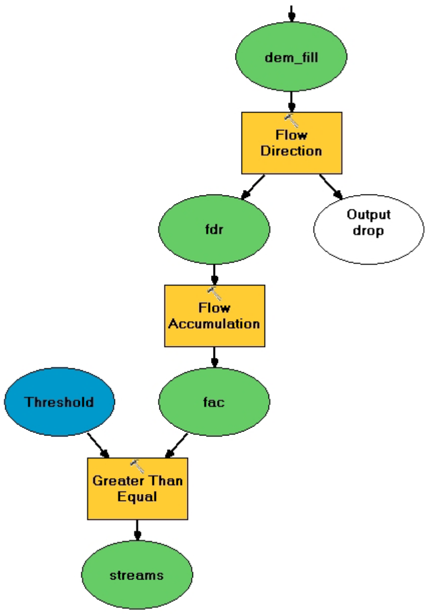

Figure 2. Second part of the error propagation model to create a stream network from the DEM.

The third and final part of the model, shown in Figure 3, stores the results of the model iteration. The stream network is added to the original Study Zero raster, with values of zero. This raster is copied and from the results a feedback loop is created. If the model runs multiple times (using iteration), the streams from the first run will be stored, and during the second run the second stream result will be added to the first. This makes it possible to store the results of multiple iterations.

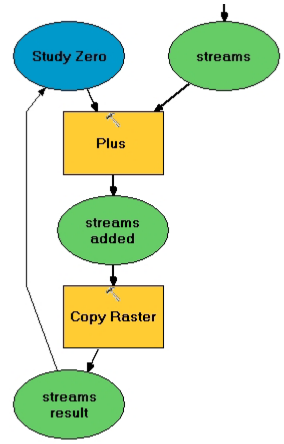

Figure 3. Third part of the error propagation model to store the results of model iterations.

The final output from the model is a raster where the cell values represent the number of times a stream is modeled as a stream. The number of model iterations can be set by the user and every model iteration employs a unique error term. The final output will therefore have value between a minimum of zero (never a stream) to a maximum of the number of iterations (always a stream. Figure 4 shows an example result of the model using 100 iterations with an RMSE value of 0.5 feet.

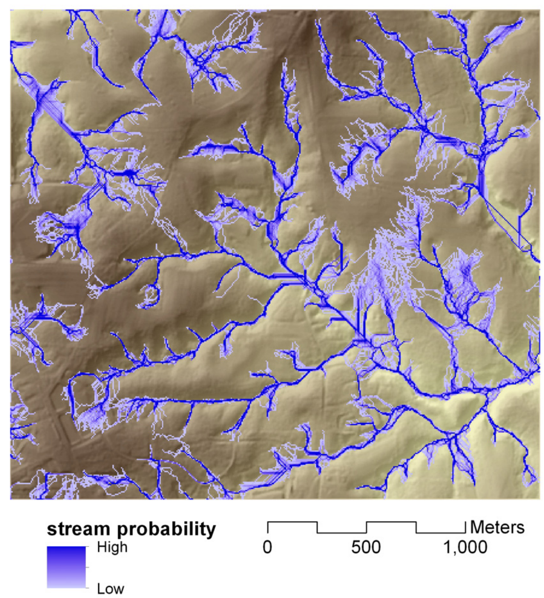

Figure 4. Results of error propagation analysis for stream delineation for RMSE of 0.5 feet and 100 model iterations.

The model result provides an indication of the uncertainty in the stream delineation. When running regular stream delineation without any error, the results do not come with any measure of the confidence in the location of the streams. The error modeling identifies areas where there is substantial uncertainty over the exact location of the streams and where additional information sources may be required to confirm the accuracy of the result.

The true error in this particular DEM is approximately 0.5 feet. This is quite small and the overall accuracy of this DEM can be qualified as very good. The results suggest most of the stream channels are quite well defined with little uncertainty over their location. The greatest uncertainty can be found in the channel heads (i.e. where 1st order streams start). There are also a few areas of very low relief where uncertainty is quite substantial.

Discussion and Conclusions

The error propagation model is ArcGIS Modelbuilder has a number of advantages. First, it presents a very intuitive and visual interface to developing and using a model, which makes it easier to follow for those who are new to this type of modeling. Very limited scripting or programming experience is required. Second, the model is fully extendible – specific parameters can be modified by the user and specific model elements can be replaced or updated. For example, modifying the spatially autocorrelated error model is relatively simple for those with Modelbuilder skills. Finally, the model can easily be shared with the broader ArcGIS user community. The model itself can be copied and any of its elements can be modified and run within ArcGIS.

There are a number of limitations as well. First, the approach to develop the spatially autocorrelated is relatively simple and does not incorporate more robust geostatistical simulations as employed by [9]. Second, some more advanced improvements cannot be accomplished using Modelbuilder alone but do in fact require Python scripting. While not very complicated, this does require additional skills. Third, the Monte Carlo simulation using a large number of iterations requires a lot of processing time. Running the analysis on large DEMs can therefore take a prohibitively long time in the order of days, not hours. The depression filling algorithm in ArcGIS is also limited in terms of the size of DEMs it can process.

Finally, similar error propagation tools for terrain analysis exist using open source tools, for example as documented by [9]. However, the Modlebuilder tool presented here make these tools available to those less familiar with scripting. It also demonstrates some of the capabilities of Modelbuilder with respect to dynamic simulation modeling. This presents new opportunities for making error propagation modeling available to a wider audience.

References

[1] Kumler, M.P., 1994, “An Intensive Comparison of Triangulated Irregular Networks (TINs) and Digital Elevation Models (DEMs),” Cartographica, 31(2):1–99.

[2] Carrara, A., Bitelli, G., and R. Carla, 1997, “Comparison of Techniques for Generating Digital Terrain Models from Contour Lines,” International Journal of Geographic Information Systems 11(5):451–473.

[3] Lee, J., Snyder, P. K., and P. F. Fisher, 1992, “Modeling the Effect of Data Error on Feature Extraction from Digital Elevation Models, Photogrammetric Engineering and Remote Sensing, 58(10):1461–1467.

[4] Veregin, H. V., 1997, “The Effects of Vertical Error in Digital Elevation Models on the Determination of Flow-Path Direction,” Cartography and Geographic Information Systems, 24 (2):67–79.

[5] Oksanen, J. and T. Sarjakoski, 2005, “Error Propagation Analysis of DEM-Based Drainage Basin Delineation,” International Journal of Remote Sensing, 26(14):3085–3102.

[6] Ehlschlaeger, C.R., 1998, The Stochastic Simulation Approach: Tools for Representing Spatial Application Uncertainty, Ph.D. dissertation, University of California, Santa Barbara.

[7] Holmes, K.W., Chadwick, O. A., and P.C. Kyriakidis, 2000, “Error in USGS 30-Meter Digital Elevation Model and its Impact on Terrain Modeling,” Journal of Hydrology, 233:154–173.

[8] Hancock, G.R., 2005, “The Use of DEMs in the Identification and Characterization of Catchment over Different Grid Scales,” Hydrological Processes, 19:1727–1749.

[9] Hengl, T., Heuvelink, G.B.M., and E.E. van Loon. 2010.”On the uncertainty of stream networks derived from elevation data: the error propagation approach,” Hydrol. Earth Syst. Sci. Discuss., 7, 767–799.

[10] Wang, X. and Z. Yin, 1998, “A Comparison of Drainage Networks Derived from Digital Elevation Models at Two Scales,” Journal of Hydrology, 210:221–241.

[11] McMaster, K. J., 2002, “Effects of Digital Elevation Model Resolution on Derived Stream Network Positions,” Water Resources Research, 38(4):13-1–13-8.

[12] Clarke, S. and K. Burnett, 2003, “Comparison of Digital Elevation Models for Aquatic Data Development,” Photogrammetric Engineering and Remote Sensing, 69(12):1367–1375.

[13] Lindsay, J. B., 2006, “Sensitivity of Channel Mapping Techniques to Uncertainty in Digital Elevation Data,” International Journal of Geographical Information Science, 20(6):669–692.

[14] Lindsay, J. B. and M. G. Evans, 2008, “The Influence of Elevation Error on the Morphometrics of Channel Networks Extracted from DEMs and the Implications for Hydrological Modeling,” Hydrological Processes, 22:1588–1603.

[15] Lindsay, J.B. and I.F. Creed, 2005b, “Sensitivity of Digital Landscapes to Artifact Depressions in Remotely-Sensed DEMs,” Photogrammetric Engineering and Remote Sensing, 71(9):1029–1036.

[16] Zandbergen, P.A. 2010. “Accuracy considerations in the analysis of depressions in medium resolution lidar DEMs,” GIScience and Remote Sensing, 47(2): 187-207.

[17] Goovaerts, P., 1997, Geostatistics for Natural Resources Evaluation, New York, NY: Oxford University Press.