Available online at www.sciencedirect.com
ScienceDirect
Procedia Social and Behavioral Sciences 2 (2010) 7650–7651
Procedia
Social and Behavioral Sciences

Sixth International Conference on Sensitivity Analysis of Model Output

Identifying Important Observations Using Cross Validation and Computationally Frugal Sensitivity Analysis Methods

Laura Foglia[a], Mary C. Hill[b], Steffen W. Mehl[b], Paolo Perona[c], Paolo Burlando[c]

[a]Civil and Environmental Engineering Department, University of California Davis, One Shields Ave., 95616
[b]U.S. Geological Survey, 3215 Marine St., Boulder, CO, 80303, [c]IfU, Institute for Environmental Engineering, ETH Zurich

Abstract

Sensitivity analysis methods are used to identify measurements most likely to provide important information for model development and predictions. Methods range from computationally demanding Monte Carlo and cross-validation methods that require thousands to millions of model runs, to very computationally efficient linear methods able to account for interrelations between parameters that involve tens to hundreds of runs. Some argue that because linear methods neglect the effects of model nonlinearity, they are not worth considering. However, when faced with computationally demanding models needed to simulate, for example, climate change, the chance of obtaining insights with so few model runs is tempting. This work compares results for a nonlinear groundwater model using computationally demanding cross-validation and computationally efficient local sensitivity analysis methods.

Keywords: Local sensitivity analysis; linear and non linear methods; hydrological models; climate models

1. Comparison with cross-validation to evaluate the utility of linear methods

Sensitivity analysis, calibration, and uncertainty evaluation methods are critical to developing useful models of complex hydrologic systems for which important characteristics can not be measured accurately and(or) completely enough to define model input values fully (Saltelli et all, 2008; Hill and Tiedeman, 2007)). These methods allow the modeller to explore the relations between different types of data and the processes represented in the model, including the testing of hypotheses about system structure (alternative models).

Model nonlinearity and its consequences for model calibration and sensitivity analysis are evaluated by Kavetski1 and Kuczera (2007). Experience suggests that many models of natural systems are linear enough for local sensitivity analysis methods to be useful (see examples of groundwater flow and advective transport, conservative and reactive groundwater transport, and streamflow and transport cited by Foglia et al., 2009). This suggests that the concern expressed by Saltelli et al. [2008, p. 11] that local methods are inefficient in terms of the analyst's time is perhaps overstated, but clear comparison of nonlinear and linear methods are needed to better understand the opportunities and limitations of linear methods.

Local sensitivity analysis is based on first-order, second moment (FOSM) approximations. The linear statistics for which results are shown in this presentation are calculated using UCODE_2005 (Poeter et al., 2005). They include fit-independent and fit-dependent statistics (Hill and Tiedeman, 2007). The fit-independent statistics are

doi:10.1016/j.sbspro.2010.05.161

dimensionless scaled sensitivities (DSS), leverage, and observation-prediction (OPR). The fit-dependent statistics are DFBETAS and Cook's D. The role of parameter interdependence as measured by parameter correlation coefficients (PCC) is discussed (only DSS does not account for parameter interdependence).

The alternative to local sensitivity analysis commonly is global sensitivity analysis (Saltelli et al., 2008). Global methods most commonly used in hydrology are GLUE (Generalised Likelihood Uncertainty Estimation; Beven and Freer, 2001) and MCMC (Markov Chain Monte Carlo; Kuczera and Parent, 1998). Here, we compare linear methods with results obtained through cross-validation. In cross validation, all observations are used to produce a calibrated model and associated predictions. Then one or more observations are removed, the regression repeated, and resulting changes in parameter values and predictions are evaluated. Large changes in parameter values and predictions indicate the associated observations are important.

In this work, comparisons of local sensitivity analysis and cross-validation are conducted using a groundwater model of the Maggia Valley, Southern Switzerland; applicability to climate models is inferred. Results show that the frugal linear methods produced about 70% of the insight from about 2% of the model runs required by the computationally demanding methods. Linear methods were not always able to distinguish between moderately and unimportant observations. However, they consistently identified the most important observations. Importance both to estimate parameters and predictions of interest was readily identified.

The results suggest that it can be advantageous to consider local sensitivity analysis in model evaluation, possibly as a preliminary step to provide insights that can be used to improve the design of more demanding methods.

2. References

Beven, K.J., J. Freer, 2001, Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology, *Journal of Hydrology*, 249, 11-29.

Foglia, L., S.W. Mehl, M.C. Hill, P. Perona, and P. Burlando, 2007, Testing alternative ground water models using cross-validation and other methods, *Ground Water*, 45(5), pp. 627-641.

Foglia, L., S.W. Mehl, M.C. Hill, and P. Burlando, 2009, Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function, Water Resour. Res., 45, W06427, doi:10.1029/2008WR007255.

Hill, M.C. and C.R. Tiedeman, 2007, Effective calibration of groundwater models, with analysis of data, sensitivities, predictions, and uncertainty, John Wiley and Sons, New York.

Kavetski, D., and G. Kuczera, 2007, Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration, Water Resour. Res., 43, W03411, doi:10.1029/2006WR005195.

Kuczera, G. and E. Parent, 1998, Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis, *Journal of Hydrology*, 211, 69-85.

Poeter, E. P., M. C. Hill, E. R. Banta, S. W. Mehl, and S. Christensen, 2005, UCODE– 2005 and six other computer codes for universal sensitivity analysis, inverse modeling, and uncertainty evaluation, U.S. Geol. Surv. Tech. Methods, 6-A11.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisona, S. Tarantola, 2008, Global sensitivity analysis, The primer: Chichester, West Sussex, England, Wiley and Sons, 292p.

Torn, Ryan D.; Hakim, Gregory J., 2008,Ensemble-Based Sensitivity Analysis: Monthly Weather Review, vol. 136, issue 2, p. 663.