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

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

1. State of the art

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 essentially relies on error estimation techniques 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 of existing approaches like Kalman filtering for geospatial data (Wikle and Cressie, 1999), as well as new approaches in spatial data mining (Ester et al., 2000; Shekhar et al., 2001; Miller and Han, 2001; Shi and Wang, 2002; Hanning and Shuliang, 2004). However, these approaches have not been applied to, and nor have they been designed for, large-scale and complex computational models. Applied mathematicians and computer scientists have developed methods for sensitivity analysis and uncertainty reduction from complex computer models (e.g., Bischof et al., 1996; Barhen et al., 2004), however these have not been developed for or applied to geospatial data. Geo-scientists have developed approaches for spatial and spatio-temporal uncertainty estimation (e.g., Ganguly, 2002; Ganguly and Bras, 2003), but these are applicable in the context of the specific domain.

2. Gaps in the literature and research motivation

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. The approach adopts a systematic approach towards classification of geospatial operations, develops uncertainty estimation techniques for each geospatial operation or an entire class of such operations, accounts for uncertainty in the input variables, and proposes a bottom-up strategy where the uncertainties from the individual operations and the inputs-dependent uncertainties are combined to yield estimates of the overall uncertainty in a complex geo-computational model. 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. The approaches for problem decomposition and input-dependent or component-based uncertainty relies on recent developments in disciplines like time series analysis and complex systems.

The proposed approach requires a decomposition of the complex model into the constituent geospatial operations, as well as an understanding of how the operations combine to yield the final model and the results. The eventual requirement is to develop generic mathematical frameworks for the propagation of uncertainty in complex geo-computational models through the constituent geospatial operations. Once available, uncertainty estimates from complex geo-computational models are likely to influence the interpretation and application of the model results, including automated techniques for model calibration. In addition, estimates of uncertainty will impact judgmental updates of model parameters and model outputs. Uncertainty estimation is a first-step for risk formulations at multiple resolutions, and eventually impacts the entire decision-making process.

3. New approach and research vision

In our framework, complex, hierarchical geospatial models are viewed as being comprised of chains of simple operations at or between various levels of granularity or detail (Fig. 1). Each such operation transforms not only the data itself 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 it will be possible to estimate uncertainty as it "propagates" through a chain of geospatial operations (Fig. 2).

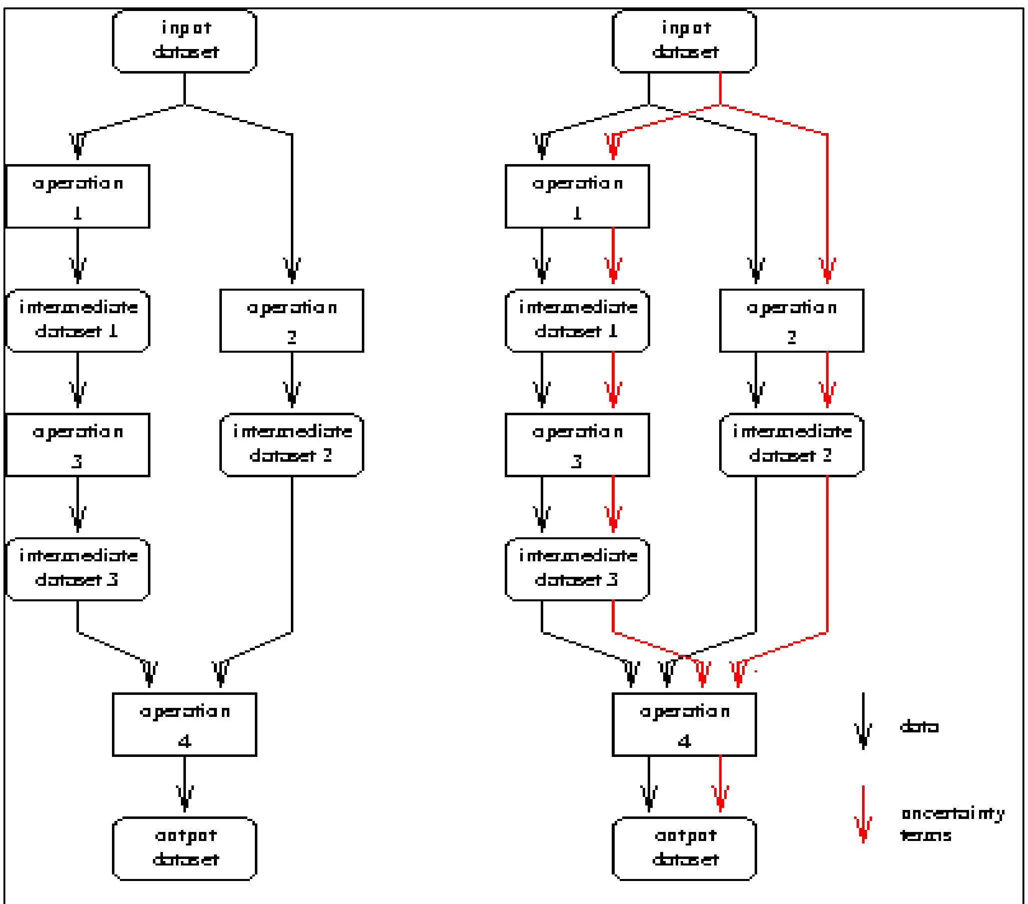

Fig. 1   Fig. 2

Our vision is to develop uncertainty estimation methods that are comprehensive enough to cover most individual geospatial operations and their interactions, and utilize these methods to develop generic framework for uncertainty quantification in complex geo-computational models. The complex uncertainty formulations may be analytically derivable as a function of the uncertainty of the individual operations in certain situations. In other situations, numerical solutions may be the only option. In specific cases, the solutions may have to be in the context of specific data sets, even though the overall approach needs to remain extensible and generic. Existing geographical information systems feature a very large number of geospatial operations and models. For example, GRASS GIS 6.0 includes about three hundred different data processing and analytical commands. However, as it was shown in (Albrecht, 1996), the majority of GIS operations can be reduced to a relatively small number of universal operations. Albrecht (1996) lists twenty such operations. Our ultimate goal is to develop methods for uncertainty estimation for all these universal operations. The vision described here comprises significant and challenging problems, and constitutes long-term research goals at the Oak Ridge National Laboratory (ORNL).

4. Research focus and results

We demonstrate the applicability of the proposed framework using a generalized geospatial model comprised of a chain of simple analytical operations of various types. The model construction simulates a representative set of commonly used geospatial operations and their combinations. The focus of this specific paper is on the operations that result in transformation of geospatial data at multiple scales or resolutions, for example, on operations like spatial and spatio-temporal aggregation and “atomization” or disaggregation processes. The formulation, estimation and propagation of uncertainty resulting from the combined effect of uncertainty in the input variables and the spatial, temporal or attribute properties (e.g., spatial or spatio-temporal correlations and dependence), as well as the uncertainty caused by the individual geospatial operations, are demonstrated. Test datasets were created using two major sources at ORNL: (1) real-world high-resolution population data and related ancillary variables and (2) global meteorological data obtained from observations and forecasts generated by numerical models of the weather. Established time series and domain-specific approaches are utilized in the formulations. Specifically, this paper extends and further develops ARMA-ARCH (Engle, 1982; Mills, 1990) type time series approaches for input dependent uncertainty in the context of geospatial data and extends the Bayesian Neural Network formulations (MacKay, 1995; Ganguly and Bras, 2003) for the estimation of the uncertainty resulting from a chain of simple geospatial operations.

4. 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

Barhen, J., Protopopescu, V., and D. B. Reister, 2004, Consistent uncertainty reduction in modeling nonlinear systems, SIAM Journal of Scientific Computing, 26, 2, 653-665.

Bischof, C., Khademi, P., Mauer, A., and A. Carle, 1996, Adifor 2.0: Automatic differentiation of Fortran 77 programs, IEEE Computing in Science and Engineering, 3, 18-32.

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.
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.
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 for their interest and help.