A bottom-up strategy for uncertainty quantification in complex geo-computational models

Auroop R Ganguly*, Vladimir Protopopescu, Alexandre Sorokine†**

*† Computational Sciences & Engineering ** Computer Science & Mathematics
Oak Ridge National Laboratory
1 Bethel Valley Road, Oak Ridge, TN 37831

* Corresponding Author: Phone: +1-865-241-1305; Email: gangulyar@ornl.gov

1. Introduction

The reliable quantification of uncertainty in complex geo-computational models is an outstanding theoretical and practical problem within geographical information systems (Chiles and Delfiner, 1999; Lowell and Jaton, 1999; Zhang and Goodchild, 2002; Couclelis, 2003; Gertner et al., 2004; Kardos et al., 2005). A commonly used approach for uncertainty quantification in relatively simple geospatial models is due to Openshaw (1989). This approach relies on estimation techniques of errors caused by individual operations and has been used successfully in the context of map algebra operations (Heuvelink, 1998). Recent advances in geospatial uncertainty estimation include new approaches in spatial statistics (Cressie, 1993; Ripley, 2004), refinement Kalman filtering techniques for geospatial data (Wikle and Cressie, 1999), as well as methods for spatial data mining (Ester et al., 2000; Shekhar et al., 2001; Miller and Han, 2001; Shi and Wang, 2002; Hanning and Shuliang, 2004). In the context of specific domains, geo-scientists have developed special methods for spatial and spatio-temporal uncertainty estimation (e.g., Ganguly, 2002; Ganguly and Bras, 2003). However, these approaches have not been applied to, nor have they been designed for, large-scale and complex geo-computational models.

2. Approach

This paper addresses an important gap in the geospatial sciences by proposing a new and generic framework for uncertainty quantification for complex geo-computational models. Within this framework, we propose *a "bottom-up strategy" whereby the uncertainties from the individual operations and the input-dependent uncertainties are combined to yield estimates of the overall uncertainty in a complex geo-computational model*. This strategy would allow us to develop a systematic approach towards classification of geospatial operations and uncertainty estimation techniques for each geospatial operation or an entire class of such operations.

Geospatial uncertainty can be expressed in various ways, and can be expected to be a function of auto- and cross-correlations in space or time; spatial, temporal or spatio-temporal outliers; spatial and spatio-temporal stationarity; as well as spatial, temporal and attribute error structures. Our approach to problem decomposition and input-dependent or component-based uncertainty relies on recent developments in time series analysis and complex systems. In our framework, complex, hierarchical geospatial models are viewed as being comprised of chains of simple operations at or between various levels of detail or granularity (Fig. 1). Each such operation transforms not only the data themselves *but also the uncertainty associated with the data*. As a result, even under stationary assumptions, uncertainty "propagates" through complex geospatial models. In the proposed framework uncertainty can be estimated as it "propagates" through a chain of geospatial operations (Fig. 2).

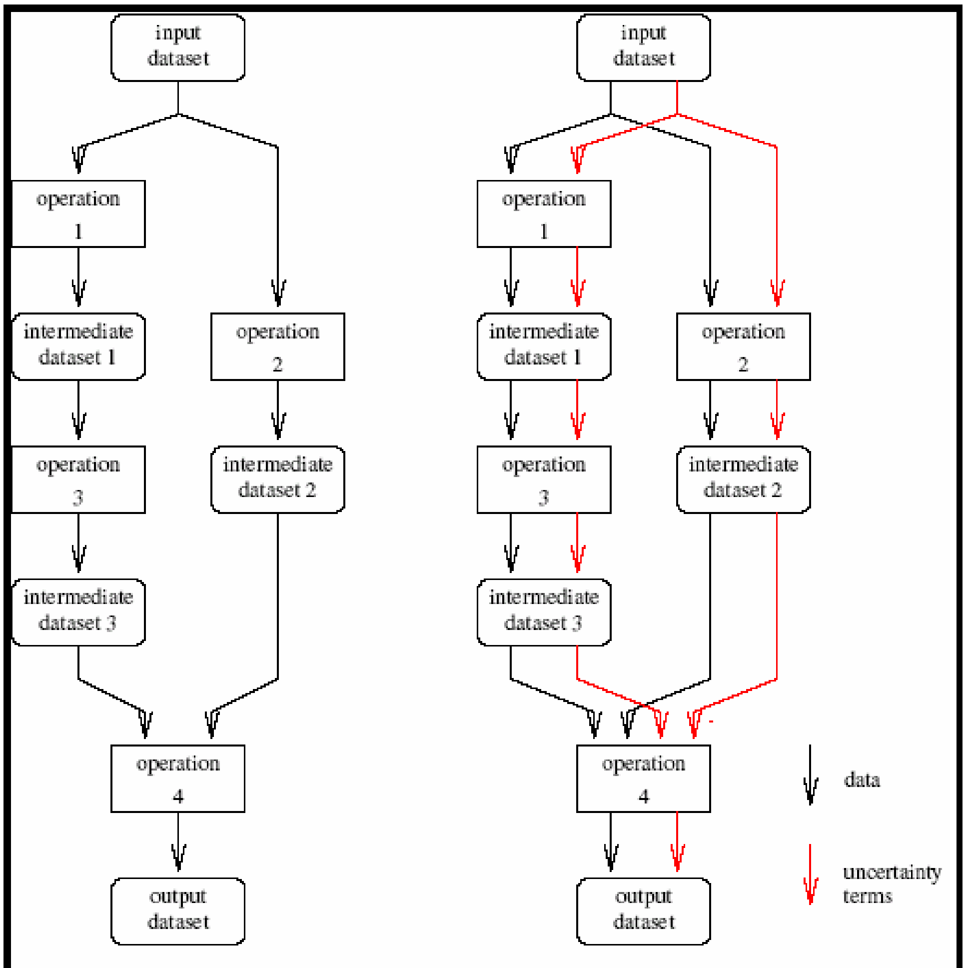

Fig. 1 Fig. 2

The proposed approach ultimately requires uncertainty estimation methodologies for each and every geospatial operation. However, while the number of such operations may be rather large (e.g., GRASS GIS 6.0 includes about three hundred different data processing and analytical commands: Neteler and Mitasova, 2004), the majority of GIS operations is reduced to a relatively small number of universal operations (e.g., Albrecht, 1996, lists twenty such operations). Our goal is to develop accurate and robust methods for uncertainty estimation for these universal operations.

The methods for uncertainty quantification utilized in this paper are capable of considering the uncertainty in the input data (or input dependent uncertainty in space and time) as well as the uncertainty caused by the individual geospatial operations and their concatenations (or geospatial model/operation dependent uncertainty). Specifically, the methods in this paper are based on a couple of broad techniques: (i) spatially-weighted linear regression formulations (Cressie, 1993), and (ii) extensions and further development of the "Ensemble Neural Network" (ENN) formulation developed by Ganguly and Bras (2003) and Ganguly (2002), which in turn was developed based on the Bayesian Neural Network approaches proposed by MacKay (1995) and the NARMA formulation of Connor et al. (1994). The regression-based and ENN methodologies are utilized to estimate the uncertainty resulting from individual geospatial operations as well as a chain of such operations. The paper also utilizes the concepts of input-dependent uncertainty, originally developed for time series data (e.g., the ARMA-ARCH and GARCH formulations: see Engle, 1982 and Mills, 1990), in the context of geospatial data and the spatially-weighted regression or the ENN formulations.

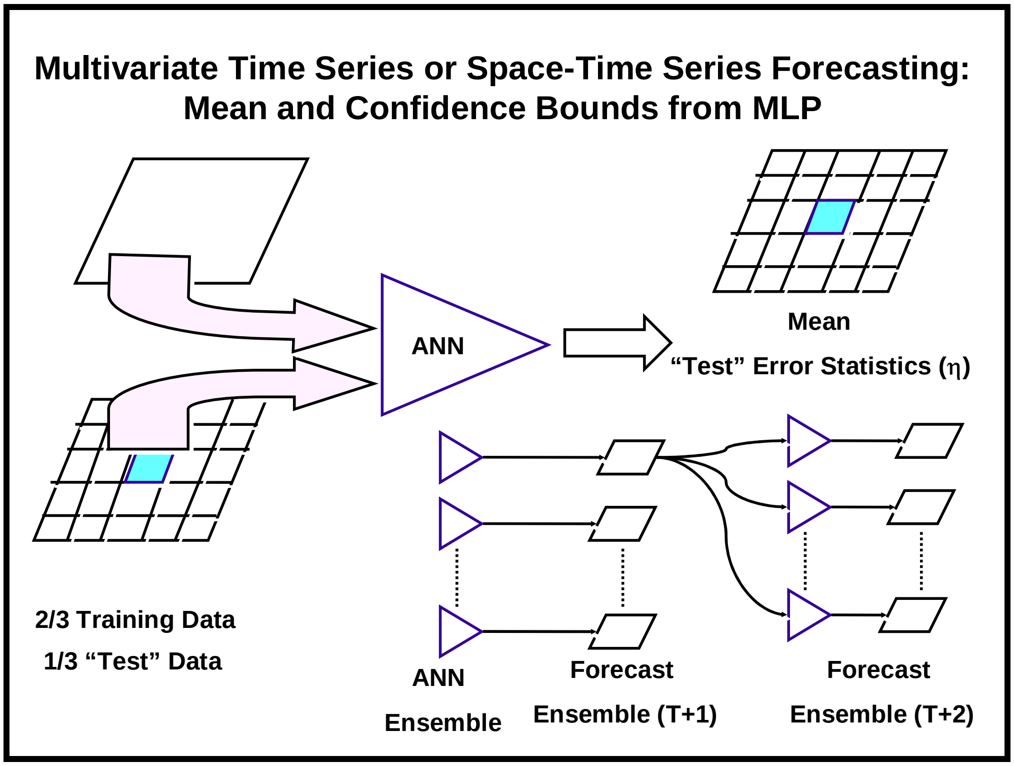

Fig. 3: Schematics of the ensemble neural network approach
(ANN: Artificial Neural Networks; MLP: Multi-Layer Perceptrons)

The new contribution of this paper is to bring together these techniques (e.g., the ENN formulation depicted in Fig. 3) into the "bottom-up strategy" schematically illustrated in Figs. 1 and 2. Specifically, this implies a consistent evaluation of both the input-dependent and the model- (or geospatial operation) dependent uncertainty within a common generic framework, in the context of complex geo-computational models.

4. Results

We illustrate the applicability of the proposed framework using two geospatial models comprised of a chain of relatively simple operations. The models simulate a representative set of commonly used geospatial operations and their combinations. Specifically, we utilize geospatial datasets available at ORNL in the areas of population distribution and meteorology. For population, we develop regression-based approaches for distributing population estimates available from low-resolution census counts or other estimates to high-resolution grid cells through the use of ancillary variables like nighttime lights and land-cover. The overall uncertainty estimates derived from the uncertainty in the constituent geospatial operations are validated by comparing with available high-resolution population datasets. For meteorology, we develop multi-scale relationships between observed variables and develop uncertainty formulations based on the constituent operations. Validation is performed using "held-out" data. Our results demonstrate the validity of our bottom-up strategy for uncertainty formulations.

References

Albrecht, J., 1996, Universal Analytical GIS Operations, A Task-Oriented Systematization of Data Structure-Independent GIS Functionality Leading Towards a Geographic Modeling Language. Dissertation, ISPA Mitteilungen – Heft 23, Vechta.

Couclelis, H., 2003, The Certainty of Uncertainty: GIS and the limits of geographic knowledge, Transactions in GIS, 7, 2, 165-175.

Chiles, J.-P., and P. Delfiner, 1999, Geostatistics: Modeling spatial uncertainty, Wiley.

Connor, J.T., Martin, R.D., and L.E. Atlas, L.E, 1994: Recurrent neural networks and robust time series prediction. IEEE Transactions on Neural Networks, 5(2), 240-254.

Cressie, N., 1993, Statistics for spatial data, revised edition, Wiley.

Engle, R.F.. 1982, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. Econometrica, 50, 987-1008.

Ester M., Frommelt A., Kriegel H.-P., and J. Sander, 2000, Spatial data mining: databases primitives, algorithms and efficient DBMS support, Data Mining and Knowledge Discovery, 4, 193-216.

Ganguly, A. R., 2002, A Hybrid Approach to Improving Rainfall Forecasts. Computing in Science and Engineering: Special Issue on Data Mining, IEEE Computer Society and American Institute of Physics, 4(4), 14-21.

Ganguly, A. R., and R. L. Bras, 2003, Distributed quantitative precipitation forecasting combining information from radar and numerical weather prediction model outputs. Journal of Hydrometeorology, American Meteorological Society, 4(6), 1168-1180.

Gertner, G. Z., Fang, S., Wang, G., and A. Anderson, 2004, Partitioning spatial model uncertainty based on joint spatial simulation, Transactions in GIS, 8, 4, 441-458.

Hanning, Y., and W. Shuliang, 2004, Factors causing uncertainty in spatial data mining, XXth ISPRS Congress, Istanbul, Turkey, July.

Heuvelink, G. B. M., 1998, Error propagation in environmental modelling with GIS, Taylor & Francis.

Kardos, J., Benwell, G., and A. Moore, 2005, The visualisation of uncertainty for spatially referenced census data using hierarchical tessellations, Transactions in GIS, 9, 1, 19-34.

Lowell, K., and A. Jaton (Eds.), 1999, Spatial accuracy assessment: Land information uncertainty in natural resources. Chelsea, MI: Ann Arbor Press.

MacKay, D. J. C., 1995, Probable networks and plausible predictions-a review of practical Bayesian methods for supervised neural networks. Network: Computational Neural Systems, 6, 469-505.

Miller H. J., and J. Han, 2001, Geographic data mining and knowledge discovery, Taylor and Francis.
Mills, T. C., 1990, Time Series Techniques for Economists, Cambridge.
Neteler, M., and H. Mitasova, 2004, Open Source GIS: A GRASS GIS Approach, Boston.
Openshaw, S., 1989, Learning to live with errors of spatial data transformations, in Michael F. Goodchild, Sucharita Gopal (Eds.), Accuracy of Spatial Databases
Ripley, B., 2004, Spatial statistics, John Wiley.
Shekhar, S., Huang, Y., Wu, W., Lu, C.T., and Chawla, S., 2001, What's spatial about spatial data mining: three case studies, In R.Grossman, C. Kamath, V. Kumar, and R. Namburu (eds), Data Mining for Scientific and Engineering Applications, Kluwer Academic Publishers.
Shi, W.Z., and S.L. Wang, 2002, Further development of theories and methods on attribute uncertainty in GIS, Journal of Remote Sensing, 6(4): 282-289.
Wikle, C.K., and N. Cressie, 1999, Dimension-reduced approach to space-time Kalman filtering, Biometrika, 86, 815-829.
Zhang, J.X., and M.F. Goodchild, 2002, Uncertainty in Geographical Information, Taylor and Francis.

Acknowledgments

We thank B. Bhaduri, P. Coleman, E. Bright and A. King of the Geographical Information Science and Technology group at the Oak Ridge National Laboratory for their interest and help with an earlier version of the paper.