Comparison of three spatial sensitivity analysis techniques

Nathalie Saint-Geours
AgroParisTech, UMR TETIS,
F-34035, Montpellier,
France
nathalie.saint-geours@teledetection.fr

Linda Lilburne
Landcare Research
PO Box 40, Lincoln 7640, Canterbury
New Zealand
lilburnel@landcareresearch.co.nz

Abstract—**This paper compares the spatial Sobol' sensitivity analysis approach to two other sensitivity analysis techniques on a model with spatially distributed inputs. The comparison is performed on AquiferSim, a model that simulates groundwater flow and nitrate transport from paddock (i.e. field) to aquifer. Some of the input layers have considerable uncertainty. Alternative soil and land-use layers were simulated through Monte Carlo simulation based on expert-derived confusion matrices. Uncertainty of the raster rainfall layer was simulated via geostatistical unconditional simulation of error fields. The three sensitivity techniques are: (1) the spatial Sobol' technique, (2) one-at-a-time (OAT) variation around base sample points, and (3) the Elementary Effects method. The results show that the spatial Sobol' approach gives the best insight on AquiferSim behavior. OAT local variations of inputs around some sample points allow checking of the robustness of model predictions around those points, but give no insight on the relative importance of inputs. The Elementary Effects method shows that land use layer is the most influential input factor, but fails to capture interactions between input factors. The spatial Sobol' approach identifies the land use layer as being the most influential. It shows that strong interactions occur between most of the inputs, explaining 43% of the output variability.**

Keywords: sensitivity analysis, Sobol, Elementary Effects

I. INTRODUCTION

Sensitivity analysis (SA) techniques can be used to study how uncertainty in model inputs influences uncertainty in model predictions (Saltelli et al., 2000). Various techniques are available to perform sensitivity analysis from a set of model evaluations (Helton, 1993; Saltelli et al., 2004). They differ mainly in how the uncertainty of the model inputs is sampled, and in how the sensitivity measures are calculated. Some of these various techniques are suitable for models with spatially distributed inputs, others, like regression-based approaches, are not (Lilburne and Tarantola, 2009).

A spatial SA has been carried out on AquiferSim, a model that simulates groundwater flow and nitrate transport from paddock to aquifer. The AquiferSim model has been designed for analysis of cumulative effects of leaching from agricultural non-point sources at a regional scale on alluvial plains in New Zealand. Some of the inputs for AquiferSim are categorical vector data (soil, land use and climate zone layers); others are continuous raster data (aquifer transmissivity and recharge). These spatial inputs, along with other scalar inputs, are subject to various sources of uncertainty (measurement errors, interpolation errors, missing data, etc.). Sensitivity analysis (SA) is needed to study how these uncertainties influence the variability of AquiferSim predictions, to check the robustness of the model predictions, and to identify the factors that account for most of the model output variability. Three different spatial SA techniques were used: one-at-a-time (OAT) local variations around some base points, the Elementary Effects method, and the variance-based spatial Sobol' method. This paper compares the results of the three methods.

II. METHODS

A. *AquiferSim Model*

AquiferSim is a steady-state model of groundwater flow and contaminant transport. It has been designed for analysis of cumulative effects of leaching from agricultural non-point sources at a regional scale under a range of land use scenarios (Bidwell et al., 2005). AquiferSim takes various GIS layers as inputs, including maps of land use, soil type and climate zone (categorical vector data) and maps of river recharge and transmissivity of the aquifer (continuous raster data). AquiferSim models the three-dimensional concentration of nitrate in the groundwater. One output is vertical 1-D profile graphs of nitrate concentration below point locations chosen by the user.

B. *Target Function of Study*

AquiferSim was used to simulate a region of 3500 km^2 with a cell size of 100 m. The objective function Y for the sensitivity analysis was chosen as the average of the maximum nitrate concentrations (mg/L) below a set of 10 points distributed evenly around the study area.

C. *Simulating Uncertainty of the Model Inputs*

Most SA methods need the model to be evaluated on multiple points of the space of the input factors. Six input factors were considered in this analysis, the first is a lookup table of nitrate and drainage values, the other five are spatial inputs: land use, soil, climate, transmissivity of the aquifer, and river recharge maps. Uncertainty for each input factor has been simulated using a specific model (Table 1).

The nitrate and drainage lookup table is derived from outputs from a paddock-scale model. The table lists nitrate concentration and vertical drainage values per unit land area for every possible combination of land use type, soil type and rainfall class. In addition to the original table, two alternatives were generated by applying global –5% and +5% variations to the nitrate concentration values only, leaving the drainage values unchanged.

Soil and land use are categorical vector layers. Derived from a soil survey, and a farmer questionnaire respectively, they identify 7 key soil and 25 land-use types over the study area. Uncertainty was modeled using two expert-derived confusion matrices. Nine alternative soil layer realizations and nine alternative land-use layer realizations were generated through random simulation of the soil and land use uncertainty by applying probabilities derived from the confusion matrices.

The climate map comes from an interpolated layer of mean annual rainfall. Three rainfall layer realizations were simulated by a geostatistic method, adding a stochastic error field to the original data. This error field was generated using the circulant embedding technique (Chan and Wood, 1997), as a zero mean stationary Gaussian field with some spatial structure, described by a variance and a semi-variogram model.

Maps of river recharge and transmissivity of the aquifer come from calibrated runs of the Modflow groundwater model (Harbaugh et al., 2000) where hydraulic conductivity has been varied to match groundwater levels. Three water-budget scenarios, based on expert opinion of the ratio of river recharge to land-surface recharge and the proportion of ocean discharge, were simulated in Modflow, resulting in paired river recharge and transmissivity layers for each of the three scenarios.

D. *Multiple Evaluations of the AquiferSim Model*

One run of the AquiferSim model takes several minutes to complete. In order to run multiple evaluations of the model, a parallel computing framework has been designed in Python to use a cluster of 10 powerful nodes, thus reducing computational time considerably. In this configuration, a Monte Carlo type approach to sensitivity analysis is practical.

TABLE I. INPUT VARIABLES OF AQUIFERSIM MODEL

Variable	Model of Uncertainty
Soil map	9 alternative layers derived from a Monte Carlo simulation using a confusion matrix
Land-use map	9 alternative layers derived from a Monte Carlo simulation using a confusion matrix
Nirate/drainage table	3 alternative tables : original table, –5%, +5%
Climate map	3 alternative layers of rainfall zone from geostatistical simulation
River recharge map and aquifer transmissivity map	3 sets of linked pairs of maps – based on three water-budget scenarios with higher/lower proportions of ocean discharge

E. *Sensitivity Analysis*

Three different approaches have been considered to perform spatial sensitivity analysis of the AquiferSim model: OAT local variations around some base points, the Elementary Effects method (Morris, 1991; Saltelli et al., 2008) and the variance-based spatial Sobol' method (Lilburne and Tarantola, 2009).

The same principle was used to handle spatially distributed inputs in the three approaches: the i^{th} spatial input factor X_i is described as a discrete uniform distribution in $\{0 ; 1 ; .. ; n_i - 1\}$ where n_i represents the total number of realizations for the i^{th} input factor. Each discrete level is associated with a single realization of the spatial input, generated from a specific spatial uncertainty model. One must note that SA techniques based on regression and the Fourier-based techniques cannot be applied in this case, as the order in which spatial realizations are ranked has no meaning.

The three different approaches considered differ on various points, including design of the sampling of input factors, measures of sensitivity, and computational cost (e.g. number of model simulation runs). The first two are mainly local, quantifying the variation of the model output due to small variations in the uncertain model inputs around a base point. The spatial Sobol' method is global, exploring more of the multidimensional input space.

1) *OAT local variations around some base points*

A first rough approach is to analyze the local influence of the variation of each input factor X_i around a few random base points $P^j = (X_1^j, X_2^j, ..., X_k^j)$ where X_i^j is the j^{th} realization (layer) of input factor X_i for k input factors. The following steps were followed for $j = 0, 1, 2$:

- select the realizations for the base point $P_\bullet^j$ from the space of the input factors
- for each input factor X_i:
  - calculate the model output for each spatial realization of X_i, with the other input factors (layers) remaining unchanged from the base point.
  - compute S_i^j, the range of model output Y for this set of runs (there are not enough runs to calculate variance).

S_i^j can be seen as a sensitivity measure of input factor X_i around base point P^j.

The total number of model evaluations is $C = p \times \sum n_i$ where n_i is the number of possible values for input factor X_i and p the number of base points studied. Here $C = 81$.

2) *Elementary Effects method*

The Elementary Effects method is a technique for sensitivity analysis that overcomes some of the limitations of the previous technique, whose results depend heavily on a single base point. It also belongs to the class of OAT sampling designs. In this study $r = 100$ different trajectories in the input space were generated from randomly selected base points. Each trajectory has (k+1) points, where k is the number of input factors, and has two key properties: two consecutive points differ in only one input factor, and each input factor has been varied exactly once in the trajectory. This approach can be adapted for the spatial case by using another realization to vary each each input factor (layer) in turn. From each trajectory t, the elementary effect EE_i^t of input factor X_i can be computed as:

$$EE_i^t = Y\left(X_1,...,X_{i-1},X_i^{\sim j},X_{i+1},...,X_k\right) - Y\left(X_1,X_2,...X_i^j...,X_k\right)$$

where $X_i^{\sim j}$ is a different layer realization to X_i^j :

For each input factor X_i, a sensitivity measure μ_i^* can be calculated. It is defined as the mean of the absolute values of the computed elementary effects EE_i^t over the r different trajectories:

$$\mu_i^* = \frac{1}{r}\sum_{t=1}^{r}\left|EE_i^t\right|$$

The total number of model evaluations is $C = r \times (k+1)$. Here $C = 600$.

3) Variance-based spatial Sobol' method

The spatial Sobol' method is a generalization of the methods of Sobol' (1993) and Saltelli et al. (2000) to spatially dependent models (Lilburne and Tarantola, 2009). Briefly, it is based on the decomposition of the output variance in conditional variances. It uses two quasi-random sample matrices to explore the space of the input factors. Each spatial factor (layer) X_i is sampled from a discrete uniform distribution in $\{0 ; \ldots ; n_i-1\}$. Numerous permutations between the two sample matrices allow the computation of Sobol' first-order and total-order sensitivity indices S_i and S_{Ti}, for each input factor. Empirical 90% confidence intervals for the Sobol' indices have been computed from the same set of model evaluations using 100 bootstrap replicas.

The computational cost of the spatial Sobol' method is considerably higher than that of the other two techniques: the number of model evaluations needed is $C = 2 \times N \times (k+1)$ where N is the size of the quasi-random sample matrices (here $N = 512$ rows) and k is the number of input factors (here $k = 5$). Here $C = 6144$.

III. Results

The results from OAT local variations of model inputs around some base points show that this type of analysis is heavily dependent on the selected base point. Local sensitivity measures of the input factors are different for each base point and no general ranking can be inferred (Table 2).

The Elementary Effects method gives more valuable information: land use layer is the most influential input (Table 3), while no input can clearly be discarded as having little influence.

The spatial Sobol' method leads to similar conclusions: the land use layer also appears to be the most influential input (Table 4). The sum of first-order effects is only 57%, showing that there are significant interactions between the inputs. The results also show that all of the inputs have some interactions with the other inputs.

IV. Discussion

Associating randomly generated map and tabular realizations to scalar values sampled from discrete uniform distributions makes it possible to use various sensitivity analysis techniques on the spatial AquiferSim model.

Complex descriptions of spatial uncertainty can be used to generate map realizations, including geostatistical techniques and simulation from confusion matrices.

OAT variations of each input factor around base sample points depends heavily on the selected sample point and it fails to identify the factors that account for most of the model output variability. Yet it is a relatively costless method to implement, and provides valuable information on the local behavior of the model around some specific points.

The most influential input, which is the land use layer, was clearly identified by the Elementary Effects method, with a quite low computational cost (Table 5). Yet this technique does not explore the whole space of the input factors, and some specific interaction effects could have been missed by the analysis. Moreover, the adaptation of this method to a spatial context raises some difficulties: specifically, as the integer index associated with each spatial realization has no meaning, the computation of sensitivity measures μ and σ (Morris, 1991) is not possible, therefore limiting the usefulness of the Elementary Effects method.

TABLE II. Results of OAT local variation about base sample points

	Base Point 1		**Base Point 2**		**Base Point 3**	
	Rank	*Range of Y*	*Rank*	*Range of Y*	*Rank*	*Range of Y*
Land use	1	1.73	3	0.87	2	0.81
River recharge & transmissivity	2	1.23	2	1.23	5	0.65
Soil	3	1.03	1	1.71	3	0.75
Nitrate/drainage table	4	0.73	4	0.67	4	0.70
Climate	5	0.09	5	0.45	1	1.29

TABLE III. RESULTS OF ELEMENTARY EFFECTS SENSITIVITY ANALYSIS

	Rank	μ*
Land use	1	68.57
River recharge & transmissivity	2	57.04
Climate	3	56.74
Nitrate/drainage table	5	49.51
Soil	6	44.66

TABLE IV. RESULTS OF SPATIAL SOBOL' SENSITIVITY ANALYSIS

	First-Order Sensitivity Index (90% Confidence Interval)	Total-Order Sensitivity Index (90% Confidence Interval)
Land use	15% – 28%	49% – 60%
River recharge & transmissivity	6% – 19%	22% – 29%
Nitrate/drainage table	5% – 19%	12% – 18%
Climate	0.5% – 12%	32% – 44%
Soil	0% – 11%	26% – 31%

The Spatial Sobol' method has none of those shortcomings, but at the price of a much higher computational cost (in this case nearly a week). It ensures that the entire model input space is explored. Land use was identified by the Spatial Sobol' method as being the input that accounts for most of the output variability. This method gives valuable insight on the interactions between input factors, which account for 43% of the output variance. Finally it shows that no inputs can be considered as negligible. With a number of sample rows of $N = 512$, confidence intervals on the sensitivity indices are small enough to draw firm conclusions.

V. CONCLUSION

By treating spatial layers as unique random variables, various sensitivity analysis techniques have been applied to the AquiferSim spatial model. Specific and complex descriptions of spatial uncertainty were used for each spatial input, to generate sets of random realizations. OAT local variations of inputs around some sample points allow the robustness of model predictions to be checked around those points, but give no insight on the relative importance of inputs. The Elementary Effects method shows that the land use layer is the most influential input factor: it would be worthwhile to obtain better land use information in order to reduce the output variability. The spatial Sobol' method gives the most valuable information: it confirms that the land use layer is the most influential input; it also shows that strong interactions occur between most of the inputs, explaining around 40% of the output variability.

TABLE V. SUMMARY COMPARISON BETWEEN THE THREE TECHNIQUES

	Model Evaluations	Ranking of Inputs	Inter-actions
Local OAT variations around base point	81	No	No
Elementary Effects	600	Yes	No
Spatial Sobol'	6144	Yes	Yes

ACKNOWLEDGEMENTS

We thank David Pairman, Stephen McNeill and Christine Bezar for their helpful comments. Nathalie's visit to New Zealand was supported by the research program "Risque, Décision, Territoire" (EPI) funded by MEEDEM, France. Funding for this work was provided by Foundation for Research Science and Technology (NZ), contract CO9X0306.

REFERENCES

Bidwell, V.J., Lilburne, L.R. and Good, J.M. (2005). Strategy for developing GIS-based tools for management of the effects on groundwater of nitrate leaching from agricultural land use. In: A. Zerger and R.M. Argent (Eds) *MODSIM 2005 International Congress on Modelling and Simulation.* Melbourne, Australia: Modelling and Simulation Society of Australia and New Zealand.

Chan, G. and Wood, A.T.A. (1997). An algorithm for stimulating stationary Gaussian random fields. *Applied Statistics.* 46, 171–181.

Harbaugh, A.W., Banta, E.R., Hill, M.C. and Mcdonald, M.G. (2000). *MODFLOW-2000, the U.S. Geological Survey modular ground-water model: User guide to modularization concepts and the ground-water flow process.* U.S. Geological Survey.

Helton, J.C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety.* 42, 327–367.

Lilburne, L. & Tarantola, S. (2009). Sensitivity analysis of spatial models. *International Journal of Geographical Information Science.* 23, 151–168.

Morris, M.D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics.* 33, 161–174.

Saltelli, A., Chan, K. and Scott, M. (Eds.) (2000). *Sensitivity analysis.* Chichester, UK: John Wiley.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. (2004). *Sensitivity analysis in practice. A guide to assessing scientific models.* Chichester, UK: John Wiley.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008). *Global sensitivity analysis. The primer.* Chichester, UK: John Wiley.

Sobol', I.M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment.* 1, 407–414.