A Meta-Modeling Based Tool for Spatially Explicit Uncertainty and Sensitivity Analysis

Seda Şalap-Ayça*[1,2] and Piotr Jankowski[1,3]

[1]San Diego State University, Department of Geography,5500 Campanile Dr., San Diego, CA, USA 92182-4493
[2]University of California, Santa Barbara, Department of Geography, 1832 Ellison Hall, Santa Barbara, CA, 93106-4060
[3]Institute of Geoecology and Geoinformation, Adam Mickiewicz University, Poznań, Poland
*Email: ssalap@mail.sdsu.edu

Abstract

The importance of uncertainty and sensitivity analysis for spatial modeling has been recognized by several scholars and organizations including the U.S. Environmental Protection Agency and the European Union. However, the high dimensionality and complexity of spatial models make the computation of uncertainty and sensitivity computationally expensive. One possible solution to this problem is a meta-modeling approach, which reduces the number of model simulations necessary to arrive at a stable sensitivity measure. In this paper, we describe a conceptual framework for meta-modeling and illustrate it with the implementation that uses the capabilities of surrogate modeling for analyzing model uncertainty and sensitivity. The python-based tool handles raster based model simulation results and produces the uncertainty and sensitivity maps for each input parameter and for their interactions. Using the example of an urban growth model, we show how the meta-modeling approach significantly reduces computational effort required to carry out spatially explicit uncertainty and sensitivity analysis.

Keywords: uncertainty analysis, sensitivity analysis, meta-modeling, spatio-temporal model

1. Introduction

Many real world spatial problems involve some aspect of uncertainty due to their unpredictable nature, aleatory character of data, or plain lack of knowledge. Modelling of such problem can be potentially improved by explicitly addressing the sources of uncertainties. The difficulty in this, however, is that due to the non-linear nature of processes and phenomena represented in spatial and spatio-temporal models, a model output may have a much greater uncertainty than individual model inputs due to interaction effects. One solution to this impediment is variance decomposition based global sensitivity analysis (GSA), which addresses both the individual and interaction effects and is agnostic to model structure and behavior (Helton, 2008; Saltelli *et al.*, 2010).

In variance decomposition based GSA, the overall variance (uncertainty) of the model output is partitioned into first order (S) and total order (S_T) sensitivity indices. The first order indices resemble the one-at-a-time approach (which also known as local sensitivity analysis), in which each input parameter is varied separately from other parameters to observe the variance in the model output. Additionally, total order indices are computed to understand how much variability is due to higher order interactions among the model inputs. This is especially critical for complex systems, where input parameters interact with each other.

Variance decomposition based GSA has been increasingly used for spatial models (Saint-Geours and Lilburne, 2010; Mara and Tarantola, 2012; Ligmann-Zielinska and Jankowski, 2014). However, for a typical application of variance-based GSA, a large number of Monte Carlo (MC) simulations is required to reach an acceptable accuracy of sensitivity indices. Therefore, a full-order variance decomposition becomes impractical in complex spatial and spatio-temporal models unless a computationally efficient solution can be provided (Helton, 1993; O'Hagan, 2011).

As a solution to this problem, one can build a response surface / surrogate/ meta-model, which requires a moderate number of MC simulations. With strong mathematical basis and the ability to produce functional representations of models, Polynomial Chaos Expansion (PCE) (Wiener, 1938) is a meta-modelling technique, which can be integrated with the variance-decomposition based GSA in order to analyze the uncertainty and sensitivity of spatial model. The PCE approximates the model by using orthogonal polynomials that can be solved with less computational effort than the original model. These orthogonal polynomials are based on the probability density functions of the input parameters. For example, for the normal (Gaussian) distribution, Hermite polynomials are used in the approximation (various probability density functions are discussed in Xiu and Karniadakis (2002) and Sudret (2008)).

Several application areas of PCE involve water quality modeling (Moreau *et al.*, 2013), large scale socio-hydrologic modeling coupled with Agent Based Models (Hu *et al.*, 2015), groundwater hydrogeological modeling (Deman *et al.*, 2016) and crop modeling (Lamboni *et al.*, 2009). Yet, not much has been done on testing the applicability of PCE for uncertainty/sensitivity analysis of complex system models where model input and output are spatially explicit. Moreover, the existing tools/software for analyzing uncertainty and sensitivity are usually either platform dependent (e.g. UQLab (Marelli and Sudret, 2014), DAKOTA(Adams *et al.*, 2015) or based on scalar model output, producing only tabular or graphic analysis results (e.g., UQ-PyL (Wang *et al.*, 2016), OpenTURNS (Baudin *et al.*, 2015)). The objective of the study reported herein, has been to contribute to the spatially explicit integrated uncertainty and sensitivity analysis (iUSA) of spatio-temporal models by developing and testing a computationally efficient approach which can be run on any platform.

2. Conceptual Framework

The proposed PCE-based iUSA framework has three stages: the initial stage, the PCE stage, and the variance decomposition (VD) stage. The initial stage is adoptable for any spatial/spatio-temporal model, which produces distributed (raster-format) output. Following this stage, the PCE and variance-decomposition stages can be initiated to produce uncertainty and sensitivity maps (Figure 1).

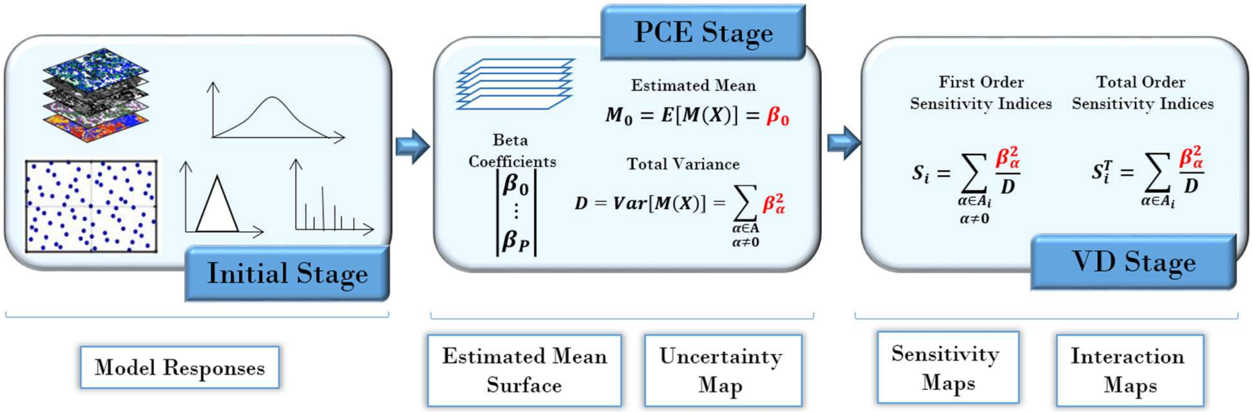

Figure 1 Polynomial chaos expansion based integrated uncertainty and sensitivity analysis (iUSA)

In the initial step of PCE based meta-modelling approach, in order to build the response surfaces (several model outputs) approximating the full model, the analyst should decide three key issues:

1) The number of input parameters that will be included in the iUSA (***M***)
2) The experimental design polynomial degree *p* (the highest order for the expansion), such that *p*<***M*** (where p=M yields the exact solution for the full degree)
3) The probability distribution functions for the selected input parameters (including the range of those functions) – this will also help to define the *orthogonal polynomial family* used in the expansion.

The estimated mean (μ) and variance (D) of the simulation output are two important statistical measures needed to analyze the sensitivity by variance decomposition. In the PCE based approach, these values are calculated based on beta coefficients; which are given by $P = \binom{M+p}{p}$, and are solved by using *N* model responses (*N=k*P where k is usually set to 2 or 3*) (Figure 2). The N model runs are represented in the *Y* output matrix, and the experimental matrix (i.e. the coefficient array, **A**) is calculated by solving orthogonal polynomials with a given experimental design polynomial degree *p*. Using the Equation 1, the beta coefficients are calculated at the PCE stage. After calculating the beta coefficients, the estimated mean (β_0), total variance (D), first order (S_i) and total order sensitivity indices (S_T) can be calculated at almost no additional cost to produce output maps.

$$|\hat{Y}|_{(P,(mxn))} = \left|\left(|A^T|_{(P,N)}|A|_{(N,P)}\right)^{-1}\right|_{(P,P)} * |A^T|_{(P,N)} * |Y|_{(N,(mxn))} = \begin{vmatrix}\beta_0 \\ \vdots \\ \beta_P\end{vmatrix} \qquad \text{Eq. (1)}$$

where (mxn) is the model output dimensions for a single run

3. Example of PCE-based Variance Decomposition

Let us consider a study region represented by a 50 rows x 100 columns matrix (5000 cells). We would like to forecast urban growth for this area by using a cellular automata (CA) based geosimulation. For this model, let us assume the cell's land use changes as a function of four behavior control parameters for each time step in the forecasting period. Given our interest in the model sensitivity to the variations in the four parameters, we have the value of *M* =4. Starting with the *p* value equal to 3, we will need to solve P= 35

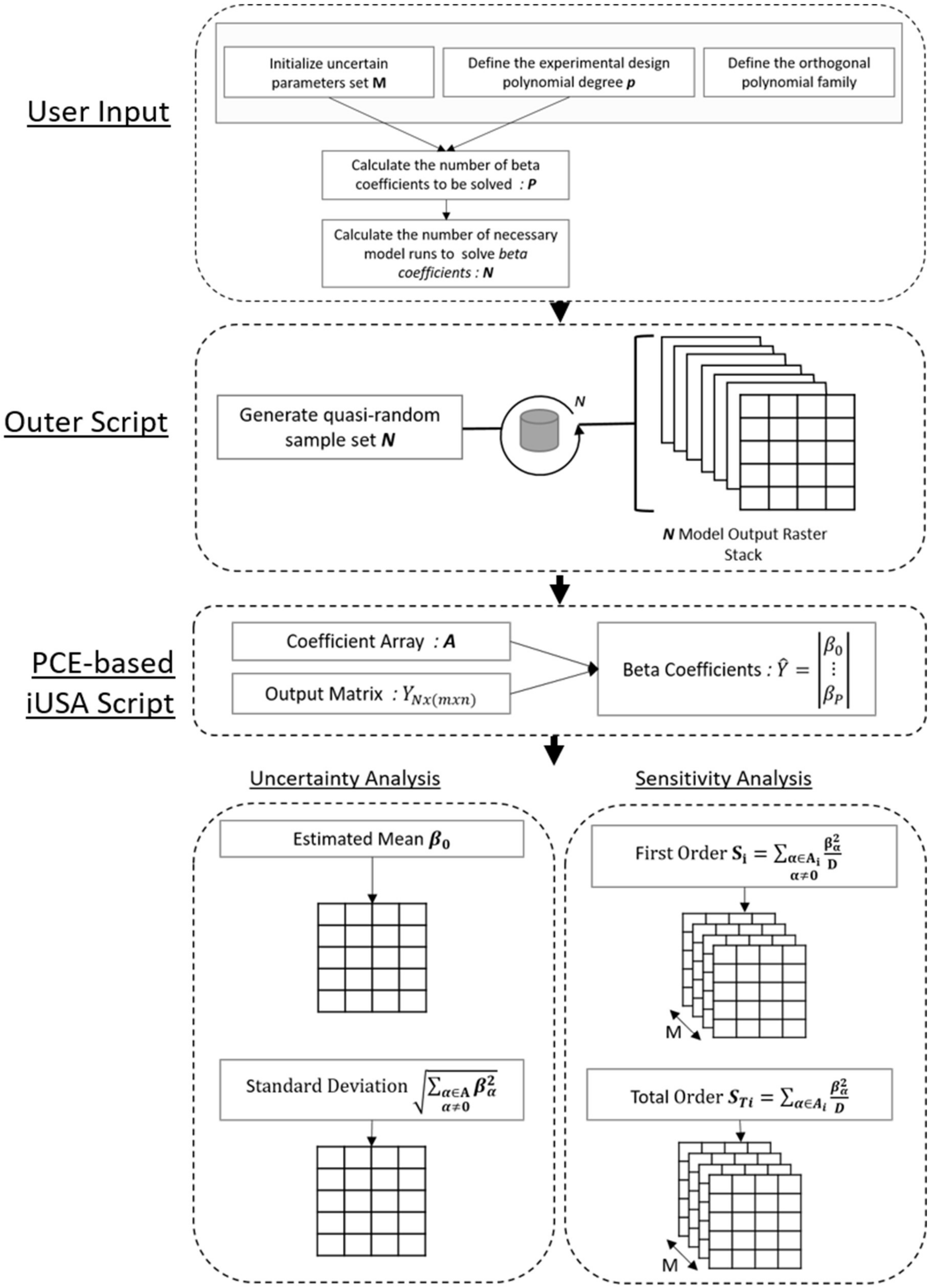

Figure 2 PCE based iUSA procedure

coefficients, where $P = \binom{M+p}{p}$. Given the number of MC simulations N=70 (for k=2), the size of the output matrix Y will be 70 rows by 5000 columns (one model run per cell) yielding 70*5000 model runs. For comparison, in the full-order variance-based decomposition, one would need to solve for M=p=4 yielding P=70 coefficients and N=140 (for k=2) simulations repeated for 5000 pixels (or 140*5000 model runs). Therefore, by applying meta-modeling and approximating the model with the third order orthogonal polynomials, we can reduce the number of calculations by 50%. A potential saving in the number of simulations will only exponentially increase with the increase in the number of model uncertain parameters since the calculation of P depends on factorial multiplication.

One computational efficiency introduced in the algorithm for PCE based iUSA procedure is through calculating the experimental matrix in a column-wise manner. This raster data division approach enables the algorithm to run for larger initial datasets even with smaller RAM configurations. Each member in the coefficient array is calculated by taking columns one at a time and saving the output in the corresponding column of the experimental matrix in Equation 1. For example, continuing our example of 50 x 100 area, for column one (simulation output for time t) the rows from 1 to n (50) are collected and for the 50 raster cell values the 1st column of the coefficients is calculated from N (70) simulation outputs. This is repeated m-1 times (m = 100). With this procedure, instead of taking all 70 simulation outputs and the full size of the output array (50*100) in a single computation, the process is divided into single raster column chunks enabling the calculation of PCE coefficients even for a small memory system. The procedure is depicted in Figure 3.

In the final stage of the procedure, the total model output variance, which is the sum of the squares of beta coefficients, is partitioned into the first and the total order sensitivity indices to produce sensitivity and interaction maps for each parameter. The output of the PCE based iUSA for this example will be comprised of one overall uncertainty map, one suitability map from the estimated means, four sensitivity maps for each parameter and four interactions maps for their higher order interactions.

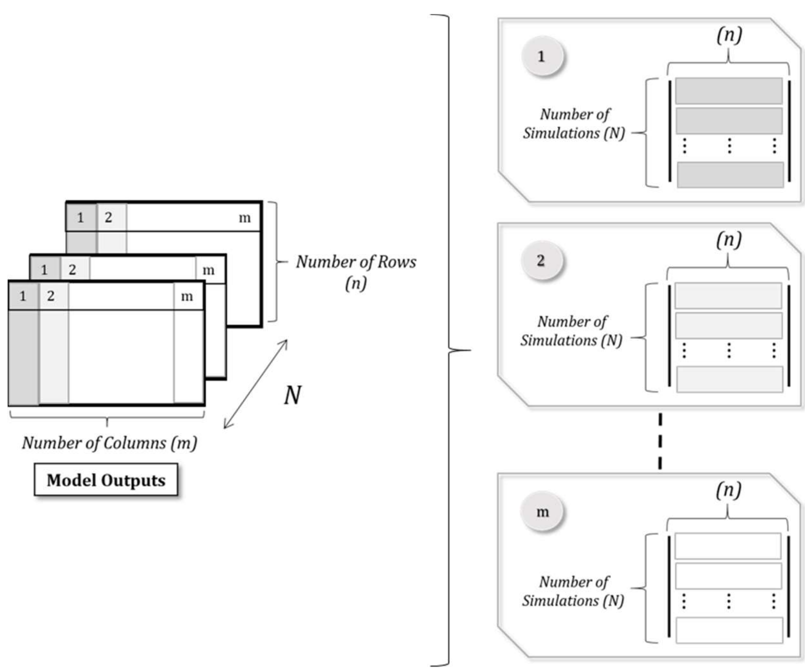

Figure 3 Column-wise computation partitioning for PCE implementation

4. Concluding Remarks

Variance decomposition is a powerful approach to gain confidence in model results, as well as a method to analyze partial and interactive effects of model input parameters where uncertainty is expected to occur. However, the major drawback of variance decomposition based methods is that a model must be run multiple times for each uncertain input parameter. Moreover, the number of model runs necessary to obtain the satisfactory level of confidence exponentially increases with the number of parameters. For complex spatial models with large input datasets, which are computationally expensive to evaluate even for a single run, the total computational cost increases drastically. A meta-modeling approach can reduce computational cost while delivering stable sensitivity estimates and helping to account for the dependence between model output variability and its inputs.

5. References

Adams, B. M., Ebeida, M. S., Eldred, M. S., Jakeman, J. D., Swiler, L. P., Stephens, J. A., Vigil, D. M., Wildey, T. M., Bohnhoff, W. J., Dalbey, K. R., Eddy, J. P., Hu, K. T., Bauman, L. E. and Hough, P. D. 2015. *Dakota , A Multilevel Parallel Object-Oriented Framework for Design Optimization , Parameter Estimation , Uncertainty Quantification , and Sensitivity Analysis : Version 6 . 2 User's Manual.*

Baudin, M., Dutfoy, A., Iooss, B. and Popelin, A.L.2015. *Open TURNS: An industrial software for uncertainty quantification in simulation.* Available at: http://arxiv.org/abs/1501.05242.

Deman, G., Konakli, K., Sudret, B., Kerrou, J., Perrochet, P. and Benabderrahmane, H. 2016.Using sparse polynomial chaos expansions for the global sensitivity analysis of groundwater lifetime expectancy in a multi-layered hydrogeological model.*Reliability Engineering and System Safety*. **147** (2016), pp. 156–169.

Helton, J. C.1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal.*Reliability Engineering & System Safety*.**42**(2–3), pp. 327–367.

Helton, J. C. 2008.Uncertainty and sensitivity analysis for models of complex systems.In Graziani, F.ed(s). *Computational Methods in Transport: Verification and Validation*,Springer.pp. 207-228.

Hu, Y., Garcia-Cabrejo, O., Cai, X., Valocchi, A. J. and DuPont, B.2015.Global sensitivity analysis for large-scale socio-hydrological models using Hadoop. *Environmental Modelling & Software*.**73**(2015), pp. 231–243.

Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B. and Monod, H.2009.Multivariate global sensitivity analysis for dynamic crop models.*Field Crops Research*.**113**(3),pp. 312–320.

Ligmann-Zielinska, A. and Jankowski, P. 2014. Spatially-explicit integrated uncertainty and sensitivity analysis of criteria weights in multicriteria land suitability evaluation. *Environmental Modelling & Software*. **57**(2014), pp. 235–247.

Mara, T. A. and Tarantola, S. 2012. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering and System Safety*. **107**(2012), pp. 115–121.

Marelli, S. and Sudret, B. 2014. UQLab : A framework for uncertainty quantification in Matlab. *Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014)*. Liverpool, United Kingdom. pp. 2554–2563.

Moreau, P., Viaud, V., Parnaudeau, V., Salmon-Monviola, J. and Durand, P. 2013.An approach for global sensitivity analysis of a complex environmental model to spatial inputs and parameters: A case study of an agro-hydrological model.*Environmental Modelling & Software*. **47**(2013), pp. 74–87.

O'Hagan, A. 2011.Polynomial Chaos: A Tutorial and Critique from a Statistician's Perspective.*SIAM/ASA J. Uncertainty Quantification*,20, pp. 1–16.

Saint-Geours, N. and Lilburne, L. 2010.Comparison of three spatial sensitivity analysis techniques', *Accuracy 2010,10 July 2010,Leicester, UK.* pp.421-424

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. and Tarantola, S. 2010.Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*. **181**(2), pp. 259–270.

Sudret, B. 2008. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*. **93**(7), pp. 964–979.

Wang, C., Duan, Q., Tong, C. H., Di, Z. and Gong, W. 2016. A GUI platform for uncertainty quantification of complex dynamical models. *Environmental Modelling and Software*.**76** (2016), pp. 1–12.

Wiener, N. 1938.The Homogenous Chaos. *American Journal of Mathematics*, **60**(4), pp. 897–936.

Xiu, D. and Karniadakis, G. E. 2002. The Wiener-Askey Polynomial Chaos for stochastic differential equations. *Society for Industrial and Applied Mathematics Journal of Scientific Computing*, **24**(2), pp. 619–644.