A meta-modeling approach for spatio-temporal uncertainty and sensitivity analysis: an application for a cellular automata-based Urban growth and land-use change model

Seda Şalap-Ayça, Piotr Jankowski, Keith C Clarke, Phaedon C Kyriakidis & Atsushi Nara

Published online: 29 Nov 2017.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

ARTICLE

Check for updates

A meta-modeling approach for spatio-temporal uncertainty and sensitivity analysis: an application for a cellular automata-based Urban growth and land-use change model

Seda Şalap-Ayça [a,b], Piotr Jankowski[a,c], Keith C Clarke [b], Phaedon C Kyriakidis[b,d] and Atsushi Nara[a]

aDepartment of Geography, San Diego State University, San Diego, CA, USA; bDepartment of Geography, University of California Santa Barbara, Santa Barbara, CA, USA; cInstitute of Geoecology and Geoinformation, Adam Mickiewicz University, Poznań, Poland; dDepartment of Civil Engineering and Geomatics, Cyprus University of Technology, Lemesos, Cyprus

ABSTRACT

The paper presents a computationally efficient meta-modeling approach to spatially explicit uncertainty and sensitivity analysis in a cellular automata (CA) urban growth and land-use simulation model. The uncertainty and sensitivity of the model parameters are approximated using a meta-modeling method called polynomial chaos expansion (PCE). The parameter uncertainty and sensitivity measures obtained with PCE are compared with traditional Monte Carlo simulation results. The meta-modeling approach was found to reduce the number of model simulations necessary to arrive at stable sensitivity estimates. The quality of the results is comparable to the full-order modeling approach, which is computationally costly. The study shows that the meta-modeling approach can significantly reduce the computational effort of carrying out spatially explicit uncertainty and sensitivity analysis in the application of spatio-temporal models.

Introduction

Increasing urban population directly affects the pressure on the available urban and suburban land with negative impacts on biodiversity in favor of urbanization. To represent these land-use dynamics, a cell-based spatial diffusion approach called cellular automata (CA) has been frequently used for urban growth and land-use change (UG-LUC) modeling (De Almeida *et al.* 2003, Torrens and Nara 2007, Clarke 2008, Moreno *et al.* 2008, Van Vliet *et al.* 2016). CA-based models mostly involve many input factors including variables and structural elements (e.g. transition rules). Any confidence in the results depends highly on these input factors. It is therefore expected that some of the initial conditions may be the drivers of model uncertainty (Kocabas and Dragićević 2006, Li *et al.* 2014, Dahal and Chow 2015). In the same context, it is critical to evaluate the influence of input factors on model's output so as to provide decision makers with confidence in their actionable information. To this end, uncertainty and sensitivity analysis offer a way forward.

Uncertainty analysis (UA) quantifies the model output variability, and sensitivity analysis (SA) investigates how this uncertainty is apportioned among the model input factors (Saltelli *et al*. 2000, Crosetto and Tarantola 2001). In particular, an uncertainty and SA (U-SA) approach that is both independent from model structure and capable of handling interaction effects is important for complex, nonlinear models (Halls 2002, Chen *et al*. 2010, Roura-Pascual *et al*. 2010, Moreau *et al*. 2013, Xu and Zhang 2013, Saint-Geours *et al*. 2014). Moreover, for spatio-temporal models, where the simulation results are spatially distributed, it is additionally important to identify not only the source(s) of uncertainty but also its location(s) at a specific time (Herman *et al*. 2013, Abily *et al*. 2016). The resulting uncertainty maps can assist in locating uncertainty hot spots, and the sensitivity maps help further in identifying the spatial pattern of influential input factors behind the areas of high uncertainty (Ligmann-Zielinska 2013, Şalap-Ayça and Jankowski 2016). However, the traditional Monte Carlo-based methods for U-SA require a large number of model evaluations; such a requirement makes these approaches intractable for computationally expansive models (Saltelli *et al*. 2010).

One approach to overcoming the computational bottleneck is high-performance computing in the form of distributed and/or parallel computing (Tang and Jia 2014, Hu *et al*. 2015). Tang and Jia (2014) and Erlacher *et al*. (2017) accelerated U-SA with graphics processing units (GPU) and Ligmann-Zielinska and Jankowski (2014) used a supercomputer to overcome intractable models. Yet another approach without the need of access to supercomputers or computationally powerful hardware is the use of surrogate or meta-models (Saltelli *et al*. 2000). The meta-model, which is fitted to the model by a set of experiments or model runs, replicates the behavior of the original model in the domain of its influential input parameters (Oakley and Hagan 2000, Marrel *et al*. 2011).

Although it is common to calibrate and validate CA models for achieving a desirable land-use pattern accuracy, only limited attention has been paid to spatially explicit U-SA. In this study, we aim to contribute to the body of knowledge on spatio-temporal U-SA by using a meta-modeling technique called Polynomial Chaos Expansion (PCE), which can reduce the computational effort of performing a U-SA. We implemented our approach on a CA-based model, called SLEUTH (Clarke *et al*. 1997), which has been widely used for simulating UG-LUC. We also compared the results of PCE to the full-order Monte Carlo (MC) approach to show how close the PCE reproduces uncertainty measures with fewer model computations than MC. For the implementation, we first analyzed model output uncertainty on UA maps to observe where and when the range of variations is most influential. We followed up on the UA by studying sensitivity maps to identify the most influential inputs.

In the remainder of the paper, we provide background information on UG-LUC forecasting, including SLEUTH modeling, and on the spatially-explicit approach to integrated U-SA. In section three, we explain the basis of the PCE method. We then show, in section four, how PCE can be applied in concert with U-SA to reveal the influence of the spatio-temporal variations of input parameters on the model output.

Background

This section provides a brief introduction to CA as a simulation method for UG-LUC modeling, followed by an introduction to U-SA and meta-modeling.

CA-based urban growth and land-use forecasting

CA, pioneered by Ulam and Von Neumann (Ulam 1952, Von Neumann 1966), is based on the generated dynamic behaviors under defined rules and constraints for certain configurations (neighborhoods) (Holland 1999). Due to the similarity between CA and field representations (i.e. a grid), especially for spatial systems based on grid representations, the applications of CA proliferated in the understanding of complex geographical systems (pioneering examples include: Tobler 1979, Couclelis 1985, White and Engelen 1993, 1994).

Within a set of regular cells, the cellular automaton, A, is defined by the set of states, S, that change in discrete time steps based on transition rules, T, within a specified neighborhood, N. Therefore, the basic formulation of a cellular automaton can be written as:

$$A \sim left(S, T, N)$$

For urban growth, the cell's state is guided by transition rules, which reflect the complexity of the environment. These rules act as a link between the spatial patterns and the underlying spatial process. The transition rules are applied within the cell's neighborhood in temporal increments. The rules are not necessarily applied uniformly to all cells since the evolution does not encapsulate each land form (i.e. water bodies in UG-LUC modeling). However, the rules capture the intrinsic variability of the model's nature. As a result, CA model design is highly sensitive to the transition rules (White and Engelen 1993, Torrens and O'Sullivan 2000, Ménard and Marceau 2005, Kocabas and Dragićević 2006, Samat 2006, Yeh and Li 2006, Pan *et al*. 2010, Pontius and Neeti 2010).

SLEUTH for urban growth and land-use change modeling

SLEUTH is a widely used CA-based UG-LUC model, which is available along with a data repository for various cities on the United States Geologic Survey (USGS) affiliated, National Center for Geographic Information and Analysis (NCGIA) Project Gigalopolis website (NCGIA n.d.). Comprehensive reviews on its usage and application areas can be found in the literature (Clarke *et al*. 2007, Clarke 2008, Chaudhuri and Clarke 2013).

SLEUTH is the acronym for slope, land use, exclusion, urban extent over time, transportation, and hill-shaded backdrop layers. The model is initialized with these 6 input layers and a scenario file containing all the necessary parameters for simulation (i.e. UG forecast parameters and time steps for forecast) (Figure 1). The model is run for each year of the simulation time frame $[t_1 \ldots t_T]$ and the output is the urban growth/land use for each year.

SLEUTH simulates UG by using four transition (growth) rules, which include spontaneous growth $F_S(\delta, \gamma)$, new spreading center $F_{NS}(\beta, \gamma)$, edge growth $F_E(\chi, \gamma)$ and road influenced growth $F_R(\beta, \rho, \delta, \gamma)$ where, δ symbolizes a value for diffusion, β breed, χ spread, γ slope resistance, and ρ road gravity coefficients given as UG forecast parameters (Candau *et al*. 2000). Urban growth transition rules are bounded by suitability defined by the exclusion layer and the slope gradient. The exclusion areas (such as water bodies, preserved areas, or parks) or slopes greater than a critical level (often 21%) are considered as less likely to be urbanized. The resulting urban growth, achieved with four
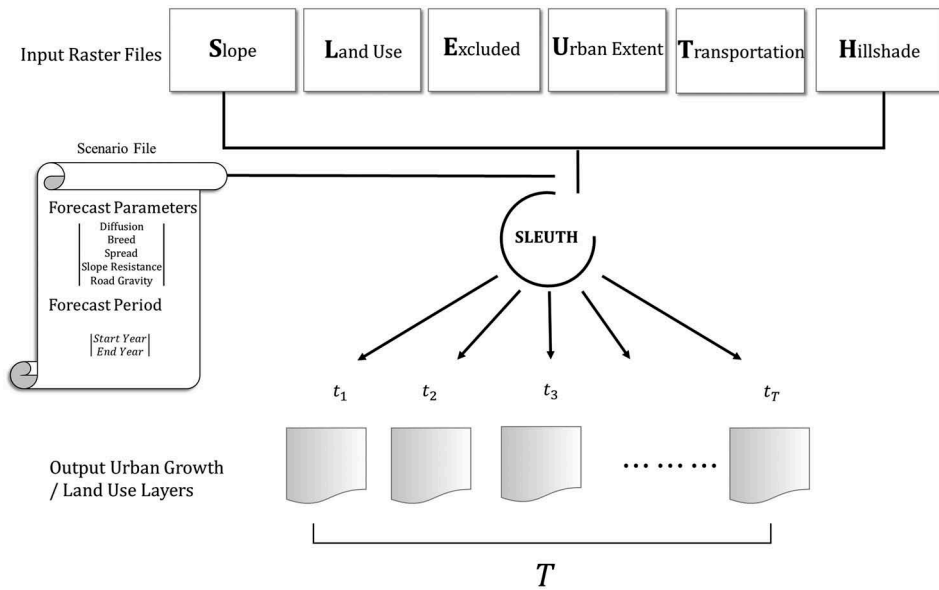
Figure 1. SLEUTH workflow.

different transition rules, serves as the basis for calculating the growth rate expressed by the ratio of the number of growth pixels to the total urban pixel population in SLEUTH (Clarke *et al*. 1997).

Uncertainty and global SA for complex system models

In spatial models, UA is usually referred to as error propagation, and SA is considered as the analysis which aims to understand the behavior of the error/uncertainty and its reflection on the predictions made by the models (Heuvelink 1989, 1998, Heuvelink and Burrough 1993, 2002).

For CA-based UG-LUC models, early examples of SA focused on cell by cell inspection (White *et al*. 1997), finding an optimum neighboring size and type (Ménard and Marceau 2005, Kocabas and Dragićević 2006), investigating the variations in the output by applying different cell sizes and cellular configurations (Chen and Mynett 2003, Samat 2006, Pan *et al*. 2010) or comparing different scenario maps for land-use change and validating results with historical data (Pontius and Neeti 2010) . However, relatively little attention has been paid to the transition rules.

SLEUTH, like any other CA model, is subject to various uncertainties associated with the transition rules. The model employs a calibration process to define the best fit range for each forecast parameter used in the transition rules, however, the parameter values highly depend on the data and the user's expertise. The reliability of SLEUTH simulations depends on the forecast parameters, the model's sensitivity to these parameters and their interactions. A simple one-at-a-time (OAT) approach or (mono-looping) approach to quantifying the influence of different forecast parameter values on simulation results fails to catch the higher-order effects resulting from parameter interactions nor does it

reveal their partial contribution to the overall variance of the model results. In light of the above mentioned studies, U-SA is essential to increase the credibility of the UG-LUC models, as well as to test the model robustness in producing realistic outputs for future implementations (Jantz *et al*. 2003, Jantz and Goetz 2005, Kim 2013). Additionally, a spatially-explicit approach is also crucial for spatial models that do not entirely depend on scalar inputs but that are also affected by spatial relationships (Herman *et al*. 2013, Abily *et al*. 2016),

Variance, which is a central measure in error propagation and SA, allows partitioning the effects of uncertainty in the input factors on model output (Kyriakidis and Goodchild 2006). There is a range of different theoretical and methodological approaches to variance-based SA in the literature. Generally, SA is discussed under two categories: local and global (Saltelli *et al*. 2000, Helton and Davis 2002). In a local approach, one investigates the input variations by estimating partial derivatives of model input factors as varying one and keeping others constant (OAT approach) (Crosetto and Tarantola 2001). OAT is computationally efficient and it does not need a large number of model executions. However, since it only investigates individual variations, it is mostly applied when no higher-order interactions are expected (Crosetto *et al*. 2000, Gómez-Delgado and Tarantola 2006, Lilburne and Tarantola 2009).

Interactions among model inputs and nonlinear output responses can be addressed by global SA (GSA) techniques where the sensitivity of the model input factors is not only determined individually (first-order effects), but also as overall interactions (total-order effects). Specifically, in variance-based GSA, the model factors influence on model output variability is represented by first-order (S) and total-order (S_T) sensitivity indices. The first-order index (S), sometimes referred to as the main effect sensitivity, estimates directly the selected single variable's portion of the overall variance (Sobol' 2001). The total-order index (S_T) measures the total contribution of pairs of factors, i.e. first- and higher-order interactions of the selected single variable (Saltelli *et al*. 2010, Saint-Geours *et al*. 2014). Since CA models are characterized by interactions among input factors and non-linear output responses, the variance-based GSA is potentially an attractive SA method for different types of spatio-temporal models, including for the CA-based SLEUTH, where higher-order interactions result from the transition rules.

Variance-based global SA and polynomial chaos expansion

GSA variance-decomposition-based methods (e.g. *Fourier Amplitude Sensitivity Test* (FAST) (Cukier *et al*. 1973), *Extended FAST* (E-FAST) (Saltelli *et al*. 1999), *Sobol Decomposition* (Sobol' 1993)) use MC simulation (Metropolis and Ulam 1949) as a sampling method during uncertainty propagation to draw multiple random samples from a given input distribution. A model is then run for each random sample set to obtain simulated outputs. The computational cost of input sample generation (required to run MC) is insignificant in comparison to the cost of model simulations (required to quantify the output uncertainty), especially for complex models with large input datasets (Crosetto *et al*. 2000). Consequently, the high number of simulations results in an expensive uncertainty evaluation and SA becomes impractical (Helton 1993, Heuvelink 2003).

As a practical solution to this problem, meta-models can evaluate a model's response by using a mathematical model approximation. SA based on meta-modeling starts with the selection of the ranges and distribution for each input factor and continues with the development of the experimental design to define the combinations of factor values, on which the model evaluation will be based. Therefore, although the sequence of steps seems to be exactly the same as in the comprehensive SA, the experimental design that uses the selection of design points (points representing parameter values in sampling space) makes the meta-model approach distinct (Saltelli *et al.* 2000). The approximation is based on these design points and their selection is determined according to the presence of higher-order effects, the number of variables under consideration, and the computational effort required to evaluate the model. Different types of experimental designs are available such as multiple linear, nonlinear regression (e.g. Ratto and Pagano 2012), neural networks (e.g.Villa-Vialaneix *et al.* 2012), cubic splines (Rutherford *et al.* 2015), Gaussian processes (e.g. Marrel *et al.* 2011), and orthogonal polynomials (e.g. Sudret 2008).

One of the most frequently used meta-modeling techniques, Polynomial Chaos Expansion (PCE), originally proposed by Norbert Wiener in 1938, propagates uncertainty (which the 'chaos' term refers to) by expanding a complex function into orthogonal polynomials that could be solved for relatively more easily than the original function (Wiener 1938). These polynomials are selected from a family of polynomials which are orthogonal with respect to the probability distribution of the corresponding input parameter (Burnaev *et al.* 2017). For example, Legendre polynomials, which are defined over the range [−1, 1], are orthogonal to a uniform distribution. Therefore, if PCE is applied for an input parameter set that has a uniform distribution, Legendre polynomials are used for expansion. PCE is preferable compared to other meta-modeling methods due to its applicability for dynamic stochastic systems, where there is an unavoidable uncertainty term in the system parameters. Theoretical studies with different polynomial functions and the applicability of various probability density functions are discussed by Xiu and Karniadakis (2002), Sudret (2008), and Crestaux *et al.* (2009).

Since PCE reduces the computational expense of uncertainty propagation, it has been widely applied in complex environmental problems including 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 in other dynamic modeling examples such as crop modeling (Lamboni *et al.* 2009) and seawater intrusion (Rajabi *et al.* 2015). Moreover, an extensive review of basic principles and applications of PCE in computational fluid dynamics was conducted by Najm (2009). PCE has also been used in the geostatistical literature under the term bivariate isofactorial models in the context of disjunctive kriging (Wackernagel 2003). The objective in those applications was to decompose a bivariate distribution into a series of orthogonal polynomials, therefore, finding isofactorial representations which help later to determine recurrence relations between factors (Armstrong and Matheron 1986, Chiles and Delfiner 1999).

There are only a few studies that have employed PCE in spatio-temporal modeling. In one study, PCE was applied for a CA-based lava flow forecasting model to evaluate the model's input parameters (Bilotta *et al.* 2012). They improved the use of the forecasting model by assessing the impact of measurement errors on the results of simulation and

subsequently identified the model's critical parameters. In another study, a PCE-based SA was combined with high-performance computational techniques for a spatially explicit, large-scale socio-hydrological model developed by Hu *et al.* (2015). The combination of the PCE with high-performance computing made possible the analysis of the effect of spatio-temporal variations of the input parameters on the output of a model even with big multi-dimensional data (Hu *et al.* 2015). Their result of PCE-based variance decomposition helped to identify the influential parameters, quantify their interactions, and prioritize the factors.

The polynomial chaos expansion method

PCE starts with the selection of input parameters where the largest uncertainty is expected, or to which the model is suspected to be most sensitive. The number of selected input parameters is represented by M, which determines the dimension of the random variable matrix **X**. The random variables are independently defined based on a probability distribution (e.g. uniform, normal, log-normal, etc.) and the model output Y is defined as $Y = f(\boldsymbol{X})$. This expansion can be represented by a series of independent random variables and orthogonal polynomials (Wiener 1938):

$$Y = f(\boldsymbol{X}) \approx f_{PC}(\boldsymbol{X}) = \sum_{j=0}^{P-1} \beta_j \psi_j(\boldsymbol{X}) \tag{1}$$

where ψ_j denotes the type of the *j*-th (out of *P-1*) orthogonal polynomials, β_j denotes the unknown response coefficient, and P is the number of unknown response coefficients (β) to be estimated. The variable set β is also referred to as the design points and these coefficients are used during the post-processing stage when calculating the results for SA. A **β** matrix is calculated as in Equation (2):

$$\begin{vmatrix} \beta_0 \\ \vdots \\ \beta_P \end{vmatrix}_{(P,(mxn))} = \left| \left(\left|A^T\right|_{(P,N)} |A|_{(N,P)} \right)^{-1} \right|_{(P,P)} * \left|A^T\right|_{(P,N)} * |Y|_{(N,(mxn))} \tag{2}$$

where (*mxn*) is the model output size (*m* = number of columns and *n* = number of rows) for a single run, **A** is the experimental matrix and **Y** is the output matrix coming from *N* model simulations (i.e. the number of runs necessary to solve β coefficients). PCE is applied in this work at the local spatial level, which means calculation is done for each pixel independently in matrix **Y**. Following Equation (2), SA results depend on the specific model output type, and presumably, they can be different for different model output constituents. Likewise, the order of importance of input factors could vary for different output constituents.

The value selected for *N* (number of model runs) should at least be equal to $k*P$ *where* $k \in [2, 3]$ (the selection of *k* is explained in detail in Sudret 2008) where $P = \begin{pmatrix} M + p \\ p \end{pmatrix} = \frac{(M+p)!}{p!M!}$ which means that the number of unknown coefficients to be solved depends on the number of input variables under interest (*M*) and the experimental degree of the polynomial (*p*).

Therefore, to build the meta-model which approximates the full approach and to calculate the estimated mean and variance for U-SA, there are five key steps:

(1) Determining the input parameters and the probability distribution functions for the selected parameters (which also defines the orthogonal polynomials)
(2) Selecting an experimental design degree, p, and calculating the number of unknown response coefficients, P
(3) Forming the experimental matrix, $\mathbf{A}$ and calculating the unknown response coefficients, $\boldsymbol{\beta}$
(4) Computing mean and total variance for UA
(5) Computing Sobol indices

Orthogonal polynomials

The type of orthogonal polynomials, ψ_j, is derived from the orthogonal bases of input factor probability density functions (PDFs). This affords the reduced dimensionality by taking advantage of the resemblance of weight functions (i.e. inner product vector) to the PDF of certain random distributions (Xiu and Karniadakis 2002). Therefore, depending on the PDF of input parameters, a corresponding orthogonal polynomial scheme is selected for PCE. The most frequently used orthogonal polynomials are Legendre for the uniform and Hermite for the Gaussian (normal) distribution, and these polynomial functions for a 5th degree model are given in Table 1 (Xiu et al. 2002, Sudret 2014).

Experimental design degree and expansion terms

For a full analytical solution, experimental design degree equals model degree. The difference between full-order solution, where P is calculated as $M = p$, and any p value where $p < M$, gives the overall computational gain. Sudret (2008) suggested a two-step strategy starting with a low-order expansion degree for factor prioritization, then continuing with a higher-order design degree to compute the main sensitivity indices. The relative error between the full analytical solution and its approximation can be obtained by experimenting with different p values (several functional model comparisons are presented in Sudret (2008)). However, increasing p values does not necessarily guarantee that the distribution is well approximated (O'Hagan 2011). Therefore, we checked the PCE approximation results with MC simulation for estimated mean and standard deviation surfaces.

Experimental matrix and computation of the response coefficients

After selecting p and calculating N, a random sample set of $\xi = \{\xi_1, \ldots, \xi_n\}$ from the corresponding distribution function of each input parameter $\boldsymbol{X} = \{\xi^{(1)}, \ldots, \xi^{(N)}\}$ is generated, and the model is evaluated for each sample to produce the output $\boldsymbol{Y} = \{Y(\xi^{(1)}), \ldots, Y(\xi^{(N)})\}$. Then, for each pixel, the experimental matrix $\boldsymbol{A}_{ij}$ can be computed using the previously defined orthogonal polynomials (Table 1) as follows:

Table 1. Orthogonal polynomials for selected probability distribution types (Sudret 2014).

Distribution type	Orthogonal polynomials	First five polynomials
Uniform	Legendre $P_k(x)$	$P_0(x) = 1$
		$P_1(x) = x$
		$P_2(x) = \frac{1}{2}(3x^2 - 1)$
		$P_3(x) = \frac{1}{2}(5x^3 - 3x)$
		$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$
		$P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x)$
Gaussian (Normal)	Hermite $H_k(x)$	$H_0(x) = 1$
		$H_1(x) = x$
		$H_2(x) = x^2 - 1$
		$H_3(x) = x^3 - 3x$
		$H_4(x) = x^4 - 6x^2 + 3$
		$H_5(x) = x^5 - 10x^3 + 15x$

$$\boldsymbol{A}_{ij} = \psi_j(x_n) \ n = 1, \ldots, M, \ i = 0, \ldots, P - 1 \ and \ j = 1, \ldots, N$$

$$\boldsymbol{A}_{ij} = \begin{vmatrix} \psi_0(x_1, x_2, x_3, x_4, x_5)^{\ 1} & \ldots & \psi_{P-1}(x_1 x_2, x_3, x_4, x_5)^1 \\ \vdots & \ddots & \vdots \\ \psi_0(x_1, x_2, x_3, x_4, x_5)^N & \ldots & \psi_{P-1}(x_1, x_2, x_3, x_4, x_5)^N \end{vmatrix} \ for \ M = 5$$

Finally, after calculating $\boldsymbol{A}_{ij}$ and gathering simulation results as $\boldsymbol{Y}$, response coefficients can be estimated from Equation 2. From the first response coefficient, β_0, the estimated mean, $\mu = E\left[\sum_{j=0}^{P-1} f_i \psi_j(\boldsymbol{X})\right]$, can be obtained at almost no additional cost, since $\mu = \beta_0$.

Computation of Sobol sensitivity indices for variance decomposition

Sobol's variance-based decomposition (usually referred to as Sobol decomposition) decomposes the output variance into fractions so that the contribution of each input can be traced with the first-order and total effects (Saltelli *et al.* 2010). Sensitivity indices derived from PCE are also referred to as PCE-based Sobol indices (*SU*) (Sudret 2008). First-order *SU* and total-order SU^T PCE-based Sobol indices can be computed using Equations (3) and (4). The first-order index represents the main influence of the *i*th term on the total variance for all the input parameters from $(i_1, \cdots, i_s)$ set and the SU^T is the influence from higher-order interactions among all the parameters from the integer sequence of $\mathcal{J}_{(j_1, \cdots j_t)} = \{(i_1, \cdots, i_s), (j_1, \cdots, j_t) \subset (i_1, \cdots, i_s)\}$ (for more details see Sudret 2008):

$$SU_{i_1, \cdots, i_s} = \sum_{a \in \psi_{i_1, \cdots, i_s}} f_a^2 E[\psi_a^2] / D_{PC} \tag{3}$$

$$SU^T_{j_1, \cdots, j_t} = \sum_{(i_1, \cdots, i_s) \in \mathcal{I}_{j_1, \cdots j_t}} SU_{i_1, \cdots, i_s} \tag{4}$$

where D_{PC} corresponds to the total variance determined with PCE and can be computed from Equation (5):

$$D_{PC} = Var\left[\sum_{j=0}^{P-1} f_i \psi_j(\boldsymbol{X})\right] = \sum_{j=1}^{P-1} f_i^2 E\left[\psi_j^2(\boldsymbol{X})\right] \quad (5)$$

Application of PCE-based GSA in a probabilistic CA-based UG-LUC model

SLEUTH set-up

Santa Barbara, California was selected for the case study due to: (1) availability of data, and (2) the authors' familiarity with the study site. SLEUTH produces distributed (spatially-explicit) output for each time step selected in the forecast scenario. UG in SLEUTH takes place in a three-step cycle: setting up the growth coefficients, applying the growth rules, and finally evaluating the growth rate. The general function governing SLEUTH simulations can be written as:

$$y(t) = f(x, \varepsilon, t)$$

where x represents the vector of input variables (input layers), ε is the vector of behavior control parameters (forecast parameters) and t is the time step (each year in time span T). The input vector (x) is composed of an array of size (486x2074) for the Santa Barbara area and kept constant for every simulation. ε is replaced by the $\boldsymbol{X}$ quasi random-sample matrix. The forecast period, T, is set from 1999 to 2016 (18 years).

The uncertainty-SA

Quasi-random sequences produce random numbers but the selected numbers 'know' the positions of previously sampled numbers and therefore, they do not form clusters or gaps in the sample space. Since there is no a priori information about the distribution of these parameters, each is assumed to be uniformly distributed. The lower and upper bounds for each parameter are used as ranges for the uniform distribution for quasi-random sample generation. The quasi-random sample matrix is generated based on the ranges in Table 2. The uniform distribution is also used when selecting the polynomial basis for PCE, which corresponds to Legendre orthogonal polynomials. The samples generated with the quasi-random sampling scheme are normalized to a [−1, 1] uniform distribution interval.

The model is run for N times for *(mxn)* pixels and produce the urban growth land-use extent for each year of the simulation time frame [$t_1...t_T$] as output . The model outputs are collected to analyze urban growth and change by comparing the intermediate simulation results. Since we are interested in how the forecasted urban growth changes at each time step, the volume of data required to estimate the Sobol sensitivity indices at each particular pixel at each time step is *T x N*. The variance decomposition is expected to show the influence of forecast parameters on the urban land-use output.

Table 2. Selected input parameters for SA.

	Diffusion	Breed	Spread	Slope	Road
Calibrated/Optimum	40	41	100	1	23
Lower Bound	36	36.9	90	0.9	20.7
Upper Bound	44	45.1	110	1.1	25.3

Calculation of PCE-based Sobol sensitivity indices

Since there are five forecast parameters (diffusion, breed, spread, slope and road gravity), the degree of the model becomes $M = 5$, and if we select our experimental design degree as $p = 3$ and take $k = 2$, the number of model simulations necessary to solve the unknown coefficients becomes $= k * P = 2 * \frac{(5+3)!}{3!5!} = 112$ yielding the Y output matrix of (112*(486(rows)*2074(columns))) size. Therefore, for PCE-based U-SA, 112 runs are enough to calculate the sensitivity indices for the time period (which yields 18*112 = 2016 model outputs). For a full-order approach, where $p = 5$, N will be $= k * P = 2 * \frac{(5+5)!}{5!5!} = 504$, which requires 78% more computations for each pixel. Another comparison can be made for the Monte Carlo-based variance-decomposition described by Saltelli *et al.* (2010), where the number of simulations required to compute the Sobol indices equals $(k + 2)$ *N where *k* is the number of uncertain model factors, and *N* is the number of factor samples. *N* is recommended to be large enough (≥ 1000) to give reliable estimates for the sensitivity indices. However, large values for *N* are computationally expensive and require more processing memory and time. Even for the number of samples $N = 112$ used in PCE-based approximation, at least $N = (5 + 2) * 112 = 784$ simulation samples would be required to perform the full variance-based decomposition described by Saltelli *et al.* (2010).

Although 112 simulations were found to be enough to conduct PCE-based U-SA, SLEUTH was run 1000 times to have enough runs to compare the PCE approximation with the sample-based MC simulation for the mean and total variance. After solving the polynomials and calculating coefficients for the five input parameters, the PCE-based Sobol decomposition resulted in the calculation of first, second, third, fourth and fifth-order indices, which were summed to yield the total Sobol indices.

Results and discussion

Uncertainty analysis

For PCE, 112*18 simulation outputs were used to compute statistical moments of estimated mean, total variance, first-order and total-order indices. Figure 2 depicts an example of these outputs for 2016 just showing the urban land-use class extent. Starting with the initial land-use layer, the urban extent was calculated from the cumulative result of equally probable growth cycles completed in SLEUTH.

For the mean value comparison, the first (1999), middle (2008) and the last simulation year (2016) were selected to show the PCE-based estimated mean compared with the mean resulting from 1000 MC model simulations. Figures 3–5 show that the estimated
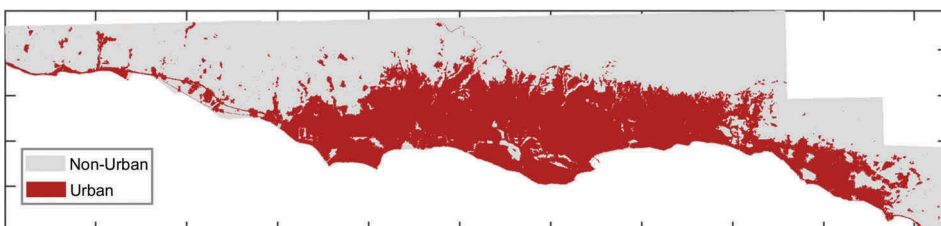

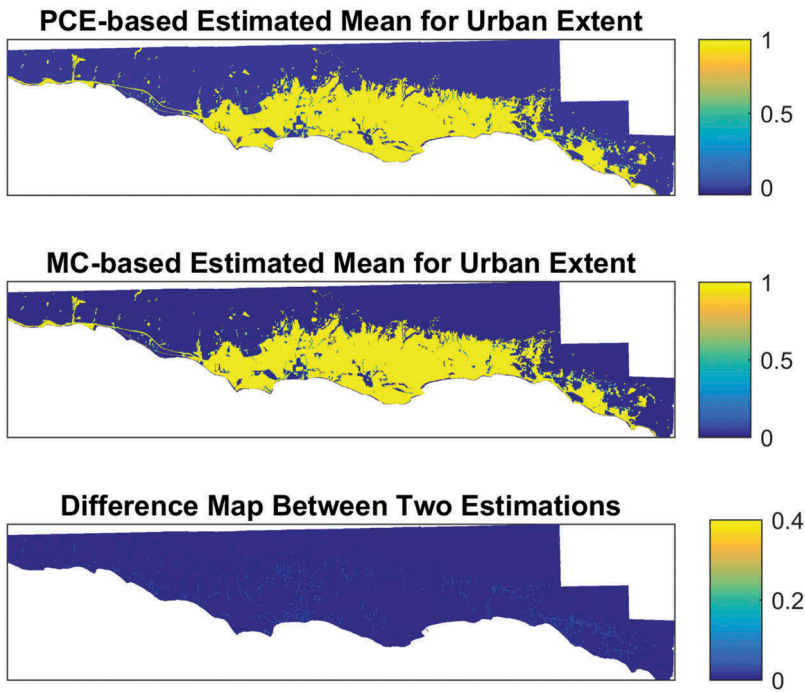**Figure 2.** Forecast Urban Extent for 2016.

Figure 3. The estimated means for the probability of a cell to change into urban land use for the year 1999: PCE-based Estimated Mean (top), MC-based Estimated Mean (middle), difference between the PCE-based Estimated and MC-based Estimated Mean Surfaces (bottom).
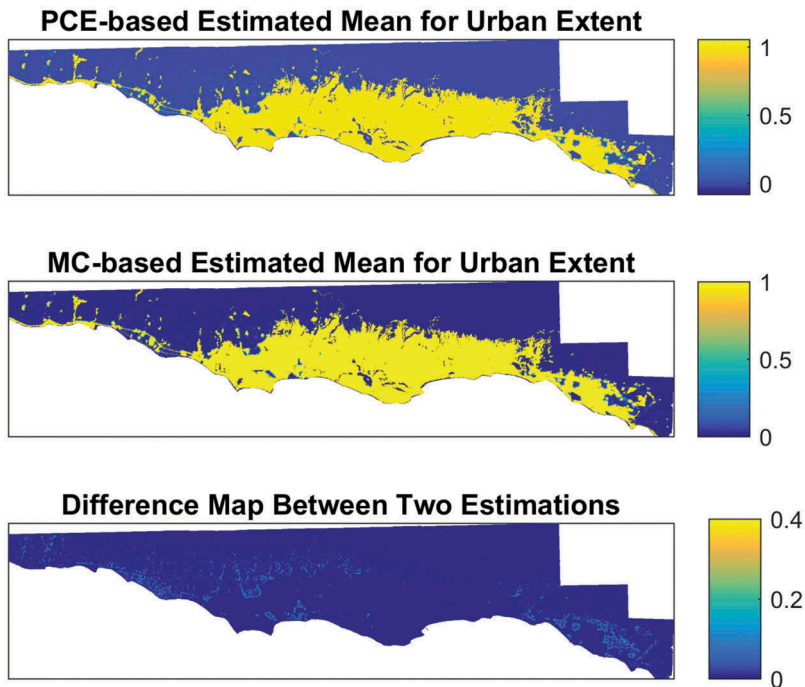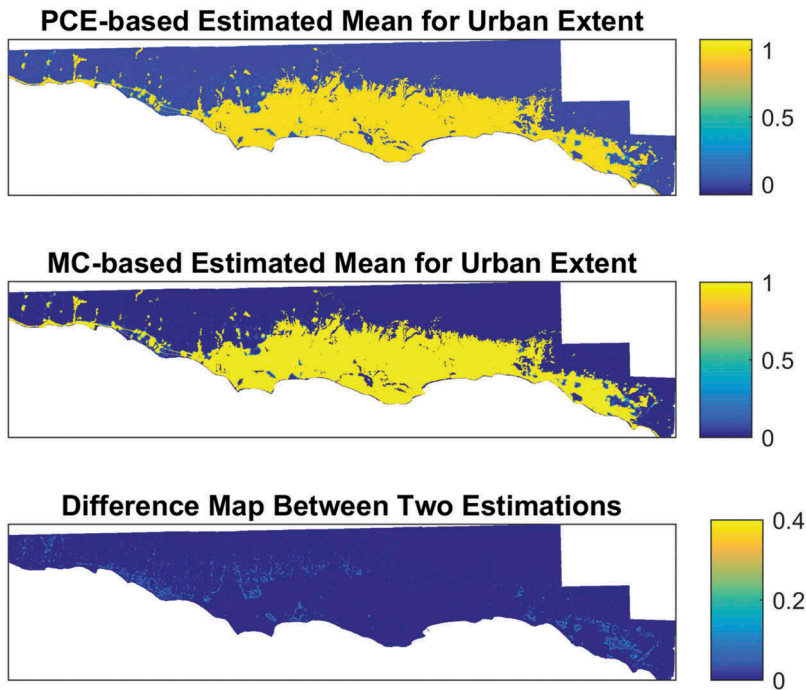

Figure 4. The estimated means for the probability of a cell to change into urban land use for the year 2008: PCE-based Estimated Mean (top), MC-based Estimated Mean (middle), difference between the PCE-based Estimated and MC-based Estimated Mean Surfaces (bottom).

Figure 5. The estimated means for the probability of a cell to change into urban land use for the year 2016: PCE-based Estimated Mean (top), MC-based Estimated Mean (middle), difference between the PCE-based Estimated and MC-based Estimated Mean Surfaces (bottom).

mean surfaces, calculated from only 112 model simulations, is a very close approximation to the mean of the 1000 simulations. As can be seen in the difference maps in Figure 3 (bottom) to 5 (bottom), the maximum value of the difference between the two approaches is 0.4, which is only observable for less than 10% of the urban pixels. Still, the accuracy of the estimated mean can be enhanced by experimenting with different p values for PCE approximation. Increasing p values will require a higher number of model simulations, which results in longer computational time. There is a trade-off between the computational time required to execute 1000 simulations and the desired accuracy of PCE-based approximation of mean and standard deviation, which can be assessed.

The comparison can be extended by plotting the model's forecast accuracy over the simulation time interval. For this comparison, the difference between the mean computed with 1000 Monte Carlo samples and the mean obtained with 112 PCE samples was calculated and plotted against the simulation time interval (Figure 6). The graph shows that the difference between the means is converging throughout the forecast period and ranges from 0 to 0.2. To better understand this difference, one can examine the distribution of pixels comprising the study area. The number of pixels with the observed difference greater than 0.12 is less than 1000 and for the overall urban area, this difference of 1.3% can be considered negligibly small.

Another interesting observation can be drawn from the decreasing pattern of differences. Neither the parameters nor the procedure to calculate the PCE or MC measures differ as the simulation progresses. As expected, the differences occur in places where
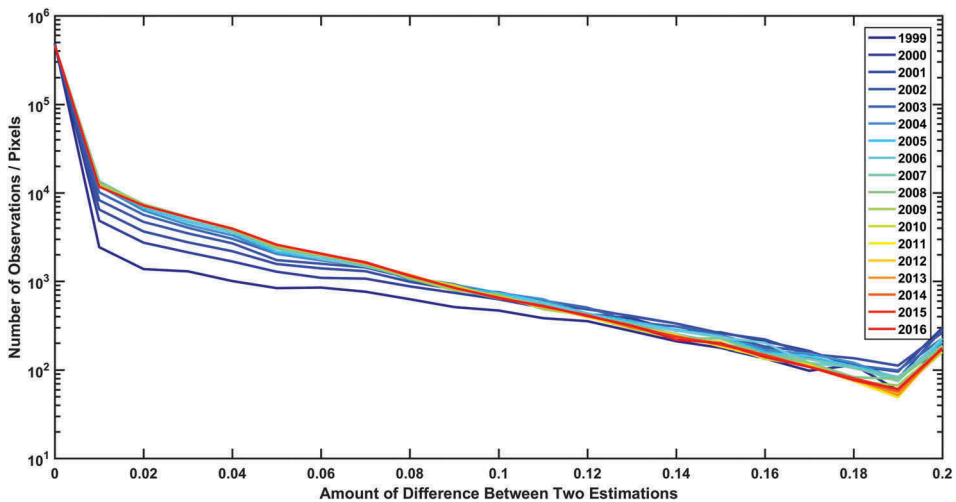
Figure 6. The comparison of difference between PCE-based estimated mean and MC-based mean.

SLEUTH forecasts a high probability of a cell changing its state to urban. As discussed in the methodology section, PCE estimation is based on the produced output to approximate the model. Therefore, the differences between MC and PCE estimates not only depend on the characteristics of the methods but also on the model output as well.

The estimated mean maps are helpful to locate the regions where urban growth is expected to occur. They summarize the forecasting results of all simulations based on the calculated mean for each pixel. Similar to the mean maps, the simulation years 1999, 2008 and 2016 were selected to depict the uncertainty (represented by standard deviation from the mean) resulting from transition parameters (Figures 7–9). The uncertainty maps, showing a similar pattern, can be further investigated in concert with the mean maps.

In order to categorize the study area based on high/low probability of change to urban land use and high/low uncertainty (standard deviation) of the change, a categorical map can be compiled to show four distinct categories: (1) high probability of changing into the urban class with high uncertainty (HH), (2) high probability of changing into the urban class with low uncertainty (HL), (3) low probability of changing into the urban class with high uncertainty (LH), and (4) low probability of changing into the urban class with low uncertainty (LL) (Ligmann-Zielinska and Jankowski 2014). The thresholds for high/low probability and high/low uncertainty can be selected by examining the logarithmically scaled histograms of mean probability for land-use change (Figure 10) and standard deviation (Figure 11) for years 1999, 2008, and 2016. A threshold value of 0.5 was selected for the mean probability (Figure 10), and the value of 0.5 was selected for the standard deviation (Figure 11). The selection is based on the distributions of mean urban land-use change probability and standard deviation of the mean urban land-use change probability; however, the ultimate selection of thresholds is an arbitrary choice (reflecting the mean value of the frequency distribution) made by the analyst.

The regions (pixels) falling into the 2nd category (HL) are considered as reliable predictions (low uncertainty) of urban land-use change. The regions falling into the
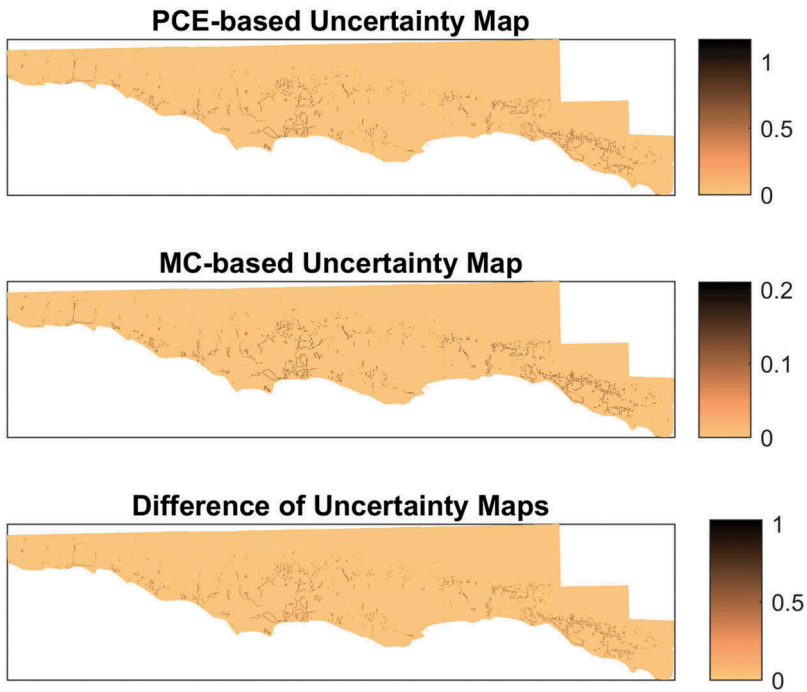
Figure 7. The standard deviations of the forecasted urban land-use change for the year 1999: PCE-based Uncertainty (top), MC-based Uncertainty (middle), difference between the PCE-based and MC-based Uncertainty Surfaces (bottom).
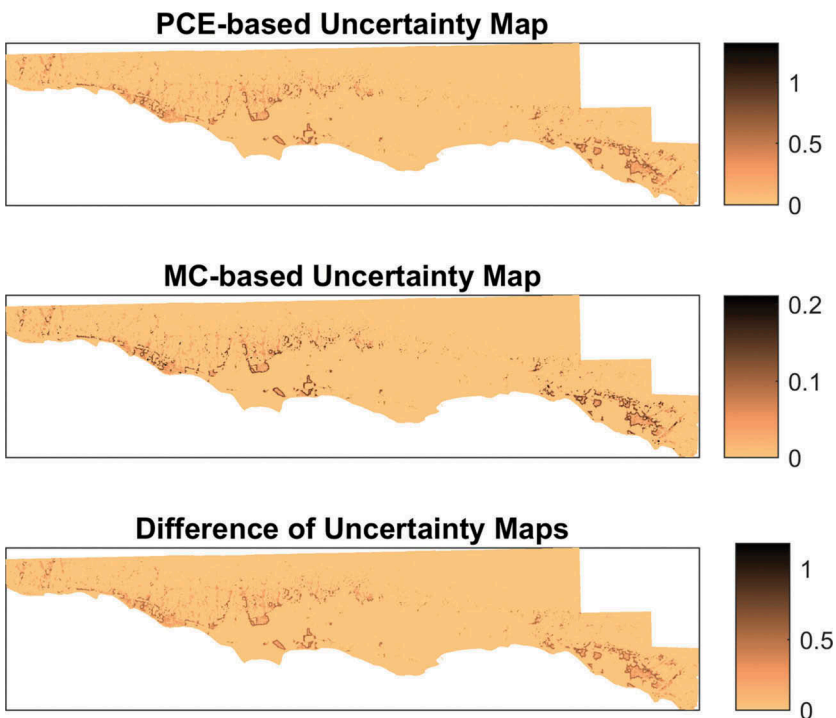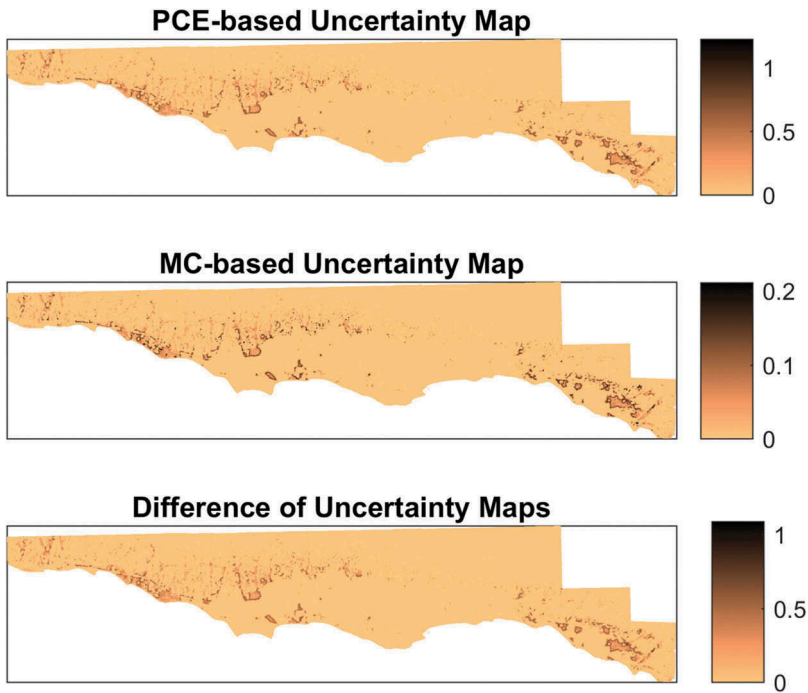

Figure 8. The standard deviations of the forecasted urban land-use change for the year 2008: PCE-based Uncertainty (top), MC-based Uncertainty (middle), difference between the PCE-based and MC-based Uncertainty Surfaces (bottom).

Figure 9. The standard deviations of the forecasted urban land use for the year 2016: PCE-based Uncertainty (top), MC-based Uncertainty (middle), difference between the PCE-based and MC-based Uncertainty Surfaces (bottom).
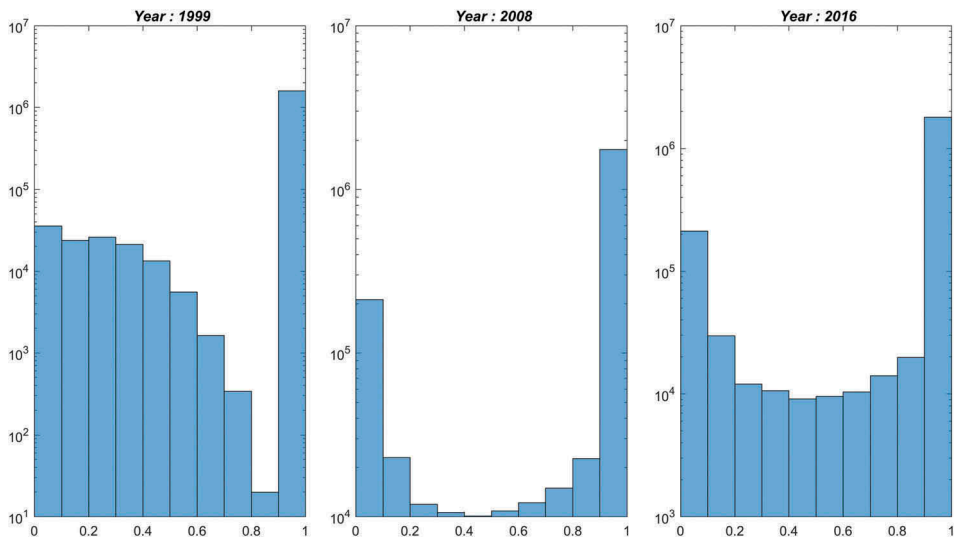

Figure 10. Mean land-use change probabilities for years 1999, 2008 and 2016.

3rd and 4th categories can be discarded due to low confidence in the result (LH) and low probability of land-use change (LL). For SA, the analysis continues by investigating the regions in the 1st category (HH) to establish how the relatively high uncertainty in these regions is apportioned to the transition parameters.
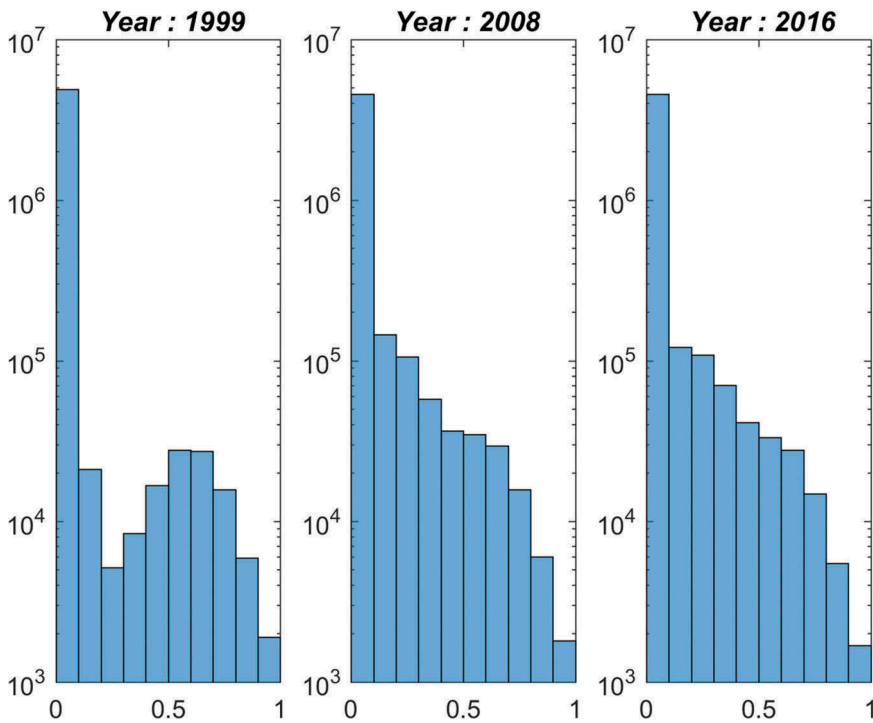
Figure 11. Standard deviation of mean land-use change probability for years 1999, 2008 and 2016.

After applying the threshold values, categorical maps were produced for three selected simulation years (Figure 12). The clusters of the high probability of land-use change accompanied by high uncertainty are depicted in red and these zones are particularly important when evaluating the results of SA due to the first and total-order interactions of transition parameters. An area in the lower-right of the map (Figure 12) where the HH category seems clustered, was selected for further investigation with SA.

Sensitivity analysis

Sobol first and total-order sensitivity indices were computed for the five forecast parameters for each simulation year resulting in 90 maps (5 parameters * 18 years) per each sensitivity index. To reduce the cognitive load of visualizing 90 maps, each sensitivity index is represented by its mean value with the confidence intervals and plotted for the simulation period for each analyzed parameter (Figure 13). After having compared these indices, one can return the produced spatial sensitivity maps for each year for each parameter to see spatial patterns of sensitivities. For different forecast years, the effect of each input parameter on the model's output varies. The effect of breed, slope and road parameters is relatively small (the model is insensitive to these parameters when taken singly or in isolation from one another) during the simulation time interval. This cannot be said about diffusion and spread parameters, which show much larger and oscillating influence on the model's output. Moreover, one can observe that during the forecast
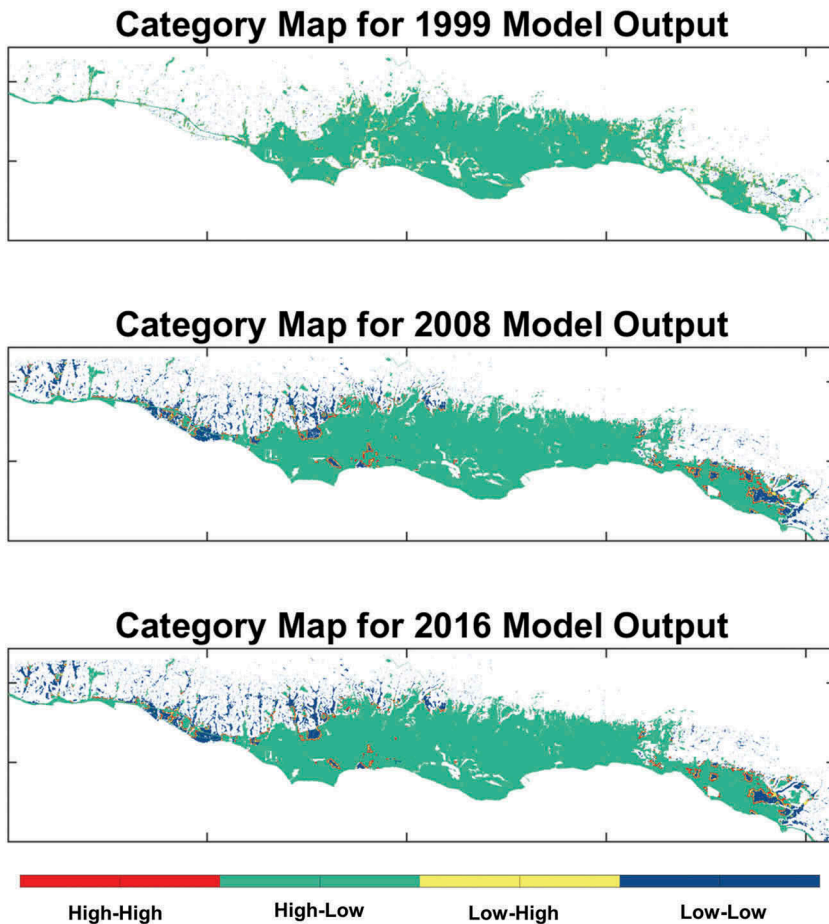
Category Map for 1999 Model Output

Category Map for 2008 Model Output

Category Map for 2016 Model Output

High-High   High-Low   Low-High   Low-Low

Figure 12. Categorical maps based on the probability of land-use change (High/Low) with the associated uncertainty (High/Low) for years 1999, 2008, and 2016.

period, most of the parameters show a steady dispersion after 2010. For the simulation period before 2010, diffusion and spread show a decreasing pattern, whereas breed, slope and road show an increasing pattern.

To visualize the interaction among forecast parameters and interaction dynamics, higher-order interactions, calculated as the difference between total-order indices and first-order indices, were plotted (Figure 14). The relatively high values represent high interactions among the parameters. For the interval 1999 – 2010, the higher-order interactions follow an increasing trend for diffusion, slope and road parameters and a decreasing trend for the breed and spread parameters. After 2010, there is no change in the level of interactions among the parameters. These graphs show that the sensitivity of slope and road parameters increases as the simulations progress for both SA measures. However, for diffusion, breed and spread, the sensitivity dynamics change between no interaction and interaction measures.

In order to understand the spatial pattern of parameter sensitivity, we also examined the sensitivity maps for a region located in the lower-left of the study area, where the
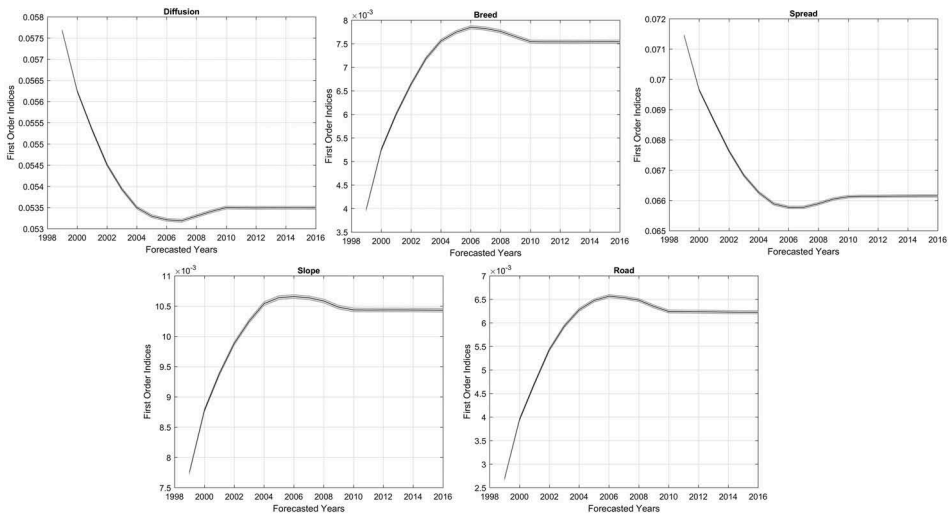
Figure 13. PCE-based Sobol first-order indices for years 1999 to 2016.
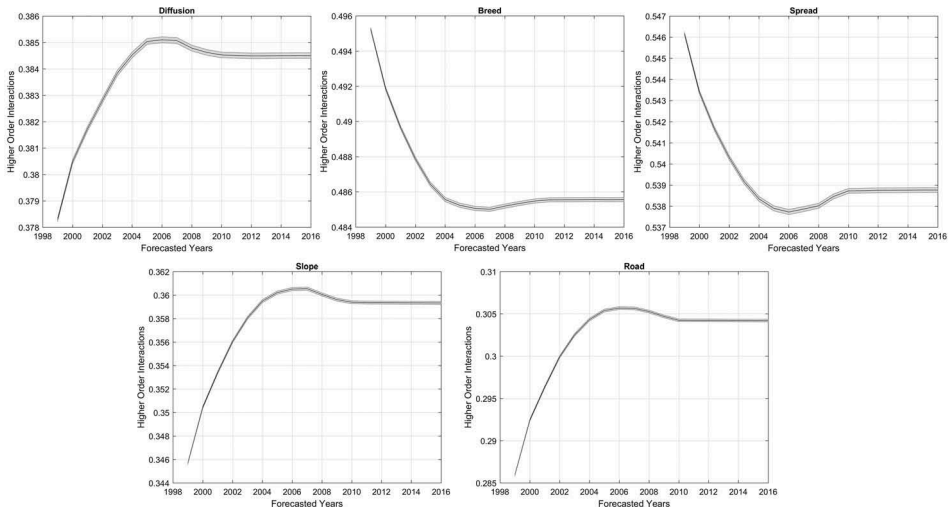

Figure 14. PCE-based Sobol higher-order indices (ST-S) for years 1999 to 2016.

model prediction has high uncertainty. The reason of this selection is the HH category pixel cluster for this region. We mapped the dominant sensitivity indices for years 2005, 2007 and 2010 where the forecast parameters change behavior (Figure 15) (Ligmann-Zielinska and Jankowski 2014). Out of five parameters, spread has the major influence on the output variability when the first-order indices are considered. For higher-order indices, diffusion dominates the other parameters in influencing output variability. Spread is a parameter driving edge growth and diffusion is effective during spontaneous growth and road influenced growth. One possible explanation for the dominance of spread and diffusion is a cyclical pattern of urbanization driven by edge and road influenced growth. The road parameter, associated with road influenced growth, is the
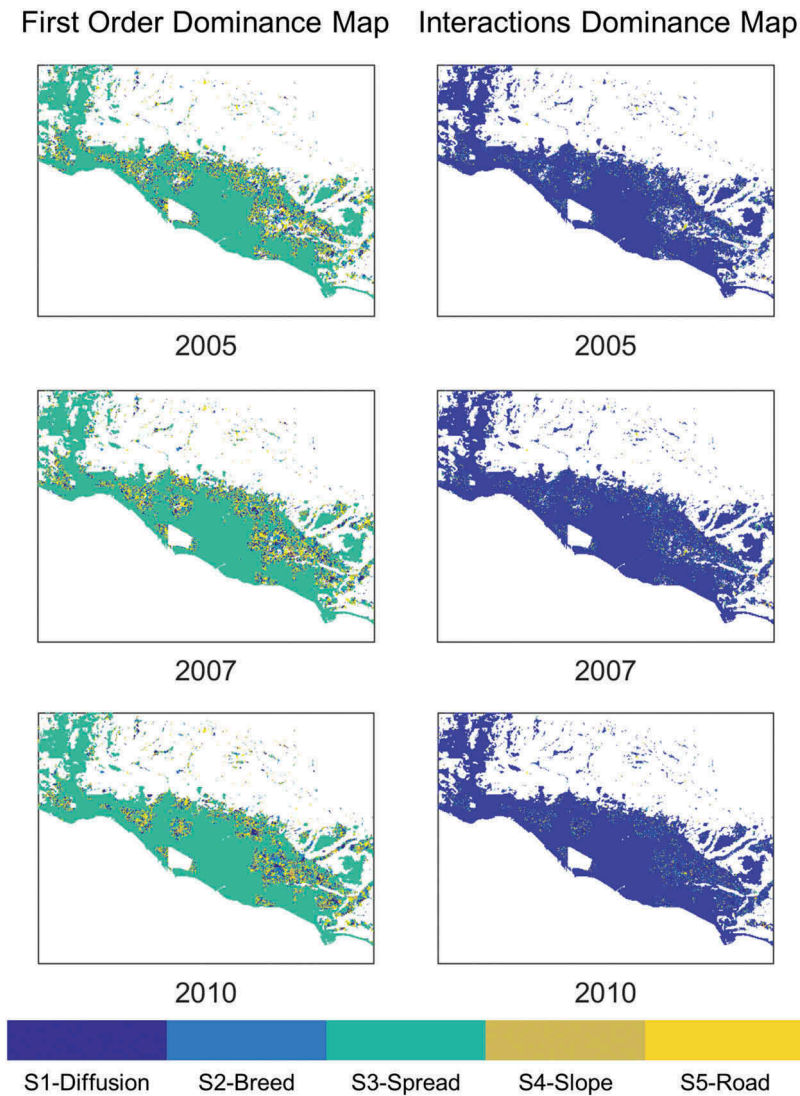
First Order Dominance Map   Interactions Dominance Map

Figure 15. First-order and higher-order dominance maps for 2005, 2007 and 2010.

second single (no interaction) most influential model parameter after spread (see the First-Order Dominance Maps in Figure 15).

Following the examination of the sensitivity maps, the contributions of the model parameters to the model output variance can be considered during model calibration, especially considering small perturbations in the value range of input parameters resulting in different levels of model output variance. For the HH zones (high probability of land-use change and high uncertainty – Figure 12), to increase the confidence in the model output, attention must be paid to the diffusion and spread forecast parameters since they have shown the most variability within the simulation time interval.

Limitations and future work

As discussed in the methodology section, PCE is a model approximation based on design points. Compared with an analytical solution, which includes every high-order interaction, PCE approximation only takes into account the subset of interactions defined by the design points. Therefore, the decimal accuracy of the indices may be lower than that for the full-order-based variance decomposition. However, given the results of tests for various models (Sudret 2008), PCE can be a good guide to prioritize input factors with smaller computational cost compared to full-order variance decomposition. For a higher level of accuracy, the experimental design can be reset with higher p values, resulting in higher-order polynomials in the expansion.

In addition to sensitivity index accuracy, another limitation of the work presented is the visualization of U-SA outputs for spatio-temporal models. Future research efforts should focus on effective means of visualizing parameter sensitivity both in space and time including the capability of examining spatial patterns for any time period. Additional topics for future research include:

- Further parallelization of the integrated U-SA approach to reduce computation cost for larger datasets.
  Applying the PCE-based Sobol decomposition to a different type of CA neighborhood (Von Neumann) to investigate whether there is a neighborhood effect on the model forecasts.
- Changing the experimental design size (N) and polynomial degree (p) to test the effect of different N and p values on the accuracy of the SA results.
- Comparing the output sensitivity indices of PCE-based Sobol decomposition with Quasi-Monte Carlo-based Sobol decomposition to examine the convergence of analysis results obtained with these two model variance decomposition methods.
- Examining in detail the regions with the most and least sensitivity and uncertainty, to test for and explain spatial autocorrelation in the modeling results.

Conclusions

SA, especially in its global approach, has received much attention in the research community over the past decade. However, the main challenge lies in the implementation stage due to the increasing amount of data and model complexity. These issues also highlight the problem of computational cost, which is especially high for spatially explicit, integrated U-SA for dynamic models. The proposed meta-modeling approach based on PCE-enabled model variance decomposition is an attractive alternative to conducting U-SA for nonlinear, non-monotonic, and high-order spatio-temporal models. The PCE implementation presented in the paper is also valuable in terms of its applicability to any spatio-temporal model due to its model-independent workflow. For high dimensional models with potentially interacting input parameters that can influence the overall model variability, in particular, PCE can be applied as an initial screening approach to find the most influential variables. This enables further experiments with a smaller subset of model parameters.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Seda Şalap-Ayça 🆔 http://orcid.org/0000-0002-5788-9563
Keith C Clarke 🆔 http://orcid.org/0000-0001-5805-6056

References

Abily, M., *et al*., 2016. Spatial global sensitivity analysis of high resolution classified topographic data use in 2D urban flood modelling. *Environmental Modelling and Software*, 77, 183–195. doi:10.1016/j.envsoft.2015.12.002

Armstrong, M. and Matheron, G., 1986. Isofactorial models for granulodensimetric data. *Mathematical Geology*, 18 (8), 743–757. doi:10.1007/BF00899741

Bilotta, G., *et al*., 2012. Sensitivity analysis of the MAGFLOW cellular automaton model for lava flow simulation. *Environmental Modelling & Software*, 35, 122–131. doi:10.1016/j.envsoft.2012.02.015

Burnaev, E., Panin, I., and Sudret, B., 2017. Efficient design of experiments for sensitivity analysis based on polynomial chaos expansions. *Annals of Mathematics and Artificial Intelligence*, 81 (1–2), 187–207. doi:10.1007/s10472-017-9542-1

Candau, J., Rasmussen, S., and Clarke, K.C., 2000. A coupled cellular automaton model for land use/land cover dynamics. In *4th International Conference on Integrating GIS and Environmental Modeling*. Banff, Alberta.Canada.

Chaudhuri, G. and Clarke, K.C., 2013. The SLEUTH land use change model: a review. *The International Journal of Environmental Resources Research*, 1 (1), 88–104.

Chen, Q. and Mynett, A.E., 2003. Effects of cell size and configuration in cellular automata based prey-predator modelling. *Simulation Modelling Practice and Theory*, 11 (7–8), 609–625. doi:10.1016/j.simpat.2003.08.006

Chen, Y., Yu, J., and Khan, S., 2010. Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. *Environmental Modelling and Software*, 25 (12), 1582–1591. doi:10.1016/j.envsoft.2010.06.001

Chiles, J.P. and Delfiner, P., 1999. *Geostatistics: modeling spatial uncertainty*. New York: Wiley.

Clarke, K.C., *et al*., 2007. A decade of SLEUTHing: lessons learned from applications of a cellular automaton land use change model. In: P. Fisher, ed. *Classics in IJGIS Twenty Years of the*

International Journal of Geographical Information Science and Systems. Boca Raton, FL: Taylor and Francis, CRC, 413–425.

Clarke, K.C., 2008. A decade of cellular urban modeling with SLEUTH: unresolved issues and problems. In: R.K. Brail, ed. *Planning support systems for cities and regions*. Cambridge, MA: Lincoln Institute of Land Policy, Chapter 3, 47–60.

Clarke, K.C., Gaydos, L., and Hoppen, S., 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco bay area. *Environment and Planning B*, 24, 247–261. doi:10.1068/b240247

Couclelis, H., 1985. Cellular worlds: a framework for modeling micro - macro dynamics. *Environment and Planning A*, 17 (5), 585–596. doi:10.1068/a170585

Crestaux, T., Le Maître, O., and Martinez, J.M., 2009. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering and System Safety*, 94 (7), 1161–1172. doi:10.1016/j.ress.2008.10.008

Crosetto, M. and Tarantola, S., 2001. Uncertainty and sensitivity analysis: tools for GIS-based model implementation. *International Journal of Geographical Information Science*, 15 (5), 415–437. doi:10.1080/13658810110053125

Crosetto, M., Tarantola, S., and Saltelli, A., 2000. Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, Ecosystems & Environment*, 81 (1), 71–79. doi:10.1016/S0167-8809(00)00169-9

Cukier, R.I., *et al*. 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: I theory. *The Journal of Chemical Physics*, 59 (8), 3873–3878. doi:10.1063/1.1680571

Dahal, K.R. and Chow, T.E., 2015. Characterization of neighborhood sensitivity of an irregular cellular automata model of urban growth. *International Journal of Geographical Information Science*, 29 (3), 475–497. doi:10.1080/13658816.2014.987779

De Almeida, C.M., *et al*. 2003. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. *Computers, Environment and Urban Systems*, 27 (5), 481–509. doi:10.1016/S0198-9715(02)00042-X

Deman, G., *et al*., 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, 156–169. doi:10.1016/j.ress.2015.11.005

Erlacher, C., *et al*. 2017. A GPU-based parallelization approach to conduct spatially-explicit uncertainty and sensitivity analysis in the application domain of landscape assessment. *GI_Forum*, (1), 44–58. doi:10.1553/giscience2017_01_s44

Gómez-Delgado, M. and Tarantola, S., 2006. GLOBAL sensitivity analysis, GIS and multi-criteria evaluation for a sustainable planning of a hazardous waste disposal site in Spain. *International Journal of Geographical Information Science*, 20 (4), 449–466. doi:10.1080/13658810600607709

Halls, J.N., 2002. A spatial sensitivity analysis of land use characteristics and phosphorus levels in small tidal creek estuaries of North Carolina, USA. *Journal of Coastal Research*, Special Issue (36), 340–351.

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), 327–367. doi:10.1016/0951-8320(93)90097-I

Helton, J.C. and Davis, F.J., 2002. Illustration of sampling based methods for uncertainty and sensitivity analysis. *Risk Analysis*, 22 (3), 591–622. doi:10.1111/risk.2002.22.issue-3

Herman, J.D., *et al*. 2013. From maps to movies: high-resolution time-varying sensitivity analysis for spatially distributed watershed models. *Hydrology and Earth System Sciences*, 17 (12), 5109–5125. doi:10.5194/hess-17-5109-2013

Heuvelink, G.B.M., 1989. Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information Systems*, 3 (4), 303–322. doi:10.1080/02693798908941518

Heuvelink, G.B.M., 1998. Uncertainty analysis in environmental modelling under a change of spatial scale. *Nutrient Cycling in Agroecosystems*, 50, 255–264. doi:10.1023/A:1009700614041

Heuvelink, G.B.M., 2003. Analysis uncertainty propagation in GIS: why is it not that simple?. In: G.M. Foody and P.M. Atkinson, eds. *Uncertainty in remote sensing and GIS*. West Sussex, England: John Wiley & Sons, Ltd, 155–165.

Heuvelink, G.B.M. and Burrough, P.A., 1993. Error propagation in cartographic modelling using Boolean logic and continuous classification. *International Journal of Geographical Information Systems*, 7 (3), 231–246. doi:10.1080/02693799308901954

Heuvelink, G.B.M. and Burrough, P.A., 2002. Developments in statistical approaches to spatial uncertainty and its propagation. *International Journal of Geographical Information Science*, 16 (2), 111–113. doi:10.1080/13658810110099071

Holland, J., 1999. *Emergence : from chaos to order*. Cambridge, MA: Perseus Book.

Hu, Y., *et al.* 2015. Global sensitivity analysis for large-scale socio-hydrological models using Hadoop. *Environmental Modelling & Software*, 73 (2015), 231–243. doi:10.1016/j. envsoft.2015.08.015

Jantz, C.A. and Goetz, S.J., 2005. Analysis of scale dependencies in an urban land-use-change model. *International Journal of Geographical Information Science*, 19 (2), 217–241. doi:10.1080/ 13658810410001713425

Jantz, C.A., Goetz, S.J., and Shelley, M.K., 2003. Using the SLEUTH urban growth model to simulate the impacts of future policy scenarios on urban land use in the Baltimore-Washington metropolitan area. *Environment and Planning B: Planning and Design*, 31 (2), 251–271. doi:10.1068/b2983

Kim, J.H., 2013. Spatiotemporal scale dependency and other sensitivities in dynamic land-use change simulations. *International Journal of Geographical Information Science*, 27 (9), 1782–1803. doi:10.1080/13658816.2013.787145

Kocabas, V. and Dragićević, S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban Systems*, 30 (6), 921–953. doi:10.1016/j.compenvurbsys.2006.01.001

Kyriakidis, P.C. and Goodchild, M.F., 2006. On the prediction error variance of three common spatial interpolation schemes. *International Journal of Geographical Information Science*, 20 (8), 823–855. doi:10.1080/13658810600711279

Lamboni, M., *et al.* 2009. Multivariate global sensitivity analysis for dynamic crop models. *Field Crops Research*, 113 (3), 312–320. doi:10.1016/j.fcr.2009.06.007

Li, X., Liu, X., and Yu, L., 2014. A systematic sensitivity analysis of constrained cellular automata model for urban growth simulation based on different transition rules. *International Journal of Geographical Information Science*, 28 (7), 1317–1335. doi:10.1080/13658816.2014.883079

Ligmann-Zielinska, A., 2013. Spatially-explicit sensitivity analysis of an agent-based model of land use change. *International Journal of Geographical Information Science*, 27 (9), 1764–1781. doi:10.1080/13658816.2013.782613

Ligmann-Zielinska, A. and Jankowski, P., 2014. Spatially-explicit integrated uncertainty and sensitivity analysis of criteria weights in multicriteria land suitability evaluation. *Environmental Modelling and Software*, 57, 235–247. doi:10.1016/j.envsoft.2014.03.007

Lilburne, L. and Tarantola, S., 2009. Sensitivity analysis of spatial models. *International Journal of Geographical Information Science*, 23 (February 2015), 151–168. doi:10.1080/ 13658810802094995

Marrel, A., *et al.* 2011. Global sensitivity analysis for models with spatially dependent outputs. *Environmetrics*, 22 (3), 383–397. doi:10.1002/env.1071

Ménard, A. and Marceau, D.J., 2005. Exploration of spatial scale sensitivity in geographic cellular automata. *Environment and Planning B: Planning and Design*, 32 (5), 693–714. doi:10.1068/ b31163

Metropolis, N. and Ulam, S., 1949. The Monte Carlo method. *Journal of American Statistical Association*, 44 (247), 335–341. doi:10.1080/01621459.1949.10483310

Moreau, P., *et al.*, 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 and Software*, 47, 74–87. doi:10.1016/j.envsoft.2013.04.006

Moreno, N., Wang, F., and Marceau, D., 2008. An object-based land-use cellular automata model to overcome cell size and neighborhood sensitivity. Proceedings of GEOBIA 2008–pixels, objects, intelligence GEOgraphic object based image analysis for the 21st century, 5–8 August, Calgary. Available from: http://www.isprs.org/proceedings/XXXVIII/4-C1/Sessions/Session6/6753_ Marceau_Proc_pap.pdf [Accessed 2 September 2016].

Najm, H.N., 2009. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annual Review of Fluid Mechanics*, 41 (1), 35–52. doi:10.1146/annurev.fluid.010908.165248

NCGIA. SLEUTH - Project Gigapolis. Available at: http://www.ncgia.ucsb.edu/projects/gig/ [Accessed 2 September 2016].

O'Hagan, A., 2011. Polynomial chaos: a tutorial and critique from a statistician's perspective. *SIAM/ASA J. Uncertainty Quantification*, 20 (2013), 1–20.

Oakley, J. and Hagan, A.O., 2000. Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89 (4), 769–784. doi:10.1093/biomet/89.4.769

Pan, Y., *et al*. 2010. The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems*, 34 (5), 400–408. doi:10.1016/j.compenvurbsys.2010.03.003

Pontius, R.G. and Neeti, N., 2010. Uncertainty in the difference between maps of future land change scenarios. *Sustainability Science*, 5 (1), 39–50. doi:10.1007/s11625-009-0095-z

Rajabi, M.M., Ataie-Ashtiani, B., and Simmons, C.T., 2015. Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations. *Journal of Hydrology*, 520, 101–122. doi:10.1016/j.jhydrol.2014.11.020

Ratto, M. and Pagano, A., 2012. State dependent regressions: from sensitivity analysis to meta-modeling. In L. Wang, H. Garnier, eds. *Environmental modeling, and control system design*. London: Springer. 171–190.

Roura-Pascual, N., *et al*. 2010. Spatially-explicit sensitivity analysis for conservation management: exploring the influence of decisions in invasive alien plant management. *Diversity and Distributions*, 16 (3), 426–438. doi:10.1111/j.1472-4642.2010.00659.x

Rutherford, M.J., Crowther, M.J., and Lambert, P.C., 2015. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation*, 85 (4), 777–793. Available at. doi:10.1080/00949655.2013.845890

Saint-Geours, N., *et al*., 2014. Multi-scale spatial sensitivity analysis of a model for economic appraisal of flood risk management policies. *Environmental Modelling & Software*, 60, 153–166. doi:10.1016/j.envsoft.2014.06.012

Şalap-Ayça, S. and Jankowski, P., 2016. Integrating local multi-criteria evaluation with spatially explicit uncertainty-sensitivity analysis. *Spatial Cognition & Computation*, 16 (2), 106–132. doi:10.1080/13875868.2015.1137578

Saltelli, A., *et al*. 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181 (2), 259–270. doi:10.1016/j.cpc.2009.09.018

Saltelli, A., Chan, K., and Scott, E.M., 2000. *Sensitivity analysis*. New York: Wiley.

Saltelli, A., Tarantola, S., and Chan, K.P.S., 1999. A quantiative model-independent method for global sensitivity analysis of model output. *Techniques*, 41 (1), 39–56. doi:10.1080/00401706.1999.10485594

Samat, N., 2006. Characterizing the scale sensitivity of the cellular automata simulated urban growth: a case study of the Seberang Perai Region, Penang State, Malaysia. *Computers, Environment and Urban Systems*, 30 (6), 905–920. doi:10.1016/j.compenvurbsys.2005.11.002

Sobol', I.M., 1993. Sensitivity analysis for non-linear mathematical models. *Mathematical Modeling & Computational Experiments*, 1, 407–414.

Sobol', I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55 (1–3), 271–280. doi:10.1016/S0378-4754(00)00270-6

Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93 (7), 964–979. doi:10.1016/j.ress.2007.04.002

Sudret, B., 2014. Polynomial chaos expansions and stochastic finite element methods. In: K. Phoon and J. Ching, eds. *Risk and reliability in geotechnical engineering*. Boca Raton, FL: CRC Press, 265–300.

Tang, W. and Jia, M., 2014. Global sensitivity analysis of a large agent-based model of spatial opinion exchange: a heterogeneous multi-GPU acceleration approach. *Annals of the Association of American Geographers*, 104 (3), 485–509. doi:10.1080/00045608.2014.892342

Tobler, W.R., 1979. Cellular geography. In: S. Gale and G. Olsson, eds. *Philosophy in Geography*. Dortrecht, NL: Dortrecht Reidel, 379–386.

Torrens, P.M. and Nara, A., 2007. Modeling gentrification dynamics: a hybrid approach. *Computers, Environment and Urban Systems*, 31 (3), 337–361. doi:10.1016/j.compenvurbsys.2006.07.004

Torrens, P.M. and O'Sullivan, D., 2000. Cities, cells, and complexity: developing a research agenda for urban geocomputation. In *5th international conference on GeoComputation*. University of Greenwich, UK, pp. 1–14.

Ulam, S., 1952. Random processes and transformations. *Proceedings of the International Congress of Mathematics*, 2, 264–275.

van Vliet, J., *et al.* 2016. A review of current calibration and validation practises in land-change modeling. *Environmental Modelling and Software*, (82), 174–182. doi:10.1016/j.envsoft.2016.04.017

Villa-Vialaneix, N., *et al.* 2012. A comparison of eight metamodeling techniques for the simulation of N 2O fluxes and N leaching from corn crops. *Environmental Modelling and Software*, 34 (August 2017), 51–66. doi:10.1016/j.envsoft.2011.05.003

Von Neumann, J., 1966. In: A.W. Burks, ed. *Theory of self-reproducing automata*. Champaign, IL: University of Illinois Press.

Wackernagel, H., 2003. *Multivariate geostatistics: an introduction with applications*. New York: Springer Berlin Heidelberg.

White, R. and Engelen, G., 1993. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A*, 25 (8), 1175–1199. doi:10.1068/a251175

White, R. and Engelen, G., 1994. Urban systems dynamics and cellular automata: fractal structures between order and chaos. *Chaos, Solitons and Fractals*, 4 (4), 563–583. doi:10.1016/0960-0779(94)90066-3

White, R., Engelen, G., and Uljee, I., 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Urban Analytics and City Science*, 24 (3), 323–343. doi:10.1068/b240323

Wiener, N., 1938. The homogenous chaos. *American Journal of Mathematics*, 60 (4), 897–936. doi:10.2307/2371268

Xiu, D., *et al.* 2002. Stochastic modeling of flow-structure interactions using generalized polynomial chaos. *Journal of Fluids Engineering*, 124 (1), 51. doi:10.1115/1.1436089

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), 619–644.

Xu, E. and Zhang, H., 2013. Spatially-explicit sensitivity analysis for land suitability evaluation. *Applied Geography*, 45, 1–9. doi:10.1016/j.apgeog.2013.08.005

Yeh, A.G.-O. and Li, X., 2006. Error propagation and model uncertainties of cellular automata in urban simulation with GIS. *Computers, Environment and Urban Systems*, 30 (1), 10–28. doi:10.1016/j.compenvurbsys.2004.05.007