GLOBAL SENSITIVITY AND UNCERTAINTY ANALYSIS OF SPATIALLY DISTRIBUTED WATERSHED MODELS

By

ZUZANNA B. ZAJAC

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

To Król Korzu
KKMS!

ACKNOWLEDGMENTS

I would like to thank my advisor Rafael Muñoz-Carpena for his constant support and encouragement over the past five years. I could not have achieved this goal without his patience, guidance, and persistent motivation. For providing innumerable helpful comments and helping to guide this research, I also thank my graduate committee co-chair Wendy Graham and all the members of the graduate committee: Michael Binford, Greg Kiker, Jayantha Obeysekera, and Karl Vanderlinden. I would also like to thank Naiming Wang from the South Florida Water Management District (SFWMD) for his help understanding the Regional Simulation Model (RSM), the great University of Florida (UF) High Performance Computing (HPC) Center team for help with installing RSM, South Florida Water Management District and University of Florida Water Resources Research Center (WRRC) for sponsoring this project.

Special thanks to Lukasz Ziemba for his help writing scripts and for his great, invaluable support during this PhD journey. To all my friends in the Agricultural and Biological Engineering Department at UF: thank you for making this department the greatest work environment ever. Last, but not least, I would like to thank my father for his courage and the power of his mind, my mother for the power of her heart, and my brother for always being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AHF	Airborne Height Finder
CCDF	Conditional cumulative distribution function
CDF	Cumulative distribution function
CI	Confidence interval
DEM	Digital elevation model
EAA	Everglades Agricultural Area
EPA	Everglades Protection Area
ET	Evapotranspiration
FAST	Fourier amplitude sensitivity test
FOSM	First-order second-moment
GSA	Global sensitivity analysis
GUA	Global uncertainty analysis
GUA/SA	Global uncertainty and sensitivity analysis
HSE	Hydrologic Simulation Engine
IFSAR	Interferometric Synthetic Aperture Radar
IK	Indicator Kriging
LiDAR	Light Detection and Ranging
MC	Monte Carlo
MSE	Management Simulation Engine
NSRSM	Natural Systems Regional Simulation Model
PDF	Probability distribution function
RF	Random function
RMSE	Root mean square error
RSM	Regional Simulation Model

RV	Random variable
SA	Sensitivity analysis
SGS	Sequential Gaussian simulation
SIS	Sequential indicator simulation
SK	Simple Kriging
SS	Sequential simulation
SVD	Singular value decomposition
UA	Uncertainty analysis
WCA-2A	Water Conservation Area-2A
XML	Extensible markup language

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

GLOBAL SENSITIVITY AND UNCERTAINTY ANALYSIS OF SPATIALLY DISTRIBUTED WATERSHED MODELS

By

Zuzanna Zajac

August 2010

Chair: Rafael Muñoz-Carpena
Cochair: Wendy Graham
Major: Agricultural and Biological Engineering

With spatially distributed models, the effect of spatial uncertainty of the model inputs is one of the least understood contributors to output uncertainty and can be a substantial source of errors that propagate through the model. The application of the global uncertainty and sensitivity (GUA/SA) methods for formal evaluation of models is still uncommon in spite of its importance. Even for the infrequent cases where the GUA/SA is performed for evaluation of a model application, the spatial uncertainty of model inputs is disregarded due to lack of appropriate tools. The main objective of this work is to evaluate the effect of spatial uncertainty of model inputs on the uncertainty of spatially distributed watershed models in the context of other input uncertainty sources. A new GUA/SA framework is proposed in this dissertation in order to incorporate the effect of spatially distributed numerical and categorical model inputs into the global uncertainty and sensitivity analysis (GUA/SA). The proposed framework combines the global, variance-based method of Sobol and geostatistical techniques of sequential simulation (SS). Sequential Gaussian simulation (SGS) is used for estimation of spatial uncertainty for numerical inputs (such as land elevation), while sequential indicator

simulation (SIS) is used for assessment of spatial uncertainty of categorical inputs (such as land cover type). The Regional Simulation Model (RSM) and its application to WCA-2A in the South Florida Everglades is used as a test bed of the framework developed in this dissertation. The RSM outputs chosen as metrics for GUA/SA for this study are key performance measures generally adopted in the Everglades restoration studies: hydroperiod, water depth amplitude, mean, minimum and maximum. The GUA/SA results for two types of outputs, domain-based (spatially averaged over domain) and benchmark cell-based, are compared. The benchmark cell-based outputs are characterized with larger uncertainty than their domain-based counterparts. The uncertainty of benchmark cell-based outputs is mainly controlled by land elevation uncertainty, while uncertainty of domain-based outputs it also attributed to factors like conveyance parameters. The results indicate that spatial uncertainty of model inputs is indeed an important source of model uncertainty.

The land cover distribution affects model outputs through delineation of Manning's roughness zones and evapotranspiration factors associated to the different vegetation classes. This study shows that in this application the spatial representation of land cover has much smaller influence on model uncertainty when compared to other sources of uncertainty like spatial representation of land elevation.

The spatial uncertainty of land cover was found to affect RSM domain-based model outputs through delineation of Manning's roughness zones more than through ET parameters effects.

The relationship between model uncertainty and alternative spatial data resolutions was studied to provide an illustration of how the procedure may be applied

for more informed decisions regarding planning of data collection campaigns. The results corroborate a proposed hypothetical nonlinear, negative relationship between model uncertainty and source data density. The inflection point in the curve, representing the optimal data requirements for the application, is identified for the data density between 1/4 and 1/8 of original data density. It is postulated that the inflection point is related to the characteristics of the spatial dataset (variogram) and the aggregation technique (model grid size).

The framework proposed in this dissertation could be applied to any spatially distributed model and input, as it is independent from model assumptions.

CHAPTER 1
INTRODUCTION

Uncertainty and Sensitivity Analysis

In the fields of water resources management and ecosystem restoration, the decision-making process is often supported by complex hydrological models. Model predictions are associated with uncertainties resulting from input data and parameter variability, model algorithms or structure, model calibration data, scale, model boundary conditions, etc. (Beven, 1989; Haan, 1989; Luis and McLaughlin, 1992; Shirmohammadi, 2006). Often, important management decisions are based on those simulations results. The uncertainty of the model results is often a major concern, since it has policy, regulatory, and management implications (Shirmohammadi et al., 2006). Scientific information feeds into the policy process, with a tendency by all parties involved to manipulate uncertainty. Uncertainty cannot be resolved into certainty in most instances. Instead, transparency must be offered by the global sensitivity analysis. Transparency is what is needed to ensure that the negotiating parties do not throw away science as a just another contentious input (Pascual, 2005). As stated by Beven (2006) if model uncertainty is not evaluated formally, the science and value of the model as a decision-supporting tool can be undermined. Formal uncertainty and sensitivity analysis (UA/SA) can increase confidence in model predictions by providing understanding of model behavior and by assessing model reliability in a decision making framework (Saltelli et al., 2004). Uncertainty analysis involves quantification of the uncertainties in the model input data and parameters and their propagation through the model to model outputs (predictions). The role of the sensitivity analysis (SA) is to apportion model output uncertainty into the model inputs.

UA/SA provides irreplaceable insight into model behavior and should be used not just at the outset but throughout model calibration and application as a part of an iterative process of model identification and refinement (Crosetto and Tarantola, 2001). Uncertainty and sensitivity analyses can be applied synergistically for the evaluation of complex computer models (Muñoz-Carpena et al., 2006; Saltelli et al., 2004). The formal application of UA allows the modeler to evaluate the performance and reliability of the model for specific application. SA, on the other hand, allows a better understanding of a model by identifying factors' contributions to output uncertainty.

However, in spite of their strengths, formal sensitivity and uncertainty analyses are often ignored in hydrological and water quality modeling efforts (Haan et al., 1995; Muñoz-Carpena et al., 2006; Shirmohammadi et al., 2006), usually due to the considerable effort these involve as the complexity and size of the models increase and also due to the limited data available specific to the model application (Reckhow, 1994).

Global Uncertainty and Sensitivity Analysis

Global UA/SA is based on Monte Carlo (MC) simulations, which involve random sampling of model input space (defined by probability distribution), model simulations for each set of input values, and the production of an empirical probability distribution for resulting model outputs. The MC approach requires that all inputs and outputs are scalar (i.e. singular) values so the uncertainty of a variable can be characterized by a probability distribution function (PDF). The term "input factor" is used to describe scalar random variables that are used to characterize uncertainty in input data and model parameters (Crosetto and Tarantola 2001), initial and boundary conditions, etc. This term is equivalent to a model input for spatially lumped inputs.

Probability distribution functions (PDFs) of model output, resulting from multiple model simulations, are used for deriving uncertainty measures, like confidence levels, or probability of exceedance of a threshold value (Morgan and Henrion, 1992). Global analysis has many advantages over local, derivative-based, one-parameter-at-a-time (OAT) approaches (Haan, 1995). Local sensitivity measures are typically fixed to a point (base value) where the derivative is taken. The choice of the base value from a factor's range may largely influence the SA results, especially in case of nonlinear, nonmonotonic models. The global analysis, on the other hand, explores the whole potential range of all the uncertain model input factors. Therefore it can be applied to any model, irrespective of model assumptions of linearity and monotonicity. Furthermore, the global analysis considers the effects of simultaneous variation of model inputs, allowing for evaluation of input factor interactions on model uncertainty. Most complex hydrological models are of non-linear, non-monotonic nature. In this case, local, OAT methods are of limited use, if not outright misleading, when the analysis aims to assess the relative importance of uncertain input factors (Saltelli et al., 2005).

The generation of samples from input factors' PDFs can be obtained using different sampling methods such as simple random brute-force sampling or more efficient, stratified sampling, such as replicated Latin hypercube sampling (r-LHS) (McKay et al., 2000; McKay, 1995), quasi random sequences (Sobol, 1993), Fourier Amplitude Sensitivity Test, FAST (Cukier et al., 1973), extended FAST (Saltelli et al., 1999), and random balance designs (Tarantola et al, 2006). Probability distributions of input factors can be constructed based on all available information derived from

available measurements, literature review, expert opinion, physical bounding consideration, or through parameter estimation in inverse problems, etc. (Cacuci, et al. 2005; Haan, 1989; Haan et al., 1995; Haan et al., 1998; Saltelli et al.2005). When no information on a factor's variability is available, it is often varied by +/-10 or 20% of the base value.

Different types of global sensitivity methods can be selected based on the objective of the analysis, the number of uncertain input factors, the degree of regularity of the model, and the computing time for a single model simulation (Cacuci et al., 2003; Saltelli et al.,2004; Saltelli et al. 2008; Wallach et al., 2006). The global sensitivity analysis (GSA) methods can be differentiated into screening methods (Campolongo et al., 2007; Morris, 1991), regression methods (Cacuci et al., 2003; Saltelli et al. 2000) and variance-based methods (Saltelli et al., 2004, Saltelli et al., 2008). Figure 1-1 presents various techniques available and their use as a function of computational cost of the model, complexity of the model, dimensionality of the input space. Variance-based methods provide robust quantitative results irrespectively of the models' behavior, but are computationally the most demanding. Regression methods, like standardized regression coefficients (SRC) are less expensive alternatives to the variance-based methods but are only suitable for linear or quasi-linear models (Saltelli et al., 2005). Screening methods, like the Morris method, are not computationally demanding but provide only qualitative measures of sensitivity. If a model is computationally expensive (CPU above 1 hour), the application of global techniques is not feasible and local techniques like automatic differentiation (AD) techniques need to be used.

The screening methods can be applied for initial, computationally cheap, qualitative sensitivity analysis (Saltelli et al. 2005). These methods are designed to determine, in terms of the relative effect on the model output, which of the model input factors can be considered negligible (i.e. with little contribution to model output uncertainty). The screening method proposed by Morris (1991), (hereafter the method of Morris) and later modified by Campolongo et al. (2005), is used in the current study for initial screening since it is relatively easy to implement, requires very few simulations, and interpreting its results is straightforward (Saltelli et al. 2005). In addition, Morris (1991) showed that the method could be applied with a large number of input factors.

Variance-based (or variance-decomposition) methods (also referred to as ANOVA-like methods) are based on the assumption that variance of the model output can be decomposed into fractions associated with input factors and their interactions. The decomposition of model output variance is presented by equation:

$$V(Y) = \sum_{i} V_i + \sum_{i<j} V_{ij} + \sum_{i<j<m} V_{ijm} + \ldots + V_{12\ldots k} \qquad (1\text{-}1)$$

where: V(Y) – total variance of model output Y, V_i - fraction of output variance explained by the i^{th} model input factor, V_{ij} - fraction of variance due to interactions between factors i and j, k – number of inputs.

For a given factor *i*, two sensitivity measures are calculated: first-order sensitivity index S_i – measuring a direct contribution of factor *i* to the total output variance, and total sensitivity index S_{Ti}, that contains sum of all effects involving a given factor (direct effects and effects due to interactions with other factors).

The first order sensitivity index S_i is calculated from the ratio of fraction of output variance explained by the i^{th} model input (V_i) to the total output unconditional variance (V):

$$S_i = \frac{V_i}{V(Y)} \tag{1-2}$$

It can be written in form of conditional variance as:

$$S_i = \frac{V\left(E\left[Y|X_i\right]\right)}{V\left(Y\right)} \tag{1-3}$$

where X_i - factor i.

Assuming the factors are independent, the total order sensitivity index S_{Ti} is calculated as the sum of the first order index and all higher order indices of a given parameter. For example, for parameter X_i:

$$S_{Ti} = 1 - \frac{V_{-i}}{V(Y)} \tag{1-4}$$

and

$$S_{Ti} = 1 - \frac{V\left(E\left[Y|X_{-i}\right]\right)}{V(Y)} \tag{1-5}$$

where: S_{Ti} – total order sensitivity, V_{-i} – the average variance that results from all parameters, except X_i.

For a given parameter, X_i, interactions with other factors can be isolated by calculating a remainder S_{Ti} - S_i Factors that have small S_i but large S_{Ti} primarily affect model output through interactions with other input factors.

The emphasis of the SA may be placed on calculating either first or total sensitivity indices. The choice of a measure depends on the purpose of the analysis, also referred

to as a SA setting (Saltelli et al., 2004). Factor prioritization setting is used when the purpose of SA is to obtain a ranking of parameter importance. For this setting it is important that the Type I error - false positive (i.e. the erroneous identification of a factor as influential when it is not) is avoided and use of first-order sensitivity indices is recommended (Saltelli, 2004). Factor fixing setting is used for identification of factors that, if fixed, would reduce the output variance the most. For this setting, Type II - false negative (i.e. failing in the identification of a factor of considerable influence on the model) error should be avoided and the suggested measures are total order indices.

This dissertation focuses on the variance-based methods for GUA/SA (Extended FAST, Sobol). Variance-based methods provide quantitative measures of the contribution to the output variance from uncertain factors individually or from interactions with other factors. Furthermore, this group of methods provides information not only about the direct (first order) effect of the individual factors over the output, but also about their interaction (higher order) effects. The variance-based methods involve high computational costs; therefore the screening methods may be applied in order to make the analysis more computationally efficient by focusing only on the subset of important factors obtained by the screening method.

The formal application of global uncertainty and sensitivity analysis allows the modeler to:

- examine model behavior,
- simplify the model,
- identify important input factors and interactions to guide the calibration of the model,
- identify input data or parameters that should be measured or estimated more accurately to reduce the uncertainty of the model outputs,

- identify optimal locations where additional data should be measured to reduce the uncertainty of the model, and
- quantify uncertainty of the modeling results (Saltelli et al., 2005).

Incorporating Spatial Heterogenity in Global Uncertainty and Sensitivity Analysis

Spatial heterogeneity is a natural feature of environmental systems. Application of spatially distributed environmental models, which aim to reproduce such spatial variability, has become more common due to the increased availability of spatial data and improved computational resources (Grayson and Blöschl, 2001). With spatially distributed models, the spatial uncertainty of input variables is a substantial source of errors that propagate through the model and affect the uncertainty of results (Phillips and Marks, 1996). The effect of spatial uncertainty of the model inputs is one of the least understood contributors to uncertainty of distributed models. Currently, UA/SA methods generally disregard the spatial context of model processes and the spatial uncertainty of model inputs.

Spatial uncertainty should be included in the evaluation of model quality for risk assessment to be realistic and effective (Rossi et al., 1993). Furthermore, a practical implication of including spatial uncertainty of model inputs results in a more effective resource allocation, since the collection of spatially distributed data is one of the most expensive parts of distributed modeling (Crosetto and Tarantola, 2001). Identification of spatially distributed factors contributing the most to model uncertainty enables elaboration of the most effective strategies for a reduction of model uncertainty.

The GUA/SA methodology has been applied primarily to lumped models, where all input factors are scalar and generated from scalar PDFs. In the case of spatially distributed input factors, alternative input maps (rather than alternative scalar values)

need to be generated and processed by the model. The application of UA to spatial models, using geostatistical techniques and MC simulations is straightforward and requires processing of alternative spatial realizations through the model (Phillips and Marks, 1996), and constructing output probability distributions to evaluate model uncertainty (Kyriakidis, 2001).

Uncertainty associated with spatial structure of input factors may affect model uncertainty and therefore influence model sensitivity. However, examples of the application of GSA techniques that account for spatial structure of input factors are rare and limited in scope (Crosetto et al., 2000, Crosetto end Tarantola, 2001; Francos et al. 2003, Hall et al., 2005; Tang et al., 2007a). GSA methods generally have limitations that make them unsuitable for evaluation of spatially distributed models (Lilburne and Tarantola, 2009). The shortcomings of GSA applied to distributed spatial models are related to impractical computational costs and the inability to realistically represent inputs' spatial structure. GSA methods based on the MC sampling require that inputs are represented by a scalar values. Medium-size watershed models (i.e., hundreds of hectares) may have hundreds or thousands of discretization units. If GSA is performed for all cells individually (each parameter value of each discretization unit treated as input factor) the computational cost of analysis for watershed models becomes impractical and the number of sensitivity indices is intractable.

This dissertation develops a procedure for application of uncertainty and sensitivity analysis of spatially distributed models with incorporation of spatial uncertainty of model inputs. A two-step procedure based on a geostatistical technique of sequential simulation and variance-based method of Sobol is proposed for incorporation of spatial

uncertainty into GUA/SA. The procedure considers both continuous and categorical model inputs**.** Continuous inputs (also referred to as numerical) are quantitative variables while categorical inputs are qualitative variables (classified into a number of exhaustive and mutually exclusive states). Land elevation is used as an example of continuous model input while land use type is used as example of categorical model input.

The benefits of this approach are compared with results for traditional screening analysis for lumped factors, used as a reference.

Research Objectives

This study aims to explore the application of global sensitivity and uncertainty techniques as a tool to evaluate complex, spatially distributed hydrological models. The Regional Simulation Model (SFWMD, 2005a; SFWMD, 2005b) in its application to WCA-2A will be used as test bed of the methods developed in this project.

The specific objectives of this study are:

- to perform global uncertainty and sensitivity analysis (GUA/SA) using approach for spatially lumped model inputs, as a reference for more advanced methodology developed in this dissertation (Chapter 2),
- to develop a procedure for incorporation of spatial uncertainty of numerical model inputs into GUA/SA and apply it for the benchmark model – RSM (Chapter 3),
- to apply the GUA/SA with incorporation of spatial uncertainty in order to optimize numerical (land elevation) data collection for RSM application to WCA-2A (Chapter 4),
- to develop a procedure for incorporation of spatial uncertainty of categorical model inputs into GUA/SA and apply it to the RSM, using land cover type as an example of categorical model input (Chapter 5), and
- to evaluate an importance of spatial uncertainty of continuous and numerical model inputs in terms of uncertainty of hydrological, spatially distributed models' predictions.

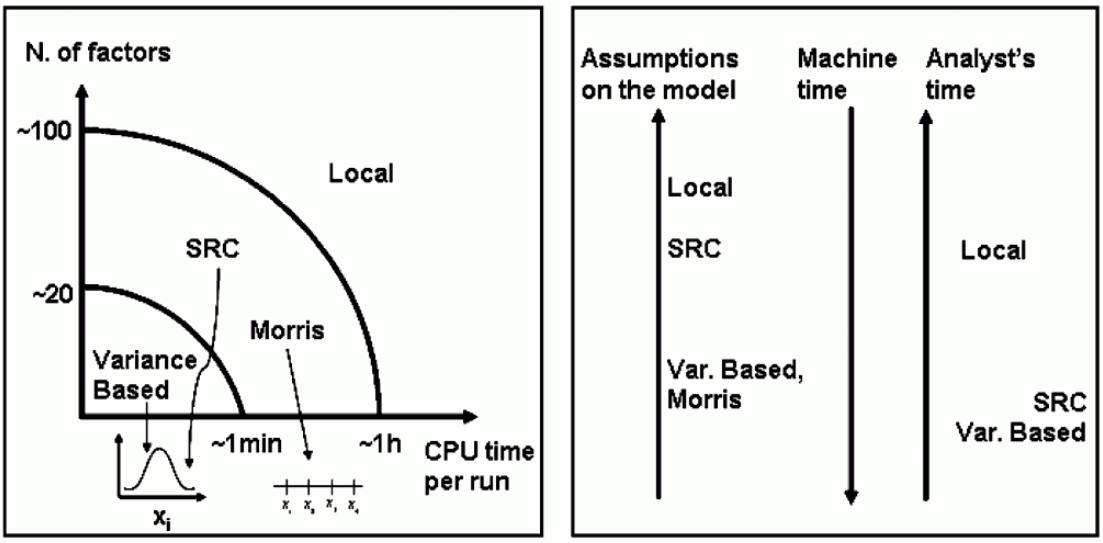


Figure 1-1. Factors influencing the use of various GSA techniques (after Saltelli et al, 2005, modified).

CHAPTER 2
EXPLORATORY GLOBAL UNCERTAINTY AND SENSITIVITY ANALYSIS, USING SPATIALLY LUMPED MODEL INPUTS

Introduction

Initially SA is performed using a screening method and spatially fixed input factors for the reference with more advanced SA methods, incorporating spatial uncertainty of model inputs, developed in further sections of this dissertation. In this chapter, the modified method of Morris is employed to initially assess the sensitivity of the Regional Simulation Model (RSM) applied to the WCA-2A conditions.

The purpose for this screening is to initially investigate the behavior of the model and indicate which input factors are important and which one are negligible. The screening test provides qualitative results (ranking of parameters importance). The computational cost of the screening SA is very low, compared to variance-based methods.

Test Case: Regional Simulation Model for Water Conservation Area-2A Application

The practical application of GUA/SA techniques proposed in this dissertation is illustrated using a spatially distributed, hydrological model Regional Simulation Model (RSM). The techniques are applied to the RSM for evaluation of model quality in a decision making framework for Water Conservation Area-2A in South Florida.

Regional Simulation Model

The Regional Simulation Model (RSM) is a spatially distributed hydrological model developed by SFWMD for evaluation of complex water management decisions in South Florida (SFWMD, 2005a). The RSM simulates physical processes in the hydrologic system, including major processes of water storage and conveyance driven by rainfall,

potential evapotranspiration, and boundary and initial conditions. RSM accounts for interactions among surface water and groundwater hydrology, hydraulics of canals and structures, and management of these hydraulic components. The governing model equations are based on the Reynolds transport theorem and the finite volume method is used to simulate the hydrology and the hydraulics of the system (SFWMD, 2005a).The governing equations for RSM are presented in Appendix A. RSM uses an unstructured triangular mesh to discretize the model domain. The model elements (cells) are assumed homogenous in terms of land elevation, land cover type, soil type, and hydraulic properties (SFWMD, 2005a).

RSM consists of the Hydrologic Simulation Engine (HSE) and the Management Simulation Engine (MSE). The HSE simulates the hydrological processes in the system. This component of the model is the focus in this study, and is referred to as the RSM. The MSE is not considered in this study. A large amount of well organized data is needed for the model to simulate the South Florida system. This is facilitated by the use of extensible markup language (XML) and geographic information system (GIS) for organizing model inputs (SFWMD, 2005a).

Model application to Water Conservation Area-2A

In this study RSM is applied to Water Conservation Area-2A (WCA-2A) in the Everglades Protection Area (EPA) (Figure 2-1). WCA-2A is a 547 km^2 natural marsh, consisting of sawgrass, sawgrass intermixed with cattail, open water sloughs and remnant drowned tree islands. It is completely surrounded by canals and levees. Surface water inflows and outflows are regulated and monitored. WCA-2 was created as a critical component of the Central and Southern Florida to provide flood protection, water supply and environmental benefits for the region. The WCA-2A area faces

ecological problems, related to shifts in vegetation communities from sawgrass (*Cladium jamaicense*) to cattail (*Typha domingensis*) caused by anthropogenic changes in water flow dynamics and increased nutrient loads. Traditional sawgrass slough vegetation has been replaced by pure cattail stands and cattail/sawgrass-slough vegetation (DEP, 1999). The dynamics and distribution of these species is controlled by nutrients and hydrologic conditions. Cattail growth is enhanced by elevated nutrients and increased flooding while sawgrass has higher capacity to resist cattail invasion in phosphorus poor conditions and shallow waters (Newman et al., 1996). Prolonged hydroperiod is conducive to cattail proliferation (Urban et al., 1993). In the WCA-2A hydrological conditions were found to be second most important (after nutrients) for controlling cattail and sawgrass communities' dynamics (Newman et al., 1998).

WCA-2A receives large inflows from agricultural runoff from the Everglades Agricultural Area (EAA) through four inflow structures (S- 10A, S- 10C, S-10D and S-10E) located along the north levee and the S-7 pump station (EPA, 1999; Urban et al., 1993) (Figure 2-1). The S-10E discharge structure has less capacity than the other S-10 structures but it does provide a way of directing water into the driest areas of WCA 2A (EPA, 1999). The southward flow of surface water from inflow structures has resulted in increased surface water and soil pore water nutrient gradient which has been documented previously (Davis, 1991; Koch and Reddy, 1992).

The current RSM application uses a model mesh with 386 triangular cells (within levee, shown in Figure 2-1) or 510 (included one layer out of the levee, not shown in Figure 2-1) varying from 0.5 km^2 to 1.7 km^2 (average of 1.1 km^2).

Model inputs and outputs

Spatial representation of model inputs used in this dissertation ranges from spatially lumped (i.e. one value is used for the whole domain), through regionalized (i.e., a group of cells is assigned the same input value) to fully distributed (i.e. each cell has an individual value assigned). Initially, in this Chapter, all model input factors for the GUA/SA are considered spatially fixed, i.e. no spatial uncertainty is considered. Later, land elevation is considered as a spatially uncertain numerical model input (Chapter 3 and 4) and finally, land cover type is considered as a spatially uncertain categorical model input (Chapter 5). The definition of all uncertain model inputs used in this study is presented in Table 2-1, together with their spatial characteristics. For more detailed description of model inputs the reader is referred to Appendix B.

In case of regionalized or fully distributed parameters, the so called level approach is used to reduce the number of input factors for the SA. In case of regionalized variable (for example parameter a, used for calculating Manning's roughness coefficient), alternative parameter values are generated from the PDF assigned to one of the zones, and values for all other zones are obtained by preserving the original ratio between zones. For more details regarding this approach, the reader is referred to Appendix C. In case of fully spatially represented hydraulic conductivity, the same "level" approach is used, only one representative cell is selected and probability distribution associated with this cell is sampled during the MC simulations, values for all other cells are obtained preserving the original ratio with the selected cell. In such way, the number of input factors is reduced significantly, and interpretation of results is easier, i.e. instead of 510 factors representing hydraulic conductivity for each cell individually, there is just one input factor representing the spatially distributed input. In case of land elevation and

aquifer bottom, an alternative approach is used for generation of alternative model input maps. The input factor is associated with the uncertainty model for error of a variable (not variable itself) and the generated values of errors are added to the base map. The same generated value of error is added for all model cells for each MC realization. The probability distributions of input factors are selected based on specific conditions of the South Florida application.

Apart from scalar input factors, the GUA/SA also requires that model outputs are scalar quantities such as a summary or aggregate objective function (Crosetto and Tarantola 2001) in order for the empirical PDFs of outputs to be constructed. Raw RSM outputs are spatially and temporarily distributed: they include water depth and stage reported for each of the model cells on a daily basis for the period of the simulation. These raw outputs need to be post-processed into objective functions that are suitable for the GUA/SA and meaningful for decision makers. The same procedure for post-processing raw model outputs is applied in all GUA/SA studies presented in this dissertation (Appendix D). The RSM performance objective functions (also referred as outputs) chosen as metrics for GUA/SA for this study were the performance measures generally adopted in the Everglades restoration studies (SFWMD, 2007): hydroperiod, water depth amplitude, mean, minimum and maximum. The GUA/SA results for two types of objective functions: domain-based approach (spatial averaging over domain), and benchmark cell-based approach are compared in this work. The benchmark cells (14 cells presented in Figure 2-1) are selected based on location in a domain and can be divided into four groups of interest: 1) cells located in the north of the domain, representing the driest areas in the domain (cell 35), 2) cells located in north-east of the

domain, representing cattail invaded areas (cell 178, 215), cells located in the south of domain, representing the wettest areas in the domain (cell 486) and 4) other cells, used for the reference to other benchmark cells (cell 224).

In all of the GUA/SA studies presented in this dissertation the simulations are performed for period 1983-2000. A one year long warm-up period (1983) is chosen to reduce the influence of the initial conditions on the model outputs. The calculated outputs are aggregate values representative for this period.

Sensitivity and uncertainty methods previously applied to RSM

Sensitivity and uncertainty analysis was previously performed on the Natural Systems RSM (NSRSM). NSRSM is a specific application of the RSM, which was designed to simulate the predevelopment hydrologic response. The model was constructed using a pre-development (i.e. pre-drainage, mid-19th century) land cover condition and predevelopment topography (Mishra et al., 2007).

The analysis of NSRSM considered only a subset of uncertain input factors that was selected subjectively by the analysts prior the analysis (Mishra et al., 2007). This is not a robust approach since sometimes the results of sensitivity analysis are very counterintuitive and it is hard to indicate a priori which factors are important with respect to the outputs and which are not. Because of this, the analysis based on subjectively chosen subset of parameters is not the optimal method for verification of the model.

For the sensitivity analysis the Singular Value Decomposition (SVD) (Doherty, 2004) was applied to NSRSM. SVD-based sensitivity analysis involves the factorization of the sensitivity matrix (Jacobian matrix of local sensitivities) to create matrices which define linearly independent groups of parameters and outputs. A vector of singular values is also created by the decomposition. These singular values indicate the relative

importance of each parameter group. The inclusion and importance of parameters in the linearly independent groups provides insight into both parameter interactions and synergies, as well as the local sensitivity of output metrics to the parameters. The SVD should be used only for linear and monotonic models (i.e. models for which input-output relation is linear or monotonic) (Mishra et al., 2007). The findings of this research were that, in general, variance of an output metric (water stage and transect flow) was controlled by the ET, crop coefficient, conveyance parameter, Manning's n, and to a lesser extent, topography.

The two uncertainty analysis techniques were applied to NSRSM: First-Order Second-Moment (FOSM) and Monte Carlo simulations. For k model inputs, the FOSM method requires only N=k+1 model simulations, as opposed to several thousand simulations for typical Monte Carlo simulations. However, the drawback of this approach is that it estimates uncertainty in model predictions only in terms of mean and standard deviation (rather than the full output distributions). These statistics may not be the most useful indicators about the model output because the information is always lost in the calculations of means and standard deviations. Also, these measures may not be adequate statistics for skewed output distributions. This analysis should only be applied to linear or mildly nonlinear problems (Mishra and Parker 1989). The FOSM analysis was not carried for the topography (considered as categorical variable with three alternative topography scenarios: "low", "base", and "high" maps), since categorical variables are not amenable to derivative calculations (Mishra et al., 2007).

Uncertainty analysis by the Monte Carlo approach (random or Latin Hypercube) consisted of the following steps: (1) selection of imprecisely known model input

parameters to be sampled, (2) construction of PDF for each of these parameters, (3) generating a sample scenario by selecting a parameter value from each distribution, (4) calculating the model outcome for each sample scenario and aggregating results for all samples (Mishra et al., 2007). By the initial examination of results, 100, 200 and 300 realization cases were examined for model stability and a sample size of 200 was found adequate to provide stable output statistics. The methods applied previously to RSM have not considered spatial distribution of input factors.

Screening Method: Morris Elementary Effects

Morris (1991) proposed an effective screening sensitivity measure to identify the few important factors in models with many factors. The method is based on computing for each input a number of incremental ratios, called elementary effects (EEs), which are then averaged to assess the overall importance of a given input factor. Campolongo (2005) proposed modifications to the original method of Morris improved in terms of the definition of the sensitivity measure. The guiding philosophy of the original elementary effects method (Morris, 1991) is to determine which input factors may be considered to have effects which are (a) negligible, (b) linear and additive, or (c) non-linear or involved in interactions with other factors. Morris (1991) proposed conducting individually randomized experiments that evaluate the elementary effects along trajectories obtained by changing one parameter at a time. Each model input X_i, i=1,.., k (where k is a number of inputs) is assumed to vary across p selected levels within its distribution. The region of experimentation Ω is thus a k-dimensional p-level grid. Following a standard practice in sensitivity analysis, factors are assumed to be uniformly distributed in [0,1] and then transformed from the unit hypercube to their actual distributions. Therefore for all model inputs, each level is associated with a given percentile of the

probability distribution). Elementary effects are calculated by varying one parameter at a time across a discrete number of levels (p) in the space of input factors. The elementary effect is calculated from:

$$EE(X_i)=\left(\frac{y\left(X_1,\ldots,X_{i-1},X_i+\Delta,\ X_{i-1},\ldots,X_k\right)-y(X_i)}{\Delta}\right) \quad (2\text{-}1)$$

where: EE(Xi) – elementary effect for a given factor Xi, Δ is a value in {1/(p-1),…,1-1/(p-1)} this value defines a “jump” in the parameter distribution between two levels considered for calculating the elementary effect , p – number of levels. The illustration of Morris sampling scheme for one input factor is presented in Figure 2-3 for p=4 and Δ of 2/3.

A number r of elementary effects is obtained for each input factor. Based on this number of elementary effects calculated for each input factor, two sensitivity measures are proposed by Morris (1991): (1) the mean of the elementary effects, μ, which estimates the overall effect of the parameter on a given output; and (2) the standard deviation of the effects, σ, which estimates the higher-order characteristics of the parameter (such as curvatures and interactions).

Campolongo noticed weaknesses of the original measure μ in the method of Morris (1996) and proposed modification of the original method in terms of the definition of this measure (2005). Since sometimes the model output is non-monotonic, the elementary effects may cancel each other out when calculating μ, this measure can be prone to the Type II error, i.e. failing in the identification of a factor of considerable influence on the model. Campolongo et al. (2005) suggested considering the mean of distribution of absolute values of the elementary effects, μ*, for evaluation of

parameter's importance in order to avoid the canceling of effects of opposing signs. The measure μ* is a proxy of the variance-based total index is acceptable and convenient (Campolongo, 2007) and can be used for ranking the parameters according to their overall effect on model outputs. Saltelli et al. (2004) suggest applying the original Morris (1991) measure, σ, when examining the effects due to interactions. Thus measures μ* and σ are adopted as global sensitivity indices in this study.

To interpret the results in a manner that simultaneously accounts for the mean and standard deviation sensitivity measures, Morris (1991) suggested plotting the points on a μ-σ Cartesian plane. The higher the measure μ* is, the more important factor is. The parameters with μ* values close to zero can be considered as negligible (non-important) ones. The parameters with the largest value of μ* is the most important one. However, the value of this measure for a given factor does not provide any quantitative information on its own and needs to be interpreted qualitatively, i.e. relatively to other factors' values. The meaning of σ can be interpreted as follows: if the value for σ is high for a parameter, X_i, the elementary effects relative to this parameter are implied to be substantially different from each other. In other words, the choice of the point in the input space at which an elementary effect is calculated strongly affects its value, which means it is sensitive to the chosen values of other parameters that constitute the remainder of the input space. Conversely, a low σ value for a parameter implies that the values for the elementary effects are relatively consistent, and that the effect is almost independent of the values for the other input parameters (i.e. no interaction).

The required number of simulations (N) to perform in the analysis results as:

$$N = r\,(k + 1) \qquad (2\text{-}2)$$

Previous studies have demonstrated that using p = 4 and r = 10 produces satisfactory results (Campolongo et al., 1999; Saltelli et al., 2000). So for example, in case of k=20 uncertain input factors, only 210 model simulations are required for the method of Morris (while variance-based methods, described in Chapter 3, would require approximately 20,000 simulations).

Despite the fact that the fundamental measure of Morris method - the elementary effect (or its absolute value) - uses local incremental ratios, this method is not considered as local. The final measure μ* is obtained by averaging the absolute values elementary effects which eliminates the need to consider the specific points at which they are computed (Saltelli et al., 2005). The method, therefore, is considered as a hybrid between local and global approaches because it samples across the input factors space yields a global measure.

Methodology

Sensitivity Analysis Procedure

The screening procedure follows the general steps required by MC based SA methods (Figure 2-4): 1) selection of input factors and construction of probability distribution functions; 2) generation of input sets by pseudo-random sampling of input PDFs according to the selected sampling scheme (in this case sampling according to the method of Morris); 3) running model simulations for each input set and obtaining corresponding model outputs; 4) performing global sensitivity (here according to the modified method of Morris).

The software package, SimLab v2.2 (Saltelli et al., 2004), is used for the SA by the modified method of Morris. SimLab is designed for pseudorandom number generation-based uncertainty and sensitivity analysis. SimLab's Statistical Pre-Processor module

executes step 2 in the procedure (Figure 2-4) based on PDFs provided by the user and the method selected and produces a matrix of sample inputs to run the model (step 3, Figure 2-4). LINUX scripts were written to automatically run RSM once for each new set of sample inputs. The scripts automatically substitute the new parameter set into the input files, run the model, and perform the necessary post-processing tasks to obtain the selected model outputs for the analysis. The outputs from each simulation are stored in a matrix containing the same number of lines as the number of samples generated by SimLab. With the input and output matrices the Statistical Post-Processor module of SimLab is used to calculate the sensitivity indices by the method of Morris (step 4). SimLab produces sensitivity measures based on the absolute values of elementary effects, proposed by Campolongo (2005), that are μ^* and σ^*.

Definition of Model Inputs and Outputs for the Screening SA

Table 2-2 shows uncertain input factors (k=20) used for the screening, together with corresponding uncertainty specifications (probability distribution functions). The PDFs are assigned based on literature review and experts opinion, having in mind conditions specific to South Florida. In case of lack of information on variability of input factor, uniform distribution with ranges ± 20% around the base value of input factor (i.e. value of a input factor from the calibrated model) is used. For the purpose of the screening analysis, all input factors are assumed spatially lumped (no spatial uncertainty is considered).

Raw RSM outputs are spatially and temporally distributed. To obtain an aggregated statistics for each simulation, raw results are post-processed using scripts in AWK programming language. Details on post-processing procedures are provided in Appendix D. Two types of model outputs are calculated: 1) domain-based outputs (by

spatial averaging of cell-based outputs over the domain), and 2) benchmark cell-based outputs. Three benchmark cells are selected for the screening exercise: cell 35 - representing drier conditions in north of the domain, cell 178 - representing cattail invaded areas in northeast of the domain and cell 486 - representing wet areas in the south of the domain (Figure 2-1).

For k=20, only N=210 model simulations are required (for r=10 in equation 2-2). The screening analysis is performed using RSM simulations for 15 years, from 1983 to 2000, with one year long warm-up period (1983).

Results

As suggested by Campolongo (2005), the ranking of importance of the input factors can be based on the relative value of μ*. Such ranking for all domain-based, as well as benchmark cell-based outputs is provided in Table 2-3. Only important parameters have assigned ranks in this table. Figure 2-5 shows the graphical representation of the Morris sensitivity measures for a selected subset of domain-based outputs (Mean Water Depth, Hydroperiod, and Maximum Water Depth). Parameters, identified as important, are separated from the origin of the μ*-σ plane are considered important. Parameters located at the origin of the plane are assumed to have negligible effect on model outputs.

In general, the number of parameters identified as important parameters is effectively smaller than the full set of model inputs studied (from original 20 inputs down to 6 main inputs for domain-based and 7 main inputs for cell-based outputs). Especially, few factors: *topo*, *a*, *det*, *kds*, *imax* are important for the majority of outputs, both domain and cell-based (except outputs for cell 486). While other factors like *leakc*, *kmd* are identified as potentially important for some outputs (Table 2-3).

Factor *topo*, associated with the uncertainty of land elevation, is found as the most important for the domain-based outputs (Figure 2-5). This factor determines how much the initial land elevation map is shifted up or down (the initial relationship between cell values is maintained for each realization). Apart from *topo*, domain-based outputs are influenced by factor *a* and *det*. Factor *a* is used for calculating mesh Manning's roughness coefficient, while factor *det* accounts for water detained in depressions within model cells, as it determines the minimum water depth that needs to be reached for overland flow from to occur one cell to the neighboring cell. Factor *imax*, specifying the interception, contributes to uncertainty of the domain-based hydroperiod. Maximum water depth for domain seem also to be slightly affected by factor *n*, which represents Manning's roughness coefficient for canals, but the effect of factors *topo* and *a* is much stronger (Figure 2-5). Some of the cell-based outputs, like mean water depth and hydroperiod for cell 35 and 178, are affected by factor *kds* (Figure 2-6). This factor specifies levee hydraulic conductivity from a dry cell to a segment. SA results for cell 486 are different than for the other two benchmark cells and indicate that the outputs for this cell are mainly affected by *topo* in case of mean and maximum water depth and the *leakc* (leakage coefficient for canals, specifies flow between aquifer and canals) in case of hydroperiod (Figure 2-6).

Discussion

The results clearly illustrate two of the products of the global sensitivity analysis: ranking of importance of the parameters for different outputs, and type of influence of the important parameters (first order or interactions).

Factor *topo*, determining the shift of land elevation for the domain is indicated as potentially the most important factor for both domain-based and cell-based outputs. This

is expected since surface water inflows and outflows in the current application are fixed and controlled by hydraulic structures. Therefore the shift of land elevation in the domain affects volume of water that can be retained in a domain. Apart from land elevation shift, model response is controlled by conveyance parameters: parameter *a* and *det*. Unlike previously performed SA studies of the NSRSM (Mishra et al., 2007) that identified the crop coefficient (*kveg*) parameter as the most important one, this ET parameter is found as non-important. However, it is important to highlight that the results of this study are specific for the WCA-2A application and selected objective functions (outputs).

The SA results for cells are affected by the specific conditions in the given section of the domain. For example results for the cell 486 reflect that this area of the domain collects all the flow, and the local water depth is conditioned on the local levee characteristics (seepage coefficient).

The modified method of Morris results indicated the additive nature of the model, since small interactions are observed (the values of σ are small for all model inputs), except for hydroperiod for cells 35 and 178, where values of σ are larger (Figure 2-6).

The proposed framework provided further validation of the model quality since no errors were detected regarding the model behavior (all the relations between inputs and outputs can be explained on the basis of the model assumptions).

The results of this study indicated which factors are of potential importance. This subset of factors (6-8 factors) could be used for the more accurate, quantitative SA analysis (as in Muñoz-Carpena et. al, 2007). For example, the reduction of parameter input set from 20 original parameters to 8 identified as important by the screening

method, may result in reduction of number of simulation required by Extended FAST from approx. 20,000 to 8,000, as explained in Chapter 3.

Furthermore, since the factor related to land elevation representation for the WCA-2A is identified generally as the most important one, this factor is going to be the focus of methodology applied in Chapters 2 and 3 of this dissertation. The rudimentary approach for describing the uncertainty of land elevation will be refined with a more advanced uncertainty description, which accounts for spatial uncertainty of land elevation and produces more realistic land elevation realizations.

Conclusions

The modified method of Morris is a screening SA method applied to RSM and WCA-2A application. This method is characterized by relatively small computational cost and it is applied for identification of important and negligible model inputs. The ranking of parameters importance is calculated based on the global measure μ* - mean of the absolute values of elementary effects. Moreover a type of influence of the important parameters (first order or interactions) may be assessed by measure σ – the standard deviation of elementary effects.

The screening performed here indicates that out of the 20 original model inputs, 8 inputs are important for the considered model outputs. Input factor *topo*, characterizing land elevation uncertainty (vertical shift of land elevation values) is identified as the most important factor in respect to most of the outputs (both domain-based and benchmark cell-based). Other factors, found important for several outputs, are conveyance parameters: *a* and *det* , interception parameter *imax*, factor *kds* (levee hydraulic conductivity from dry cell to segment), and *leakc* (leakage coefficient for

canals) for cell 486. Small interactions between parameters were observed, indicating that for the selected outputs, the model is of additive nature.

The Morris method is qualitative in nature, its sensitivity measures should not be used to quantify input factors' effects on uncertainty of model outputs. They rather provide qualitative assessment of parameter importance in form of a parameter ranking. Furthermore, this method cannot account for spatial uncertainty of model inputs because it requires that all input factors are scalar values, and uses an analytical relationship between model input and output for calculating sensitivity measures.

As land elevation is identified as one of the most important model inputs, this model input is going to be used as an example of spatially distributed numerical model input in further chapters of this dissertation.

Table 2-1. Definition of uncertain model inputs used for the GUA/SA.

#	Model Input	Definition	Units	Spatial Representation
1	$value_{shead}$	initial water head	[m]	lumped
2	topo	land elevation error	[m]	fully distributed[1]
3	bottom	aquifer bottom elevation	[-][2]	fully distributed[1]
4	hc	hydraulic conductivity	[m^2s^{-1}]	fully distributed
5	sc	storage coefficient of solid ground	[-]	lumped
6	kmd	levee hydraulic conductivity from a marsh cell to a dry cell	[m^2s^{-1}]	regionalized
7	kms	levee hydraulic conductivity from a marsh cell to a segment	[m^2s^{-1}]	regionalized
8	kds	levee hydraulic conductivity from a dry cell to a segment	[m^2s^{-1}]	regionalized
9	n	Manning's n for canals	[$sm^{-1/3}$]	lumped
10	leakc	leakage coefficient for canals	[-]	lumped
11	bankc	coefficient for flow over the canal lip	[-]	lumped
12	a	parameter "a" in equation $n_{mesh}=a*depth^{-0.77}$	[-]	regionalized
13	det	detention	[m]	lumped
14	kw	maximum crop coefficient for open water	[-]	lumped
15	rdG	shallow root zone depth [m] for grasses	[m]	lumped
16	rdC	shallow root zone depth [m] for cypress	[m]	lumped
17	xd	extinction depth below which no ET occurs	[m]	lumped
18	pd	open water ponding depth	[m]	lumped
19	kveg	ET vegetation crop coefficient	[-]	regionalized
20	imax	maximum interception	[m]	lumped

[1] in case of land elevation (topo) and aquifer bottom elevation (bottom), the input factor used for the screening SA specifies error around the original values and it is spatially lumped, the same error value is added to original maps resulting in fully distributed inputs;

[2] aquifer bottom elevation units are [m] but the error is unit less since it specifies percentage of original bottom values (this approach is easier to implement because of the structure of bottom input file);

[3] n_{mesh} – Manning's roughness coefficient for cells, calculated for each time step based on the calculated water depth (depth).

Table 2-2. Characteristics of input factors, used for screening SA.

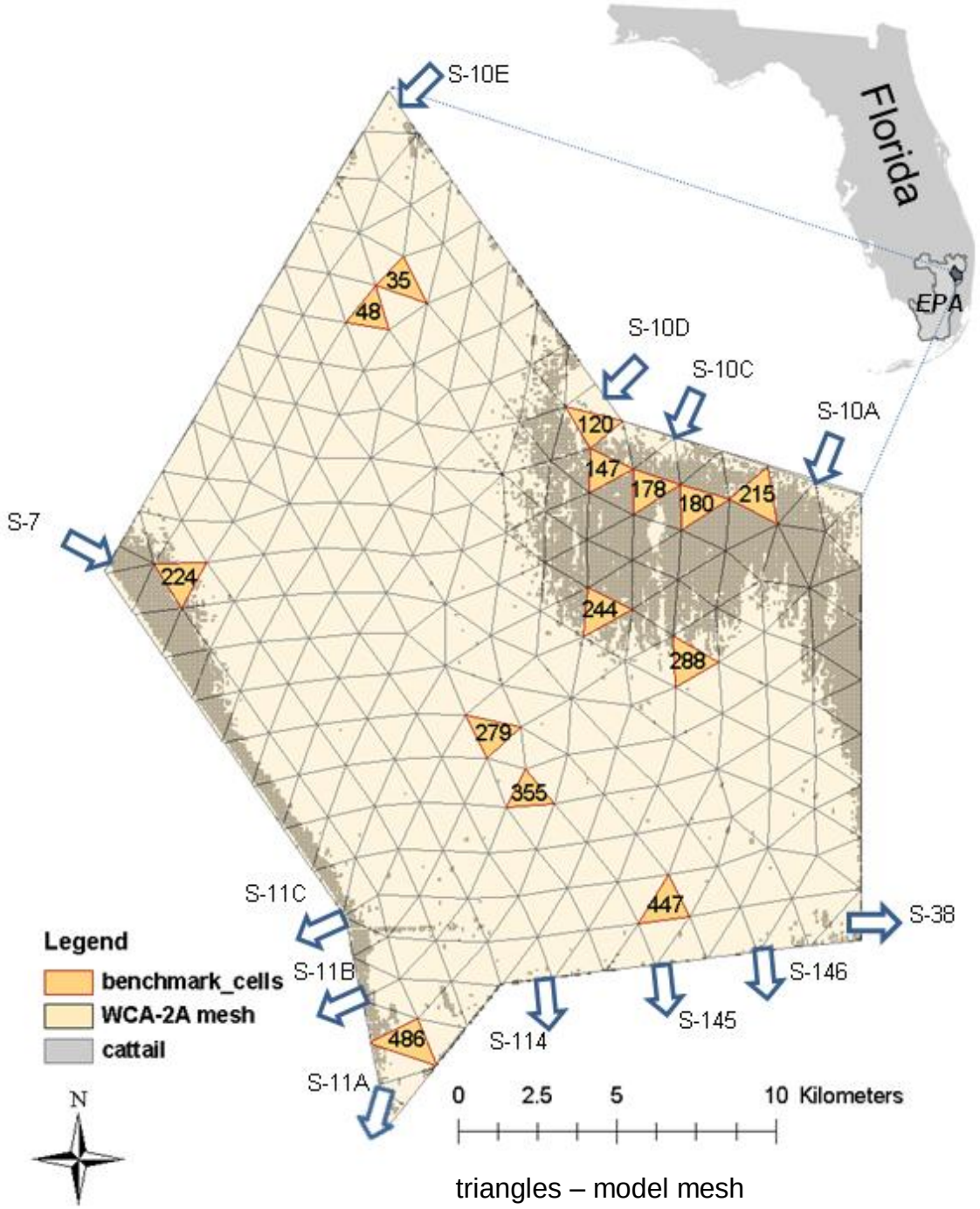
#	Input Factor	Base Value[1]	Uncertainty Model (PDF)	Source
1	$value_{shead}$	3.66	N[1](μ=3.66, σ=0.374)	Jones and Price, 2007
2	topo	-	N(μ=0, σ=0.05)	USGS, 2003
3	bottom	0	U[2] (0.8, 1)	SFWMD data
4	hc	46.5 [3]	Lognormal(μ=4.6, σ=1.2)	SFWMD data
5	sc	0.3	U (0.2, 0.3)	SFWMD expert opinion
6	kmd	0.000026 [4]	U (0.000021, 0.000032)	± 20%
7	kms	0.000011[4]	U (0.000009, 0.000013)	± 20%
8	kds	0.0000031 [4]	U (0.0000025, 0.0000038)	± 20%
9	n	0.06	Triangular (min.= 0.03, peak=0.10, max.=0.12)	SFWMD expert opinion; USGS, 1996
10	leakc	0.00001	U (0.000002, 0.001)	SFWMD data
11	bankc	0.05	U (0.04, 0.05)	SFWMD data
12	a	0.3 [5]	U (0.24, 0.36)	± 20%
13	det	0.03	U (0.03, 0.12)	Mishra et al., 2007
14	kw	1	U (0.8, 1.2)	± 20%
15	rdG	0	U (0, 0.2)	Yeo, 1964,
16	rdC	0	U (0, 1.5)	expert opinion
17	xd	0.9 [6]	U (0.7, 1.1)	Mishra et al., 2007
18	pd	1.8 [6]	U (1.5, 2.2)	± 20%
19	kveg	0.83 [6,7]	U (0.66, 0.99)	± 20%
20	imax	0	U (0, 0.03)	SFWMD expert opinion

[1] value of input from calibrated model;
[2] N - normal distribution; DU - discrete uniform distribution; U - uniform distribution;
[3-6] base values for a cell or region, used as a reference for the level approach:
[3] cell 333, [4] L38E, [5] zone 3, [6] cattail HRU;
[7] average annual value of *kveg* is used, no seasonal variation is considered.

Table 2-3. Ranking of parameters importance obtained from the modified method of Morris.

	Mean Water Depth				Hydroperiod				Minimum Water Depth				Maximum Water Depth				Amplitude			
	D[1]	35	178	486	D	35	178	486	D	35	178	486	D	35	178	486	D	35	178	486
$value_{shead}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
topo	1	2	2	1	1	1	1	2	1	4	4	1	1	2	-	1	1	2	5	1
$error_{bottom}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
hc	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
sc	-	-	-	-	6	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-
kmd	-	-	-	-	-	6	7	-	-	-	-	-	-	-	-	2	-	-	-	-
kms	-	6	-	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
kds	-	4	3	-	5	2	2	-	4	1	3	-	-	4	4	-	-	5	4	-
n	-	-	-	-	-	-	-	-	-	-	-	-	3	6	-	-	4	-	-	-
leakc	-	-	-	2	-	-	-	1	-	-	-	2	-	-	-	-	-	-	-	2
bankc	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
a	2	1	1	-	4	3	4	-	2	3	2	-	2	1	1	-	2	1	1	-
det	3	3	4	-	3	4	5	-	3	2	1	-	-	5	3	-	3	3	2	-
kw	-	-	-	-		9	-	-	-	-	-	-	-	-	-	-	-	-	-	-
rdG	-	-	-	-	7	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
rdCY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
xd	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-	-	-	-	-	-
pd	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
kveg	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
imax	4	5	5	-	2	5	3	-	-	-	-	-	4	3	2	-	5	4	3	-

[1] D – domain-based outputs, 35, 178, 486 – benchmark cell-based outputs for cells 35, 178, and 486 (Figure 2-1)



triangles – model mesh
arrows – inflows and outflows
shading – cattail-dominated areas
EPA – Everglades Protection Area

Figure 2-1. Location of the model application area: Water Conservation Area 2-A.

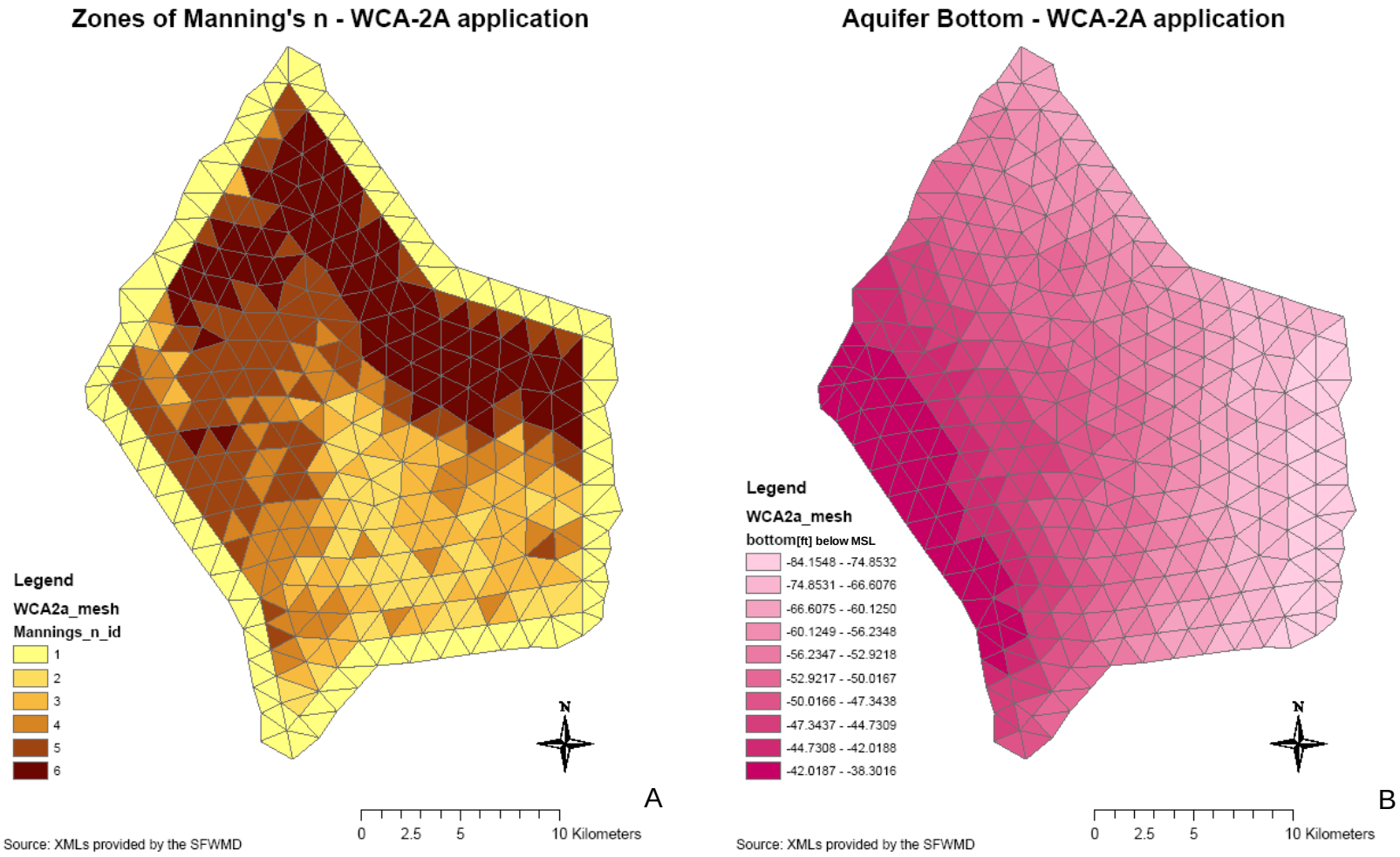
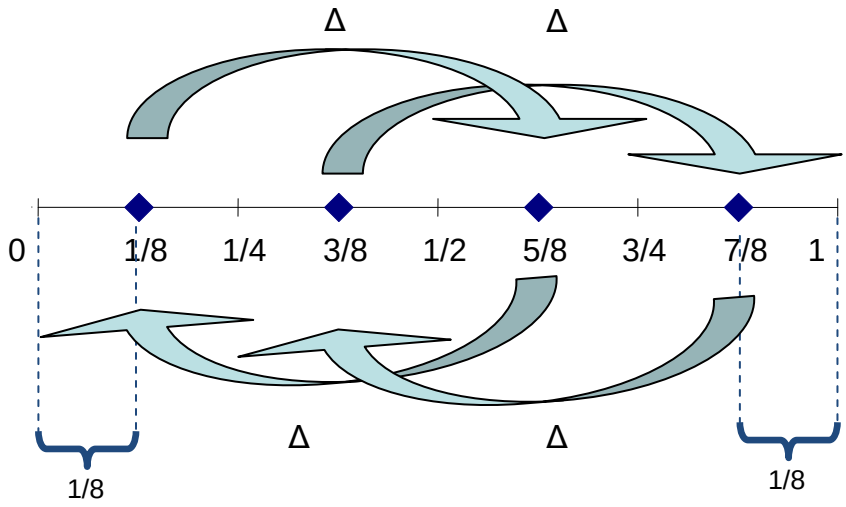
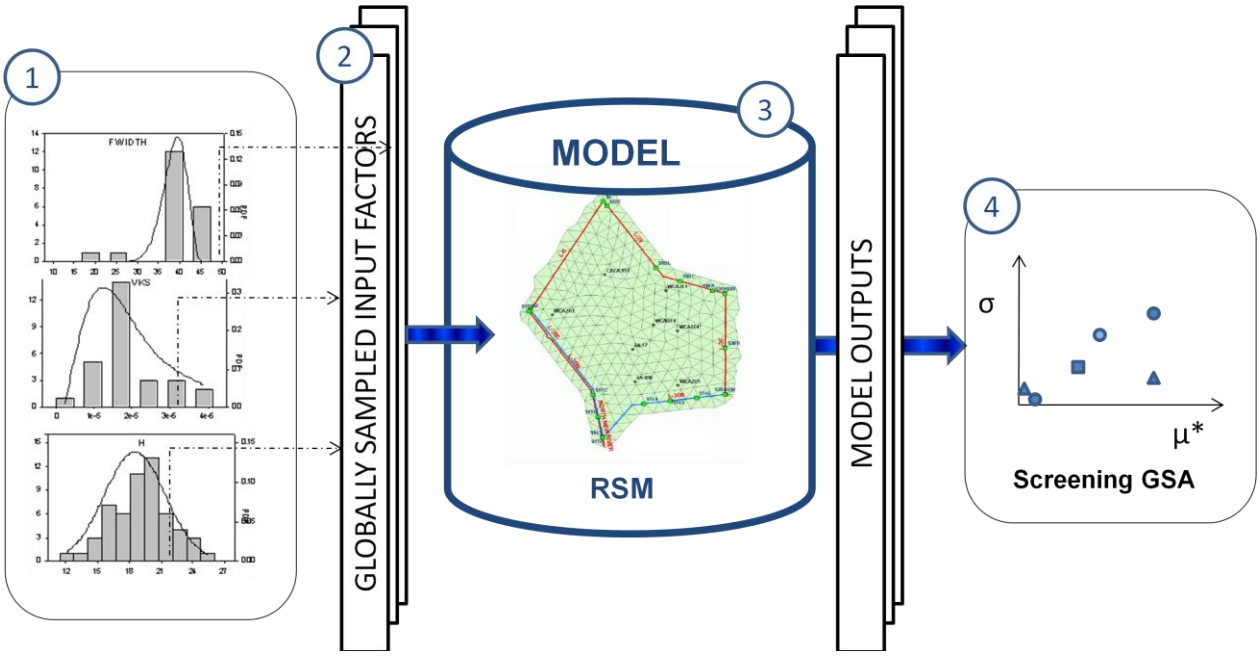


Figure 2-2. Example of spatial representation of model inputs. A) regionalized input (parameter *a* for calculating Manning's n), B) fully distributed input (elevation of bottom of aquifer).



p=4, Δ=1/2; numbers indicate percentiles of the factor's distribution (e.g. 1/8 indicates 12.5th percentile)

Figure 2-3. Illustration of Morris sampling strategy for calculating elementary effects of an example input factor, as applied in SimLab.



numbers in circles represent steps in the global evaluation procedure

Figure 2-4. General schematic for the screening GSA with modified method of Morris.

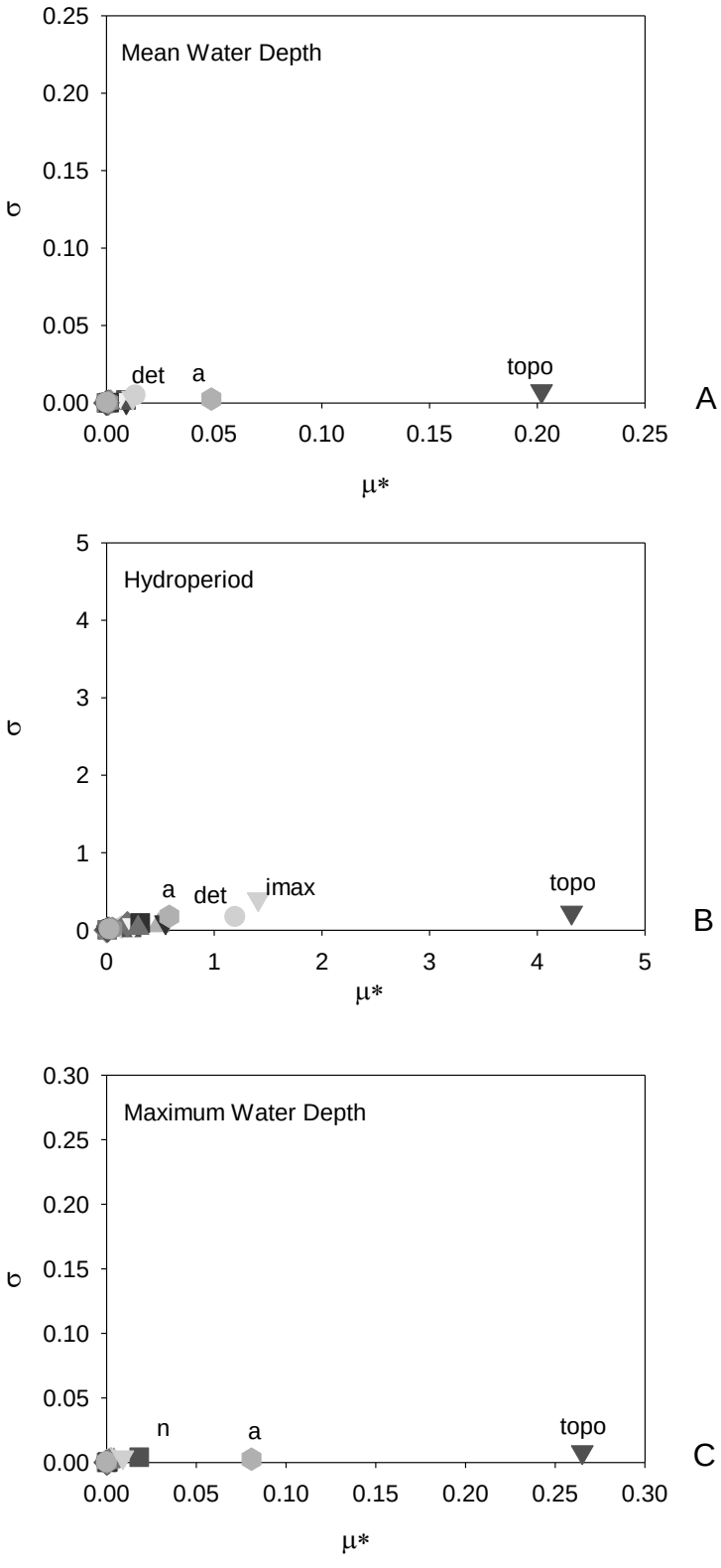


Figure 2-5. Method of Morris results for domain-based outputs. A) mean water depth, B) hydroperiod, C) maximum water depth.

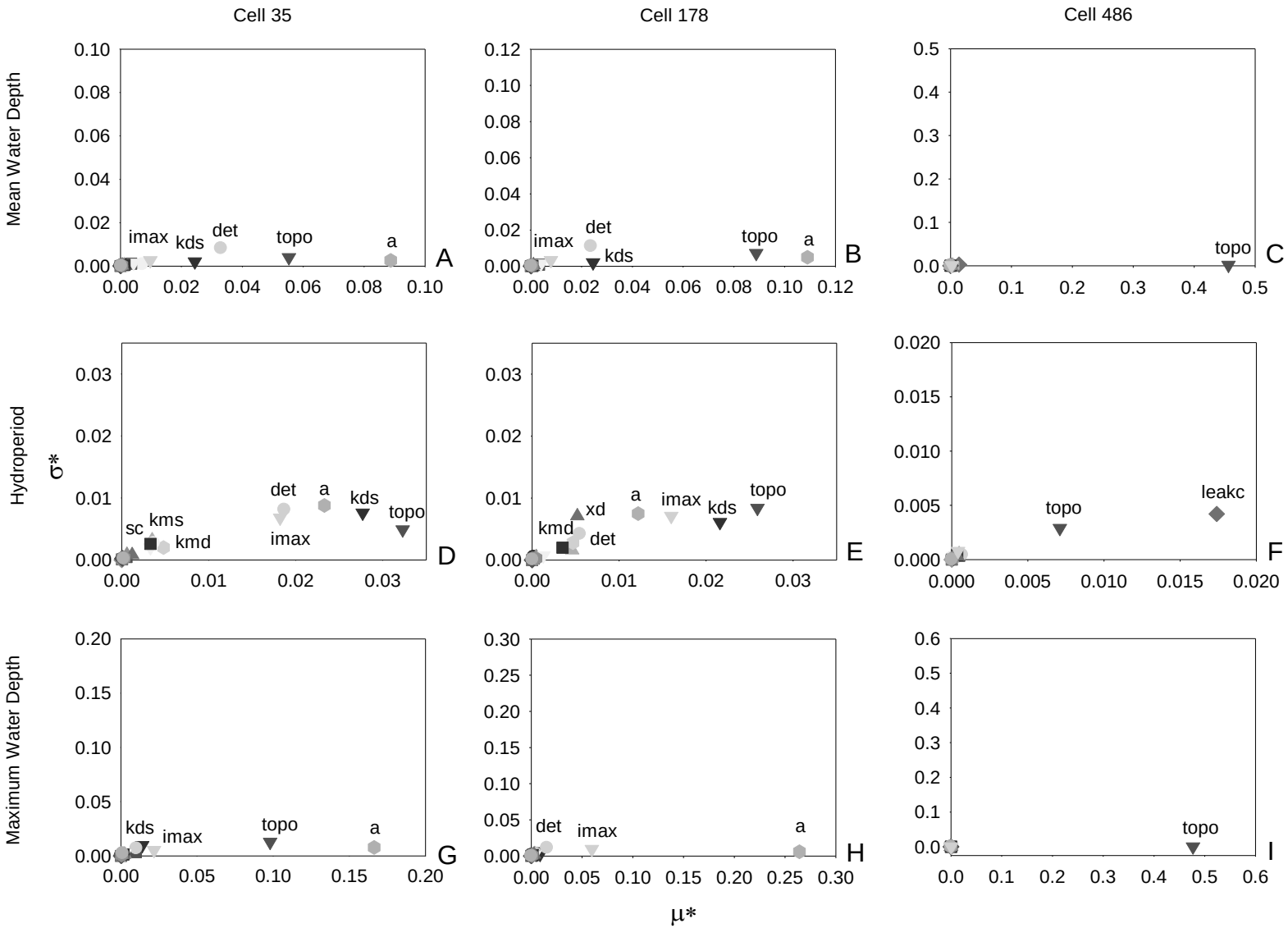


Figure 2-6. Method of Morris results for selected benchmark-cell based outputs. A), B), C) mean water depth, D), E), F) hydroperiod, G), H), I) maximum water depth.

CHAPTER 3
INCORPORATION OF SPATIAL UNCERTAINTY OF NUMERICAL MODEL INPUTS INTO GLOBAL UNCERTAINTY AND SENSITIVITY ANALYSIS OF A SPATIALLY DISTRIBUTED HYDROLOGICAL MODEL

Introduction

Incorporating Spatiality in Global Uncertainty and Sensitivity Analysis

A two-step procedure based on the geostatistical technique of sequential simulation and the variance-based method of Sobol is proposed for incorporation of spatial uncertainty into GUA/SA.

Sequential simulation (SS) provides a quantitative measure of spatial uncertainty, i.e., uncertainty regarding spatial distribution of a variable rather than location-specific uncertainty (Journel , 1989; Goovaerts, 1997). Spatial uncertainty results from the fact that knowledge of spatial distribution of phenomena is limited to measurement locations and uncertainty arises regarding spatial structure between these locations. Sequential simulation is a process of drawing alternative, equiprobable, joint realizations of the spatial variable that honor the measured data, data statistics (global histogram), and model of spatial correlation (variogram) within ergodic fluctuations (Deutsch and Journel,1998, Goovaerts, 1997). The theory behind sequential simulation has been explained thoroughly by others (Chilès and Delfiner, 1999; Deutsch and Journel, 1998; Goovaerts, 1997; Kyriakidis, 2001). Rossi et al. (1993) uses an analogy of a jigsaw puzzle, with an incomplete image in the top box, for illustration of the SS principles. Measured data are equivalent to known puzzle's pieces. Since there is only partial information about the final image on the box top, multiple equiprobable images can be constructed. These alternative final images, taken together, characterize the uncertainty about the true picture on the box top. Of the many SS techniques, Sequential Gaussian

Simulation (SGS) is often used because it is fast and straightforward (Deutsch and Journel, 1998). SGS has been applied in many studies such as remediation processes and flow simulation models, which require a measure of spatial uncertainty, rather than location-specific uncertainty (Goovaerts, 1997).

As presented in Chapter 2, the GUA/SA methodology has been applied primarily to lumped models, where all input factors were scalar and generated from scalar PDFs. In the case of spatially distributed input factors, alternative maps (rather than alternative scalar values) need to be generated and processed by the model. The application of UA to spatial models, using geostatistical techniques and MC simulations is straightforward and requires processing of alternative spatial realizations through the model (Phillips and Marks, 1996). In this way, uncertainty regarding the spatial representation of variable is transferred into consequent model uncertainty (Kyriakidis, 2001).

Uncertainty associated with spatial structure of input factors may affect model uncertainty and therefore influence model sensitivity. However, examples of the application of GSA techniques that account for spatial structure of input factors are rare and limited in scope (Crosetto et al., 2000, Crosetto end Tarantola, 2001; Francos et al. 2003, Hall et al., 2005; Tang et al., 2007a). GSA methods generally have limitations that make them unsuitable for evaluation of spatially distributed models (Lilburne and Tarantola, 2009). The shortcomings of GSA applied to distributed spatial models are related to impractical computational costs and the inability to realistically represent spatial structure. GSA methods based on the MC sampling require that inputs are represented by a scalar values. Medium-size watershed models (i.e., hundreds of hectares) may have hundreds or thousands of discretization units. If GSA is performed

for all cells individually (each parameter value of each discretization unit treated as independent factor) the computational cost of analysis for watershed models becomes impractical and the number of sensitivity indices is intractable.

This “fully distributed” spatial representation approach was used in Tang et al. (2007a), where SA is performed for all cells individually using the extended FAST. Apart from high computational and processing costs, this approach cannot account for spatial structure of inputs. Because of an assumption of factor independence inherent in variance-based methods (Saltelli et al., 2004), input factors representing cells need to be considered independent from one another for MC simulations, so spatial autocorrelation between neighboring cells cannot be accurately represented.

Several approaches have been proposed in the literature to simplify dimensionality in the problem and reduce computational demands. The crudest approach is to disregard spatial distribution of input factors (i.e., consider them as spatially lumped) (Crosetto and Tarantola, 2000; Tang et al., 2007b). Other methods propose spatial simplification of the domain to smaller number of zones (Chu-Agor et. al, 2010; Hall et al., 2005). The zones may be correlated with one another using a simple statistical model of spatial variation (Hall et al., 2005). However, the spatial structure of inputs cannot be reproduced realistically since the zones themselves are homogenous.

To address these shortcomings, Crosetto and Tarantola (2001) proposed the use of an indirect (auxiliary) input factor for GSA. The binary input factor is used as a “switch” that determines if model simulations are performed using realizations generated from a spatial uncertainty model (switch on) or if spatial structure is ignored (switch off). This approach allows for checking if the spatial representation of a given

factor has an influence on model outputs, but does not allow for the simultaneous UA. Expanding this approach, Lilburne and Tarantola (2009) proposed using the auxiliary factor approach with the method of Sobol. The auxiliary scalar factor with Discrete Uniform (DU) distribution is associated with a number of alternative spatial realizations (i.e., maps, with the number of spatial maps equal to the number of levels of an auxiliary factor), which are then used for MC simulation. When a given value from the factor's distribution is generated, the associated map is used for model runs. The specifics of calculating sensitivity indices using the method of Sobol (i.e., no analytical relation between inputs and output); allows for the incorporation of spatial uncertainty into GSA via an auxiliary input factor. There is no assumption on how alternative maps of spatial factor are produced. In the work by Lilburne and Tarantolla (2009), the alternative spatial realizations are produced without regard for the spatial correlation of variables (i.e. raster grids of 10x10 resolution are produced based on uncorrelated and uniformly distributed spatial uncertainty in the range over each pixel) but the method's potential for applicability to spatially correlated factors is discussed.

This study builds on previous work by Lilburne and Tarantolla (2009) and proposes a combination of sophisticated spatial uncertainty models produced by SGS and the method of Sobol with an auxiliary input factor. The merging of these methods represents a powerful tool for GSA of spatially distributed computer models, as it allows for incorporation of spatial uncertainty in a computationally efficient way. Furthermore, since the method relies on detailed multivariate sampling of input factors' PDFs, UA can be performed on the outputs without additional computational cost.

Theory on Sequential Gaussian Simulation

Within the geostatistical framework, spatial distribution of an attribute is modeled by a random function (RF), i.e., a collection of *J* spatially dependent random variables (RVs) *Z (x)* defined at *J* locations in a domain. A set of *I* existing, spatially distributed measurements is viewed as one potential realization of the RF model at *I* sampled locations. The purpose of geostatistical analysis is to provide the estimate for an attribute at (*J-I*) unsampled locations. The uncertainty about any unsampled attribute value z(x) can be modeled probabilistically by local conditional cumulative distribution function (CCDF), specific for a given location x. This local posterior CCDF is an updated version of global (prior) CDF, and is conditioned on the joint outcomes of nearby RVs (neighboring data). The random function's spatial variability is described by a variogram model, defining dissimilarity between random variables located at any two locations, separated by a given distance (Goovaerts, 1997).

Kriging is the most popular geostatistical estimation technique that estimates quantity at a given location as a weighted sum of the adjacent measured points. Weights depend on the exhibited correlation structure (variogram). Kriging provides the best linear local estimates, that display a lower variation than the investigated values. Therefore Kriging estimates cannot reproduce the natural spatial variability of the real media. (Goovaerts, 1997) and Kriging maps fail to represent natural heterogeneity (Goovaerts, 1997). Furthermore, the series of local posterior uncertainty models, estimated by Kriging, cannot simultaneously assess the spatial uncertainty (joint multi-point uncertainty) (Goovaerts, 2001; Kyriakidis, 2001), such as probability that z-values at a number of locations are jointly no greater than a critical threshold (Goovaerts, 1997). Joint uncertainty models are required for assessing the impact of the uncertainty

in input spatial data on the uncertainty of model's outputs (Kyriakidis, 2001). Sequential Simulation (SS), on the other hand is able to reproduce natural spatial heterogeneity of variable and provides both the local - one-point - and spatial - multi-point- uncertainty about estimates. Sequential simulation maps reproduce spatial distribution of variable more realistic than kriging maps and, several equally probable stochastic realizations together, provide estimation of spatial uncertainty (Goovaerts, 1997).

Sequential Simulation provides values for unmeasured locations (nodes) in a domain. A sampling of the joint, multipoint RF model is replaced by a sampling of a sequence of one-point models along the random path visiting all nodes in a domain. To preserve the proper covariance structure between the simulated values, each point CCDF is made conditional not only to the original data but also to all values simulated at previously visited nodes. In this way an outcome of joint spatial model for multiple locations preserves the spatial autocorrelation structure.

Sequential Gaussian Simulation (SGS) is often used, among SS techniques, because of its relative simplicity and robustness (Deutsch and Journel, 1998). SGS uses the multi-Gaussian RF model (Goovaerts, 1997), i.e. it assumes that a joint distribution of RF model is multiple normal. This is a very congenial characteristic since, under assumption of multi-normality, the local CCDF can be fully described by only two parameters: mean and variance. To avoid erroneous results, the multi-normal assumption of data needs to be checked before SGS is performed. The RF also needs to be stationary within the domain for SGS to be applied correctly, i.e., the same global CDF is assigned for all locations. RVs at all domain nodes are assumed the same prior

CDF (the same mean and variance), therefore SGS should not be applied for data exposing trends, or preferential patterns.

The foundation of sequential simulation is Bayes's theorem and Monte Carlo (stochastic) simulation (King, 2000). The idea for SS is to trade the sampling of the J-point CCDF for the sequential sampling of the *J* one-point CCDFs (Goovaerts, 1997). The sequential simulation algorithm approximates a modeling of *J*-point CCDF by a sequence of *J* univariate (one-point) CCDFs at each node *J* along the random path. To preserve the proper covariance structure between the simulated values, each point CCDF is made conditional not only to the original *I* data but also to all values simulated at previously visited locations. For a given realization, value of an attribute assigned to location is selected randomly from the local CCDF.

The simulated CCDFs are conditioned both on measured data and previously simulated values. In order for simulated values not to overshadow the measured data, the measured and simulated data may be searched separately (two-part search) within the search radii (Deutsch and Journel, 1998). In theory every previously simulated value should be used for estimation of a value in a given node. In practice only the closest conditioning data is used, up to maximum number of previously simulated data or search radius to keep CPU time reasonable. This assumes that the closest data screens further data out, and the additional information from this screened data is small enough that it can be neglected.

Sequential Gaussian Simulation (SGS) is a robust and conceptually simple parametric method. In the SGS, properties of the RF model is assumed to be multivariate normal, therefore any local CCDF is also assumed Gaussian and can be

modeled using just two parameters: Kriging mean and Kriging variance. The first condition for RF to be multivariate normal is that its univariate CDF (sample distribution) is normal (Deutsch and Journel, 1998). If data distribution fails the normality test, it needs to be transformed to standard normal distribution. The most common technique is the normal scores (nscore) transform (Goovaerts, 1997), that is a graphical, rank preserving transformation (Deutsch and Journel, 1998) (Figure 3-1). Normal score transform is presented in equation 3-1 and a back-transform, required after analysis SGS analysis is presented in equation 3-2.

$$y(x) = \varphi\{z(x)\} \quad (3\text{-}1)$$

$$z(x) = \varphi^{-1}\{y(x)\} \quad (3\text{-}2)$$

Univariate normality is a necessary but not sufficient test of multiGaussian normality, the bivariate normality - the assumption that any two RVs is joint normally distributed - for the resulting nscore values needs to be checked as well (Deutsch and Journel,1998; Kyriakidis, 2001). If the assumption of bivariate normality is retained, data can be simulated using SGS, if not other sequential simulation techniques, like nonparametric Sequential Indicator Simulation (Deutsch and Journel, 1998; Goovaerts, 1997), should be applied for determination of local CCDFs (Goovaerts, 1997). The assumption of bivariate normality can be checked by comparing experimental indicator covariance values to those obtained from theoretical expressions of the bivariate normal distribution (Deutsch and Journel, 1998). In reality, environmental data are hardly ever normally distributed, therefore normal scores transformation is required. Simulation of normal scores is done most often with Simple Kriging (SK), using the normal score semivariogram and a SK zero mean (Deutsch and Journel, 1998; Goovaerts, 1997;

Isaaks, 1991). SK determines the mean of the local Gaussian distribution at a given location (SK mean) and its variance (SK variance). Once all normal scores are simulated, they are back-transformed to original variable's space.

SGS assumes maximum spatial entropy for a given variogram model (no correlation for extreme values of a variable). When the impact of spatially connected extreme values on the process response is known to be significant, like for the paths of connected high hydraulic conductivity, the nonparametric approach like Sequential Indicator Simulation should be used (Kyriakidis, 2001), SGS requires that data in simulated area come from a single underlying distribution (global CDF used for the nscore transform). Therefore trends are not always well reproduced in SGS. If present, trends should be filtered out from the data and residuals of the original values should be used for the analysis (Deutch 2002). Furthermore, the conditional simulation assumes the values at the conditioning points are free of error, and if the measurement error should be considered the method needs to be modified (Goovaerts, 1997).

SGS has also been applied for delineating areas susceptible to soil contamination, soil erosion (Delbari et al. 2009), vegetation delineation (King, 2000) and ecological risks (Koch et al., Rossi et al., 1993).

Theory on the Method of Sobol

The method of (Sobol, 1993) estimates the sensitivity indices (variances in Equation 1-1) by approximate Monte Carlo integrations. The procedure (Lilburne and Tarantola, 2009) begins with generating 2 matrices A and B, (N,k) of quasi-random numbers, where N is a selected integer and k is a number of input factors considered in the analysis; each row of the matrices represents a sample - a set of factors values used for model simulation. Further, the matrices D_i and C_i are defined from matrices A

and B. Matrix D_i is created from matrix A, except the column i^{th}, that is taken from matrix B, (where i=1,…k); matrix C_i is defined created from matrix B, except from the i^{th} column taken from matrix A (Figure 3-2). The three vectors of model outputs $\mathbf{y}_i$ of dimensions 1xN are obtained by running the model for each of the samples from matrices A, B, C_{ii}:

$$\mathbf{y}_A = f(A), \mathbf{y}_B = f(B), \mathbf{y}_{C_i} = f(C_i) \tag{3-3}$$

The method of Sobol estimates the Monte Carlo approximation for the first order sensitivity indices as follows:

$$S_i = \frac{V_i}{V} = \frac{\mathbf{y}_A \times \mathbf{y}_{C_i} - f_0^2}{\mathbf{y}_A \times \mathbf{y}_A - f_0^2} = \frac{\frac{1}{N}\sum_{j=1}^{N} \mathbf{y}_A^{(j)} \mathbf{y}_{C_i}^{(j)} - f_0^2}{\frac{1}{N}\sum_{j=1}^{N} \left(\mathbf{y}_A^{(j)}\right)^2 - f_0^2} \tag{3-4}$$

$$\hat{f}_0^2 = \left(\frac{1}{N}\sum_{j=1}^{N} \mathbf{y}_A^{(j)}\right)^2 \tag{3-5}$$

where: $\hat{f}_0^2$ indicates the estimated average for $\mathbf{y}_A$.

The total effects can be estimated from:

$$S_{Ti} = 1 - \frac{V_{-i}}{V} = 1 - \frac{y_B}{y_A} \times \frac{y_{C_i} - f_0^2}{y_A - f_0^2} = 1 - \frac{\frac{1}{N}\sum_{j=1}^{N} y_B^{(j)} y_{C_i}^{(j)} - f_0^2}{\frac{1}{N}\sum_{j=1}^{N} \left(y_A^{(j)}\right)^2 - f_0^2} \tag{3-6}$$

With a set of (2k+2)xN simulations the first-order index and total index is obtained for each input factor, where N is a size of a sample (the same as the selected integer for matrices generation), and k is a number of factors. Saltelli et al. (2005) recommends

using N of 500 – 1000. In practice, the size of N depends on the computational cost of the model. Models that are computationally intensive to run may constrain the analyst to select small N values (e.g. N 30–100), while more computationally efficient models can allow the analyst to use larger N values (e.g. N.500) (Lilburne and Tarantola, 2009). For a given model, the larger N the more precise sensitivity estimates are obtained, complex non-linear models may require larger N to obtain stable SA estimates (Crosetto and Tarantola, 2001, Lilburne and Tarantola, 2009). The accuracy of the estimates depends also on complexity of the model under analysis (degree of linearity, additivity, etc.) (Crosetto and Tarantola, 2001).

The quasi-random sampling scheme reduces the number of simulations required for accurate SA results (compared to the brute-force random sampling). Quasi-random numbers are generated from predefined probability distributions by quasi random sequences (Sobol, 1967) (the method of Sobol employs the LPt sequence of Sobol (Sobol, 1993)), that is very efficient method of sampling parameter input space that results in homogenous sampling of multivariate input space.

Variance-based techniques assume that input factors are independent. If this is not the case other more expensive methods are available (McKay, 1995). The assumption of independence relates to the errors of input factors and this hypothesis does not forbid the possibility of performing SA with spatially correlated error fields for given geographically distributed data (Crosetto and Tarantola, 2001).

The objectives of this chapter are to: 1) incorporate spatial uncertainty of numerical inputs into a generic, model-independent global UA/SA framework based on sequential simulation and variance-based sensitivity analysis techniques; 2) apply the

framework to evaluate the effect of spatial uncertainty of land elevation data on output uncertainty and parameter sensitivities of a complex hydrological model (RSM); and 3) evaluate an effect of objective functions selection (domain averaged/cell based) on GUA/SA results.

Methodology

Land Elevation Data as an Example for Spatially Uncertain, Numerical Model Input

Topography is potentially a very important factor for all distributed hydrological models. For example, a small degree of uncertainty in land elevation may have a relatively large effect on inundation model predictions (Wilson and Atkinson, 2003). Spatial representation of land elevation may be especially important in areas of relatively flat terrain, since small variations in these areas affect surface runoff routes (Burrough and McDonnell, 1998).

The common to South Florida landscape the Water Conservation Area 2A has unique characteristics such as: vast extent, very flat topography, dense vegetation, and a thick (20-30 cm) layer of debris floating over the bottom of inundated areas. The traditional methods for obtaining high resolution and high vertical accuracy elevation data like conventional field surveys or remotely-sensed technologies such as Light Detection and Ranging (LiDAR) and Interferometric Synthetic Aperture Radar (IFSAR)) are not effective in such conditions. Therefore an unique method was developed by the USGS for the land elevation surveying of South Florida conditions (USGS, 2003). The helicopter-based instrument, known as the Airborne Height Finder (AHF) was used for obtaining high vertical accuracy land elevation data. Using an airborne GPS platform and a high-tech version of the surveyor's plumb bob, the AHF system distinguishes itself from remote sensing technologies in its ability to physically penetrate vegetation and

murky water, providing reliable measurement of the underlying topographic surface (USGS, 2003). The elevation data has a vertical accuracy not smaller than +/- 15 cm (USGS, 2003). Regularly-spaced (approx. 400x400m) land elevation measurements are available for the WCA-2A. The total number of 1,645 data points was collected in 2003 for the area of study. The topography of WCA-2A exhibits a general North-South trend and (like that of the Everglades in general) is very flat. In WCA-2A land elevation decreases from approximately 3.7 m (North American Vertical Datum 1988, NAVD88) in the north to about 2 m NAVD88 in the south over a distance of 32 km (Figure 3-3).

As it can be seen in variogram constructed for raw land elevation values (Figure 3-4), the nugget effect is 0.0125 m^2. This is a part of the land elevation variability that cannot be addressed with the current dataset and can be attributed to the measurement error and variability at distances smaller than the sampling interval (the two types cannot be distinguished in practice). The resulting standard deviation (approximately 0.11 m) is smaller than the anticipated measurement error of the USGS, AHF data (USGS, 2003).

The RSM simulations in this study were performed for a period of 18 years (January 1983 to December 2000) with a daily time step. A one-year warm-up period (1983) was chosen to reduce the influence of the initial conditions on the model outputs. Raw model outputs included time series of water depth for each cell.

Implementation of Sequential Gaussian Simulation

The workflow for the creation of spatial realizations, using SGS from measured data is presented in Figure 3-5. The steps involved in the SGS include (Deutsch and Journel, 1998; Nowak, 2005; Zanon and Leuangthong, 2005): 1) a regular data grid for which the values are to be estimated (*J* nodes) is defined and measured values are

assigned to closest grid cells; 2) a random path to visit each of the (*J* - *I*) grid nodes is generated, each node is visited just once; 3) at each node: a) measured data and previously simulated values are located within the specified neighborhood, b) the local Gaussian CCDF is defined, c) the local CCDF is sampled randomly in order to obtain simulated value for the node; 4) a successive node in the random path is visited and the procedure from step 3 is repeated, until all nodes are simulated. The above steps constitute a single realization of the procedure (one map). Multiple realizations are obtained by repeating the procedure using different random paths.

Land elevation is considered as an example of spatially distributed factor in the GUA/SA in this work. The abundance of measured land elevation data enables construction of a reliable model of the spatial variation (variogram) and global histogram for the simulations. Because of the requirement of stationarity, land elevation data (showing a North-South trend) (as seen in Figure 3-3) needed to be de-trended before the procedure is applied. For this purpose the second order polynomial model, as a function of the Y-coordinate was fitted to the data (R^2=0.79) (Figure 3-6, A) and residuals were calculated for each data point (Figure 3-6, B). Table 3-1 presents a summary of descriptive statistics for land elevation residuals. The assumption of normality of residuals is checked using the Kolmogorov-Smirnov normality test. The test results in a significant (low) p-value of 0.0016, indicating that residuals are not normally distributed at confidence level α=0.01. Therefore, a normal score transform is required. A given residual value and its normal score correspond to the same cumulative probability of residuals' CDF and standard Gaussian CDF, respectively (as illustrated in Figure 3-1). The omnidirectional semivariogram model was fitted to the experimental

semivariogram of the normal scores of elevation residuals (Figure 3-7). The omnidirectional variogram for residuals appears to be trend-free as it reaches the sill. As expected, the sill is equal to unity, i.e., the variance of a standard Gaussian distribution. The variogram model had a nugget of 0.59 (dimensionless) and two structures: exponential with sill contribution of 0.25 and range of 5,3 km; and Gaussian with sill contribution of 0.16 and range of 12 km. Anisotropic variograms were also calculated (not shown) for four directions with 45° angular increments and ±22.5 angular tolerance. The results showed no significant directional behavior of autocorrelation.

SGS was performed for land elevation data using the SGSIM routine in the GSLIB Geostatistical Library (Deutsch and Journel, 1998). Numerous (L=200) alternative land elevation scenarios were produced for land elevation over the WCA-2A domain and stored for the subsequent GUA/SA. This number was considered to be sufficient to characterize the overall uncertainty of land elevation maps, based on comparison of results for L ranging from 30 to 500. In this study, no change in SGS results was observed for L>200. Successful practical implementation of the SGS algorithms is conditioned on the setting choice that can affect analysis results and associated CPU requirements. The order of visiting nodes in the SGS algorithm was selected randomly to minimize its influence on the final model (Zanon and Leuangthong, 2005). SGS uses simple kriging (SK) with zero mean and isotropic nscore variogram model for interpolation of nscore values onto 200x200 m grid (approx. half of the measured data density). At each simulation node, the local uncertainty is determined by using 10 of neighboring simulated nodes, and 10 neighboring values of point data within 10km radius (the approximate range of the nscore variogram).

After SGS, each of the alternative realizations was aggregated to the RSM mesh scale. For this purpose, the model mesh was overlaid over the 200x200m grid generated by SGS. Values for SGS nodes that contained centroids of RSM triangular cells were extracted and used as effective land elevation values for model cells. The continuity between land elevation values for neighboring RSM cells was maintained since the centroids' values were conditioned on the measured data and SGS simulated values within the search radii. Equiprobable SGS realizations of elevation maps, aggregated to the model scale, were used as alternative inputs for RSM runs. Cell-by-cell comparison of 200 aggregated maps of land elevation provided a PDF of land elevation values for each model cell, from which estimation variance, confidence intervals, and other desired statistics were derived. The estimation variance for land elevation of model cells ranges from 0.006 m^2 to 0.027 m^2 and is 0.01 m^2 on average. The average 95%CI for all mesh cells is 0.38 m and ranges from 0.3 m to 0.59 m.

Linkage of SGS with the GUA/SA

A multi-step procedure for GUA/SA allowing for the incorporation of spatially distributed factors is presented in Figure 3-8. In the case of spatially distributed inputs, alternative pre-generated maps were at first associated with an auxiliary scalar input factor (step 1). The auxiliary input factor was characterized by a discrete uniform distribution, with the number of levels corresponding to the number of maps. For spatially lumped factors this first step was omitted and the procedure started with the definition of uncertainty model (PDFs) of scalar values (step 2). In the following step (3), numerous model runs were performed for alternative input sets, generated based on PDFs of input factors, and corresponding model outputs were mapped. Next, empirical probability distributions with desired uncertainty measures (variance, confidence

interval) were obtained for model outputs (step 4). As a final step (5), GSA was performed using the method of Sobol.

For the current study, an auxiliary factor *topo* with discrete uniform distribution (*topo* ~DU [1,200]), was associated with the 200 land elevation maps produced by SGS. This input factor was used to investigate the effect of spatial structure of land elevation maps on model output uncertainty. Other inputs were considered as spatially certain and assigned uncertainty models based on available information for south Florida wetland conditions (based on literature review and experts opinion), using the approach presented in Chapter 2 (Table 3-2). All 20 uncertain input factors were sampled pseudo-randomly (by Sobol sequences) with a sample size N = 512. This required a total of 21,504 simulation runs, i.e. (2k+2)N runs, where k – number of factors. The matrix of corresponding model results was obtained and empirical PDFs for model objective functions were constructed. The uncertainty of the model output was expressed by the 95% confidence interval (95%CI, i.e., the range between 2.5 and 97.5 percentiles) of the empirical distribution. Finally, the GSA was performed using the method of Sobol to obtain the first-order and total effect sensitivity indices.

Selected raw RSM outputs are spatially and temporally distributed; for example, water depth is calculated for each cell on a daily time step. The MC based GUA/SA procedure requires that one value for each output objective function is provided for each simulation. The RSM performance objective functions (aggregated raw outputs) chosen as metrics for GUA/SA for this study are the performance measures generally adopted in the Everglades restoration studies (SFWMD, 2007): annual hydroperiod (specified as fraction of a year that a given area is inundated); annual water depth amplitude; and

annual mean, minimum and maximum water levels. The values for objective functions were averaged so that a single value was obtained for the whole simulation period. Raw results were post-processed, using Linux scripts, following two approaches: 1) spatial averaging over the application domain (spatial and temporal average of raw outputs); and 2) benchmark cells (temporal average of raw outputs). Among the 14 benchmark cells used for this study (Figure 2-1), three benchmark cells, representing different hydrological conditions, were selected for the illustration of UA and SA results. These are: cell 35 (in the north of domain), which represents dry conditions; cell 486 (in the south), which represents very wet conditions; and cell 178 (NE of the domain), which represents wet conditions and is of special interest because the NE area of the domain has experience cattail invasion (Figure 2-1). The two kinds of objective functions (domain-based and cell-based) may be used for supporting projects of various purposes and scale. In the case of the WCA-2A application, domain-based outputs may be effective for decisions of regional scale, like regional water budget assessment. Benchmark cell-based results provide information on local hydrological conditions. Therefore, this kind of objective functions may be more meaningful for supporting decisions on ecological restoration in particular locations of the WCA-2A.

The quality of sensitivity indices depends on the number of model runs; the more runs, the more accurate the results (Sobol and Saltelli, 1995). Best practice dictates that one should continue sampling until some stable sensitivity value is reached (Pappenberger, 2008). Convergence tests were performed (for N ranging from 672 to 43,008), and 21,504 simulations produced satisfactory GUA/SA results (results for 10,753 were also acceptable). Since computational cost of the analysis is high

(accounting that one model simulation takes approximately 3 minutes), the simulations for this study were performed using the High Performance Computing Center (HPC) at University of Florida. Batch jobs utilized on average 64 computational nodes simultaneously, making possible to obtain results for each analysis (i.e. 21,504 model simulations) in approximately 17 hours. Otherwise, one analysis would take approximately 45 days on a single PC.

Results

Uncertainly Analysis Results

The summary of UA results for all domain-based outputs and benchmark cells-based outputs is presented in Table 3-3. Domain-based outputs had relatively small variability when compared to cell-based outputs (Figure 3-9). For example, the distribution of the domain's mean water depth (Figure 3-9 A-B) had a 95% CI of 0.02 m (0.28-0.30) and the distribution for the domain's hydroperiod (Figure 3-9 C-D) had a 95%CI of 3% (79%- 82%). Such small uncertainty implies that for all alternative sets of input factor's used for RSM simulations, the domain's mean water depth and hydroperiod vary by only 2 cm and 3% respectively.

Uncertainty associated with benchmark-based outputs was approximately an order of magnitude higher than for domain-based outputs (Table 3-3, Figure 3-9). For example, for benchmark cell 178, the 95%CI for mean water depth for benchmark cell 178 was 0.28 m (0.16-0.44 m), and the 95% CI for hydroperiod was 14% (83%-98%). Similar magnitudes of variability regarding water depth and inundation periods were observed for other benchmark cells (Table 3-3).

The benchmark cell results are spatially variable and reflect general hydrological conditions in domain's regions. The simulation results are in agreement with previously

described hydropatterns in WCA-2A. As described by Romantowicz and Richardson (2008), water flows into WCA-2A from the north, likely causing the water depth at the northern boundary to increase rapidly. Later, it gradually disperses through the wetland As the water flows to the southern boundary it is impounded along the southern dike until flowing out of WCA-2A. Benchmark cells located in the south of domain have generally higher values for all objective functions (Figure 3-9), the cells located in the north have smallest values, objective functions for cells in NE oscillate between these extremes. The spatial hydropattern is also reflected in the uncertainty for benchmark-based outputs. Uncertainty results for mean water depth and minimum water depth are the highest for cells in the South of the domain (Figure 3-9 B and F). For example, the 95%CI for mean water depth is 0.49 m for cell 486 , and 0.28 m for cell 35 and cell 178 (Table 3-3). The uncertainty of hydroperiod is the highest for dry cells in the North (Figure 3-9 D), with a 95% CI for hydroperiod of 3%, 14% and 32% for cells 486, 178 and 35 respectively.

In order to compare deterministic and probabilistic approaches, the model was run for base values (i.e. default values from calibrated model) of the input factors, and unique values for model output are obtained (deterministic case). For the deterministic scenario, the domain's mean water depth is 0.29 m, and domain hydroperiod is 82%, for cell 178 the mean water depth is 0.23 m and hydroperiod is 94%. These values are very similar to the median values obtained for the output PDFs (Figure 3-10, Table 3-3). Figure 3-10 illustrates the difference in information obtained using deterministic and probabilistic approach. Vertical lines indicate results obtained for factors based on nominal/base values from Table 3-2.

Sensitivity Analysis Results

Figure 3-11 illustrates first-order sensitivity indices for domain outputs. The sensitivity measure S_i represents the contribution of a factor *i* to the total variance of domain-based objective functions (y-axis). The first-order sensitivity index ranges from 0 (completely unimportant input factor) to 1 (factor entirely controlling model output variance). A subjective criterion, used in this study, is that an input factor contributing less than 5% of total output variance is not considered important.

The most important factors for the majority of domain-based outputs were: parameter *det* determining detention depth, parameter *a*, used for calculation of Manning's roughness coefficient of mesh cells, and the auxiliary factor *topo* (Figure 3-11 A and Table 3-4). Detention depth is a depth of ponding in cell below which no transfer of water from one cell the other cell occurs, even if a hydraulic gradient exists. It represents water retained in small surface depressions with a cell. Moreover, the interception parameter *imax* contributed to variability of the domain's hydroperiod, and mean and minimum water depths, though to a lesser extent (Table 3-4). Manning's roughness coefficient for canals (*n*) contributed to the variance of maximum water depth and amplitude to a small extent (Table 3-4).

The auxiliary input factor *topo*, which represents the spatial uncertainty of land elevation, contributed to 19%, 21%, 13%, and 11% of the uncertainty domain mean water depth, minimum water depth, maximum water depth and amplitude of water depth respectively (Table 3-4). This factor was the second most important (after the parameter *a*) for the domain's mean water depth, and the third most important (after *det* and *a*) for the domain's minimum water depth.

While GSA results over the model domain indicated a shared importance between *topo, det, and a* (and other input factors, to a lesser extent), results for benchmark cell-based outputs showed that spatial uncertainty of land elevation had a dominant effect over all hydrological outputs for all benchmark cells. This factor contributed to the variability of model responses directly (without interactions) since its first-order sensitivity indices were above 90% for most cell-based outputs (Table 3-4). Figure 3-11 B- D presents SA results for the three selected benchmark cells. Other parameters used for the analysis were generally unimportant, with a few exceptions. Parameter *a*, contributes to 12 to 17% of variance of water depth amplitude for cells in NE of the domain (Table 3-4), including cell 178. Parameter *leakc* affects hydroperiod and amplitude in cell 486 (sensitivity indices are 15% and 6% respectively) and may reflect a local influence of a neighboring canal.

In case of domain-based and most benchmark cell-based outputs, higher-order effects for all factors are negligible (Table 3-4) as differences between total-order effects and first-order effects (S_{Ti} -S_i) of all factors are close to zero. This indicates that there are no indirect effects of input factors on output variance (interactions between factors in influencing output variance). The exception is hydroperiod for cell 178 and amplitude for cell 486, where small interactions are observed for factors *topo* and *det*, and *topo* and *leakc* respectively (Table 3-4).

Discussion

Preserving realistic land elevation is potentially very important in hydrological modeling, as it transfers into overland flow patterns in a domain. Especially for extensive wetland systems such as WCA-2A, which has a very low slope, even small changes in land elevation can affect water flow direction and hydrological patterns. The

hypothesized importance of spatial uncertainty of land elevation on RSM results was corroborated by GSA results. Despite exacting measurement of land elevation data, and reproduction of measured data histogram and variogram, the remaining “space” of spatial uncertainty, explored using random sampling, was large enough to affect model results. The auxiliary factor *topo* was relatively important for domain-based outputs, and it dominates cell-based model responses.

The results of this study showed that the choice of objective functions used for GUA/SA has significant impact on analysis results. The smaller variation of domain-based model response can be explained by two factors: spatial averaging of raw model outputs calculated for each cell over the entire domain; and the nature of the application itself. WCA-2A wetland is confined within levees, and inflows and outflows are controlled and considered as deterministic (i.e., fixed for all model runs). Therefore the only difference between simulations was the distribution of water within domain. In such a case, differences between spatially averaged outputs were small, and consequently, the uncertainty of predictions was smaller. The higher uncertainty for benchmark cell-based outputs was related to different water distribution patterns between model simulations resulting from alternative land elevation realizations.

GSA results depend also on the selection of objective function and help to explain UA results. The domain-based outputs were controlled mainly by the overland flow parameters: *a* used for calculating Manning’s roughness coefficient for mesh cells and *det*, determining detention depth, while *topo* had a smaller contribution to uncertainty. On the other hand, benchmark cell-based outputs were controlled almost completely by the spatial uncertainty of land elevation.

Information obtained by GUA/SA should support decision making process. With UA results, transparency in the model results and assessment of model uncertainty can effectively support the decision process, rather than simply acknowledging that a model is associated with existing, but undefined, uncertainty. For example, RSM results could be used as a decision support tool for restoration of sawgrass communities in NE region of the WCA-2A. This area (Figure 2-1) was originally dominated by a sawgrass community, but is experiencing an expansion of cattail due to anthropogenic changes of hydrological conditions and nutrient loads (Newman et al., 1998). Regarding hydrological controls, sawgrass has higher capacity to resist cattail invasion in shallow waters with more variable hydroperiod (Newman et al., 1996; Urban et al., 1993). For the purpose of this example, mean water depth of 24 cm is assumed to be a threshold between sawgrass-favorable hydrological conditions (shallower water) and cattail-favorable hydrological conditions (deeper water), since water depth above 24 cm is reported as optimal for cattail (David 1996, Grace 1989). If only deterministic RSM results for benchmark cell 178 are taken under consideration (Figure 3-10, A) one may decide that hydrological conditions in this location are favorable for sawgrass restoration since mean water depth for 18-year-long simulation is 23 cm. However, if the whole PDF of mean water depth is to be considered, it can be seen that approx. 60% of output values exceed the threshold of 24 cm. Therefore probabilistic analysis could lead to conclusions that cattail invasion is encouraged by existing hydrological conditions. A similar illustration could be done for any other location in a domain, for example benchmark cell 35 (located north of domain), that does not exhibit favorable hydrological conditions for cattail expansion for approx. 70% (Figure 3-10, B) of

simulated values. The example illustrates how neglecting the variability of model predictions may lead to incorrect management decisions. The combined GUA/SA methodology, apart from providing estimation of model uncertainty, can identify the controls of hydrologic system and indicate model inputs that control model performance.

Several processes simulated by the RSM model can potentially affect hydrological patterns. From the set of processes modeled by RSM, overland flow is found to be the most important in respect with the selected objective functions in this analysis. If the model uncertainty is not acceptable, the important input factors could be better estimated to reduce the model output variance. With GSA results, resources for additional data acquisition for reduction of model uncertainty can be optimally allocated. For example, for the WCA-2A application, if variability of outputs was to be reduced, the additional measurements or parameter estimation efforts should focus on the overland flow parameters (*a* and *det*) or land elevation rather than, for example, transpiration parameters. Finally, first and total order sensitivity indices are very similar, indicating that input factors influence model outputs only by direct effects and interactions effects are weak, and that for the outputs selected RSM behaves as an additive model.

It is important to highlight that the SA results are not only specific to selected objective functions but also depend on the uncertainty (probability distributions) of input factors. Uncertainty models are generally constructed based on limited information. In the case of a sensitive factor, different uncertainty models would likely result in different sensitivity measures. Therefore the GUA/SA should be performed iteratively and uncertainty models for input factors (lumped or spatial) should be considered as dynamic and updated every time new information is available.

The proposed methodology for GUA/SA is model-independent. Application of the variance-based method of Sobol requires no assumptions on model behavior (does not have to be linear, monotonic), and both direct effects and interactions of factors are examined. The methodology presented in this study can be applied to any spatially distributed hydrological model if sufficient information for construction of a variogram model of spatially distributed inputs is available. Potential disadvantages of the framework are high computational requirements, amplified by computational cost of model simulations. If duration of model runs renders an application of variance-based methods too costly, a screening method (Campolongo et al., 2007; Morris, 1991) can be applied first, without consideration of input spatial uncertainty. The incorporation of an auxiliary input factor in a method of Sobol can be used not only for estimation of effects of spatial pattern, but also for evaluation of effects of various data scales (resolution) or aggregation techniques. It can also be applied for selecting best model structure (Lilburne and Tarantola, 2009).

Conclusions

Spatial uncertainty of model inputs has so far been omitted in the uncertainty analysis and global sensitivity analysis (GUA/SA) of hydrological models. The uncertainty regarding spatial structure of model inputs can affect hydrological model predictions and therefore its influence should be evaluated formally. The framework applied in this research enables for spatial uncertainty of model inputs to be incorporated into GUA/SA. The results of this analysis confirm that spatial uncertainty of model inputs (land elevation) can propagate through spatially distributed hydrological model and affect model predictions.

A geostatistical technique of Sequential Gaussian Simulation (SGS) was used for estimation of spatial variability of input factors. Alternative realizations of land elevation surface maps were realistic since measured data, global CDF (histogram) and variogram models were preserved. The method of Sobol, combined with an auxiliary input factor, allowed for incorporation of alternative maps into GUA/SA and an estimation of the effect of spatial variability on model uncertainty and sensitivity.

RSM, a spatially distributed hydrological model was used as a benchmark model for the framework application. Land elevation was used as an example of spatially distributed model input. The auxiliary input factor *topo* is associated with land elevation maps and represents spatial uncertainty of topography. Other uncertain inputs are considered as spatially lumped.

GUA/SA results depended on the objective function considered (domain-based and benchmark cell-based). Benchmark cell-based outputs were associated with higher uncertainty than domain-based outputs. For example, the 95%CI for mean water depth (used as uncertainty measure) was 0.02 m for the domain, and 0.28 m for benchmark cell 178. GSA results for majority of domain-based outputs indicated that the most important factors were parameters *a*, used for calculating Manning's roughness coefficient for mesh cells, and *det*, specifying detention depth. In the case of the domain's mean water depth, $S_a = 0.56$, $S_{det} = 0.13$ (where S_i –first order sensitivity index for factor *i*, measures contribution of this factor to total output variance). The factor *topo* also contributed to the variability of domain-based outputs to a considerable extent (S_{topo}=0.19 for mean water depth). The GSA results for benchmark cell, on the other hand, showed that the factor *topo* practically dominated uncertainty of cell-based

outputs for all benchmark cells (S_{topo} > 0.9 for most cases), whereas other parameters have marginal and local influence on the cell-based outputs.

The framework, based on combination of SGS and the method of Sobol, could be applied to any spatially distributed model, as it is independent from model assumptions. GUA/SA evaluates suitability of the model as a decision support tool by specifying model uncertainty. The framework identifies areas in model input space that need additional research (additional measurements, parameter estimation). With spatial uncertainty, the analysis can also optimize spatial data collection for optimal reduction of model uncertainty.

Table 3-1. Summary for sample statistics of land elevation and land elevation residuals.

Sample Statistics	Land Elevation [m][1]	Residuals of Land Elevation [m]
Mean	3.043	0.002
Variance	0.091	0.014
Skewness	-0.528	-0.308
Minimum	1.740	-0.602
Median	3.060	0.007
Maximum	3.860	0.473

[1] NAVD 88.

Table 3-2. Characteristics of input factors, used for GSA/SA.

#	Input Factor	Base Value	Uncertainty Model (PDF)	Source
1	$value_{shead}$	3.66[1]	N[3](μ=3.66, σ=0.374)	Jones and Price, 2007
2	topo[2]	-	DU[3][1,200]	USGS, 2003
3	bottom	0	U[3](0.8, 1)	SFWMD data
4	hc	46.5	Lognormal(μ=4.6, σ=1.2)	SFWMD data
5	sc	0.3	U (0.2, 0.3)	SFWMD expert opinion
6	kmd	0.000026	U (0.000021, 0.000032)	± 20%
7	kms	0.000011	U (0.000009, 0.000013)	± 20%
8	kds	0.0000031	U (0.0000025, 0.0000038)	± 20%
9	n	0.06	Triangular (min.= 0.03, peak=0.10, max.=0.12)	SFWMD expert opinion; USGS, 1996
10	leakc	0.00001	U (0.000002, 0.001)	SFWMD data
11	bankc	0.05	U (0.04, 0.05)	SFWMD data
12	a	0.3	U (0.24, 0.36)	± 20%
13	det	0.03	U (0.03, 0.12)	Mishra et al., 2007
14	kw	1	U (0.8, 1.2)	± 20%
15	rdG	0	U (0, 0.2)	Yeo, 1964,
16	rdC	0	U (0, 1.5)	expert opinion
17	xd	0.9	U (0.7, 1.1)	Mishra et al., 2007
18	pd	1.8	U (1.5, 2.2)	± 20%
19	kveg	0.83	U (0.66, 0.99)	± 20%
20	imax	0	U (0, 0.03)	SFWMD expert opinion

[1] all input factors, except *topo*, have the same PDFs as in screening SA in Chapter 2;
[2] in this chapter factor *topo* is an auxiliary input factor, associated with pre-generated land elevation maps. Unlike in the Chapter 2, where *topo* represents uncertainty of land elevation error, here factor *topo* does not have any physical meaning.
[3] N - normal distribution; DU - discrete uniform distribution; U - uniform distribution;

Table 3-3. Summary of output PDFs for domain-based and benchmark cell-based outputs.

Output	Statistics	Domain	Benchmark cells		
			35	178	486
Mean Water Depth [m]	mean	0.29	0.18	0.27	0.91
	median	0.29	0.17	0.26	0.90
	2.50%	0.28	0.07	0.16	0.72
	97.50%	0.30	0.35	0.44	1.21
	95%CI	0.02	0.28	0.28	0.50
Hydroperiod [fraction]	mean	0.80	0.81	0.94	0.99
	median	0.80	0.83	0.95	0.99
	2.50%	0.79	0.60	0.83	0.97
	97.50%	0.82	0.92	0.98	1.00
	95%CI	0.03	0.32	0.14	0.03
Minimum Water Depth [m]	mean	0.07	0.04	0.08	0.46
	median	0.07	0.02	0.06	0.45
	2.50%	0.07	0.00	0.01	0.29
	97.5%.	0.08	0.17	0.23	0.75
	95%CI	0.02	0.17	0.22	0.46
Maximum Water Depth [m]	mean	0.67	0.45	0.80	1.43
	median	0.67	0.45	0.79	1.43
	2.50%	0.65	0.29	0.66	1.24
	97.50%	0.68	0.64	0.99	1.75
	95%CI	0.03	0.35	0.33	0.51
Amplitude [m]	mean	0.60	0.42	0.73	0.97
	median	0.60	0.42	0.73	0.97
	2.50%	0.58	0.29	0.63	0.94
	97.50%	0.61	0.50	0.81	1.00
	95%CI	0.03	0.21	0.18	0.05

Table 3-4. First-order sensitivity indices (S_i) for domain-based and benchmark cell-based outputs.

Output	Factor	S_i – domain*	S_i - cells			$(S_{Ti} - S_i)$ - domain	$(S_{Ti} - S_i)$ - cells		
			35	178	486		35	178	486
Mean Water Depth	topo	0.19	1.00	0.99	0.96	-	-	-	-
	a	0.56	-	-	-	-	-	-	-
	det	0.13	-	-	-	-	-	-	-
	imax	0.07	-	-	-	-	-	-	-
Hydroperiod	topo	0.05	1.00	0.94	0.79	-	0.02	0.06	0.03
	a	0.05	-	-	-	-	-	-	-
	det	0.38	-	-	-	-	0.02	0.04	-
	imax	0.40	-	-	-	-	-	-	-
	leakc	-	-	-	0.15	-	-	-	0.02
Minimum Water Depth	topo	0.21	0.99	0.99	0.96	-	-	-	-
	a	0.24	-	-	-	-	-	-	-
	det	0.41	-	-	-	-	-	-	-
	imax	0.05	-	-	-	-	-	-	-
Maximum Water Depth	topo	0.13	1.00	0.93	0.96	-	-	-	-
	a	0.81	-	0.06	-	-	-	-	-
	n	0.06	-	-	-	-	-	-	-
Amplitude	topo	0.11	1.00	0.74	0.88	-	-	-	0.06
	a	0.59	0.05	0.17	-	-	-	-	-
	det	0.15	-	0.05	-	-	-	0.02	-
	leakc	-	-	-	0.06	-	-	-	0.06
	n	0.07	-	-	-	-	-	-	-

* only sensitivity indices with values larger than 5% are presented, but all $(S_{Ti} - S_i)$ larger than 1% are shown

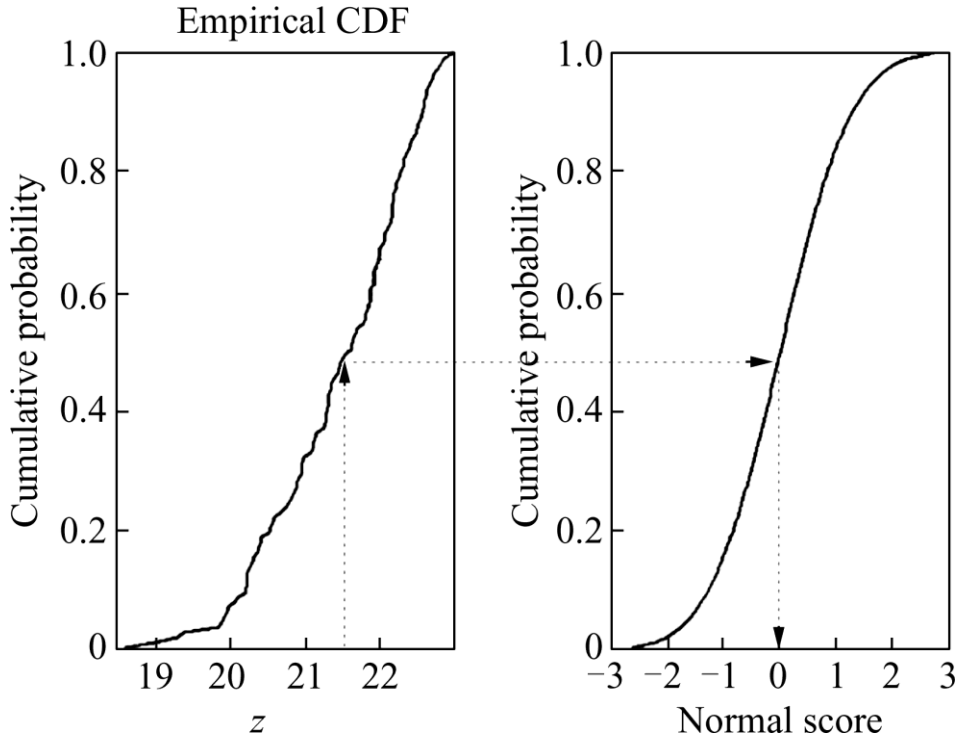


Figure 3-1. Transformation of an empirical cumulative distribution function to normal score (after Jingxiong et al., 2009).

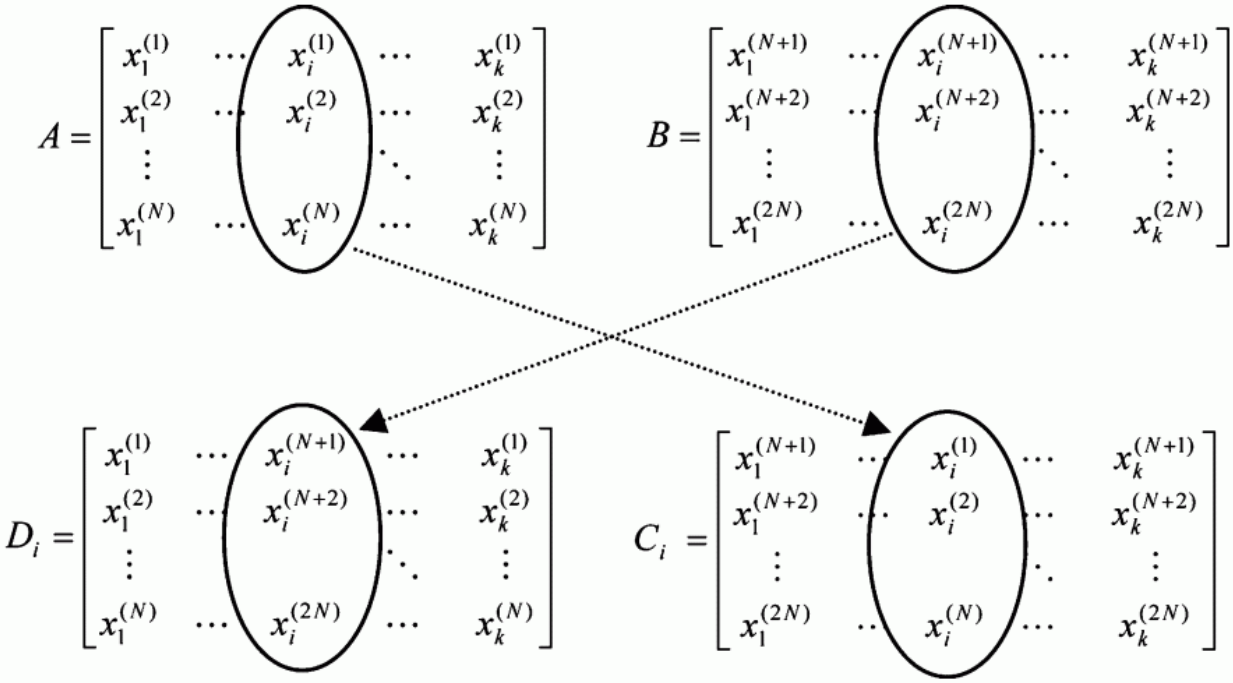


Figure 3-2. Generating matrices for the method of Sobol (after Lilburne and Tarantola, 2009).

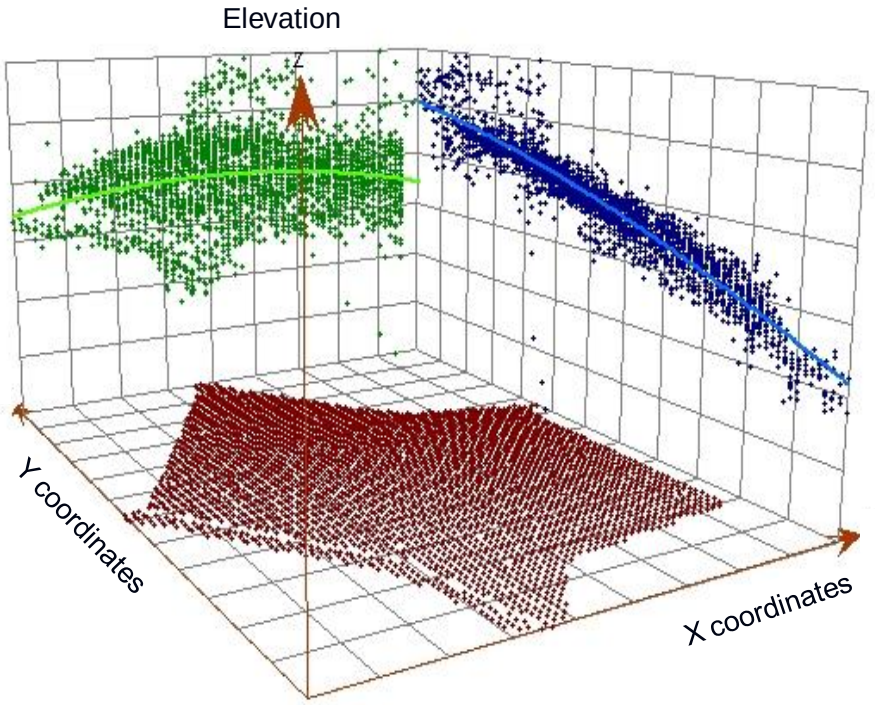
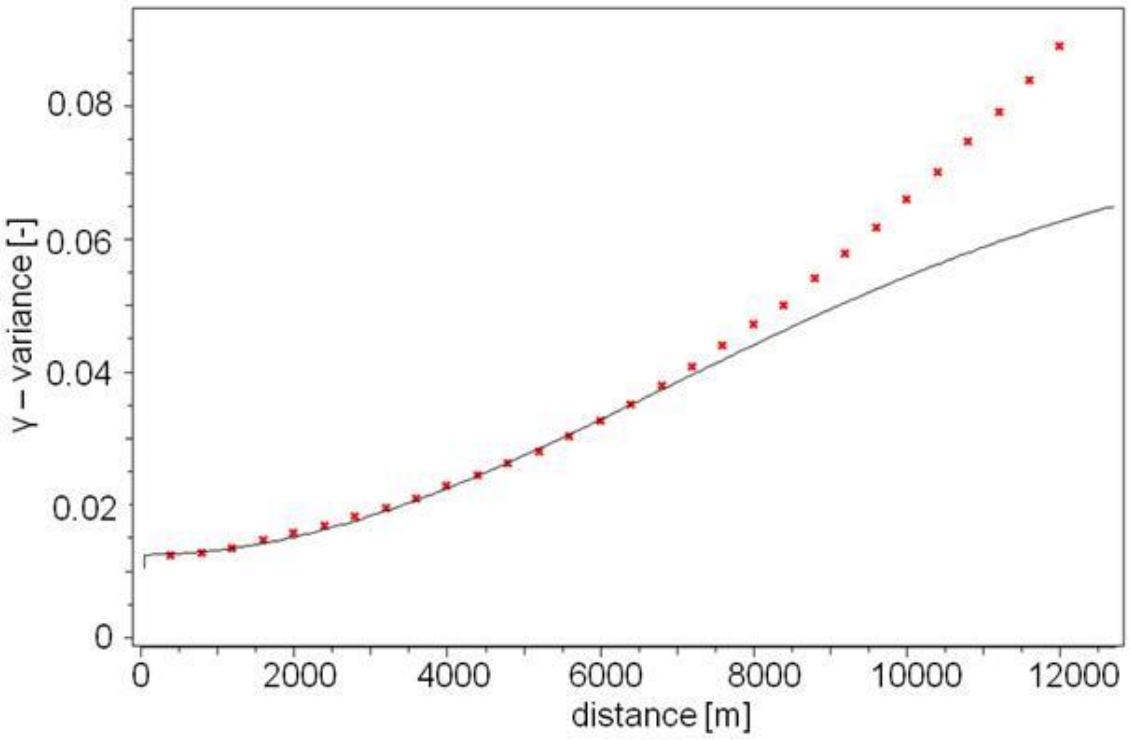


Figure 3-3. North-south trend in land elevation data for WCA-2A.



nugget = 0.0125 m^2, sill contribution=0.064 m^2, range = 16.8 km

Figure 3-4. Experimental variogram (dots) and variogram model (line) for raw land elevation data.

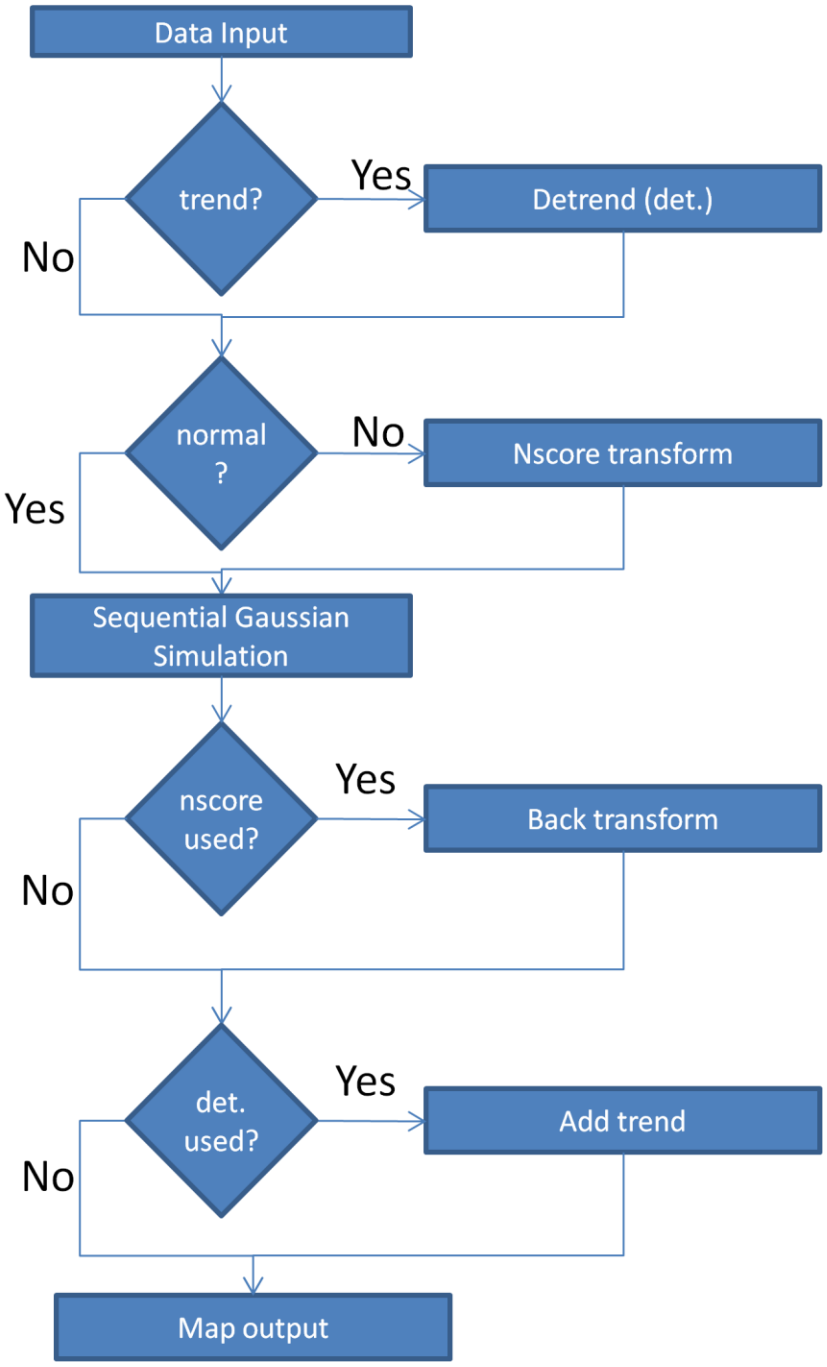
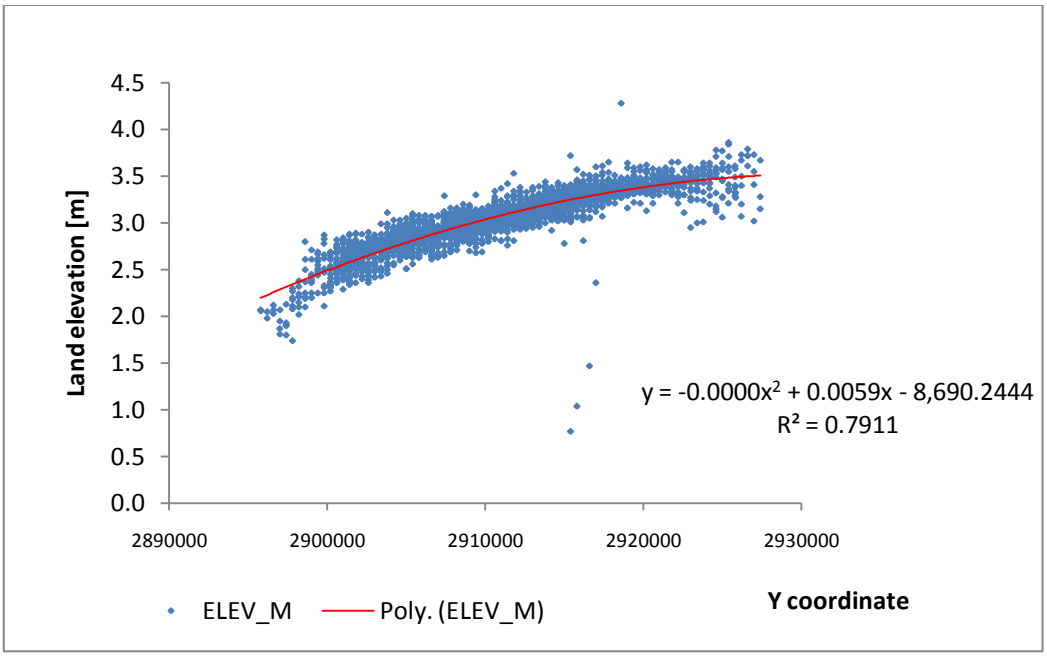
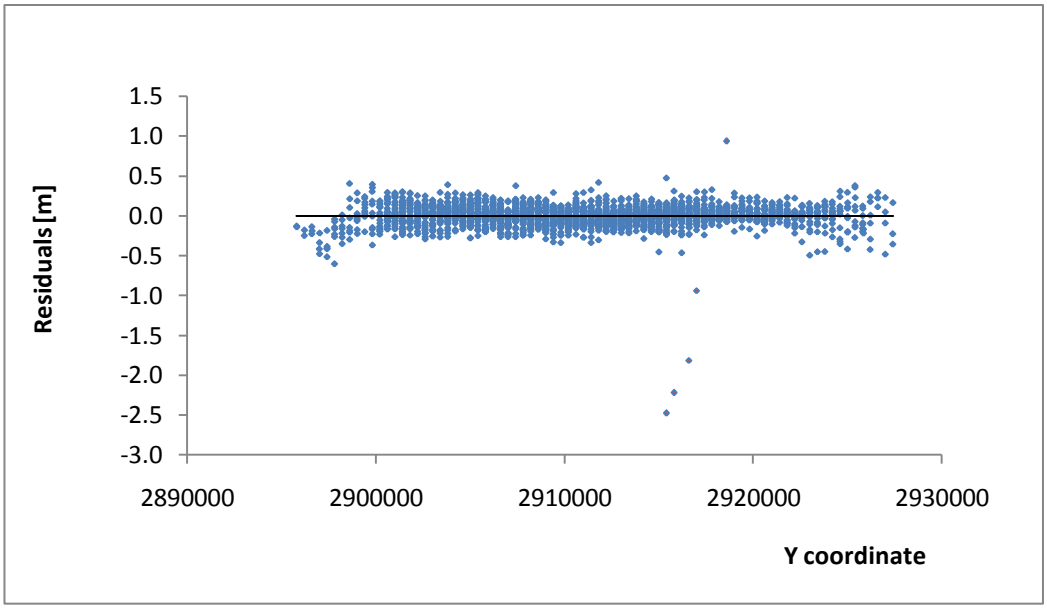


Figure 3-5. Workflow for generation of spatial realizations (maps) of spatially distributed variables from measured data, using SGS.



A



B

Figure 3-6. De-trending of land elevation data. A) polynomial trend fitted to original data as a function of Y coordinates, B) residulas obtained using the trend.

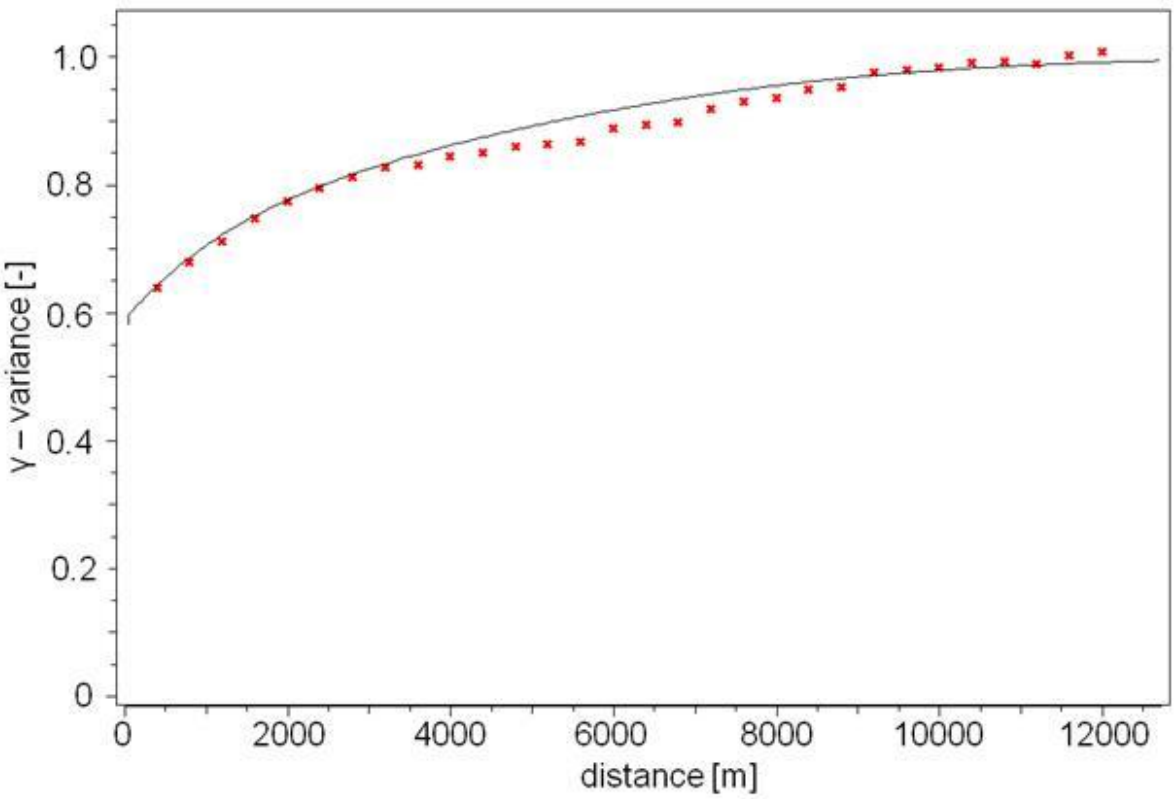


Figure 3-7. Experimental variogram (dots) and variogram model (line) for normal scores of land elevation residuals.

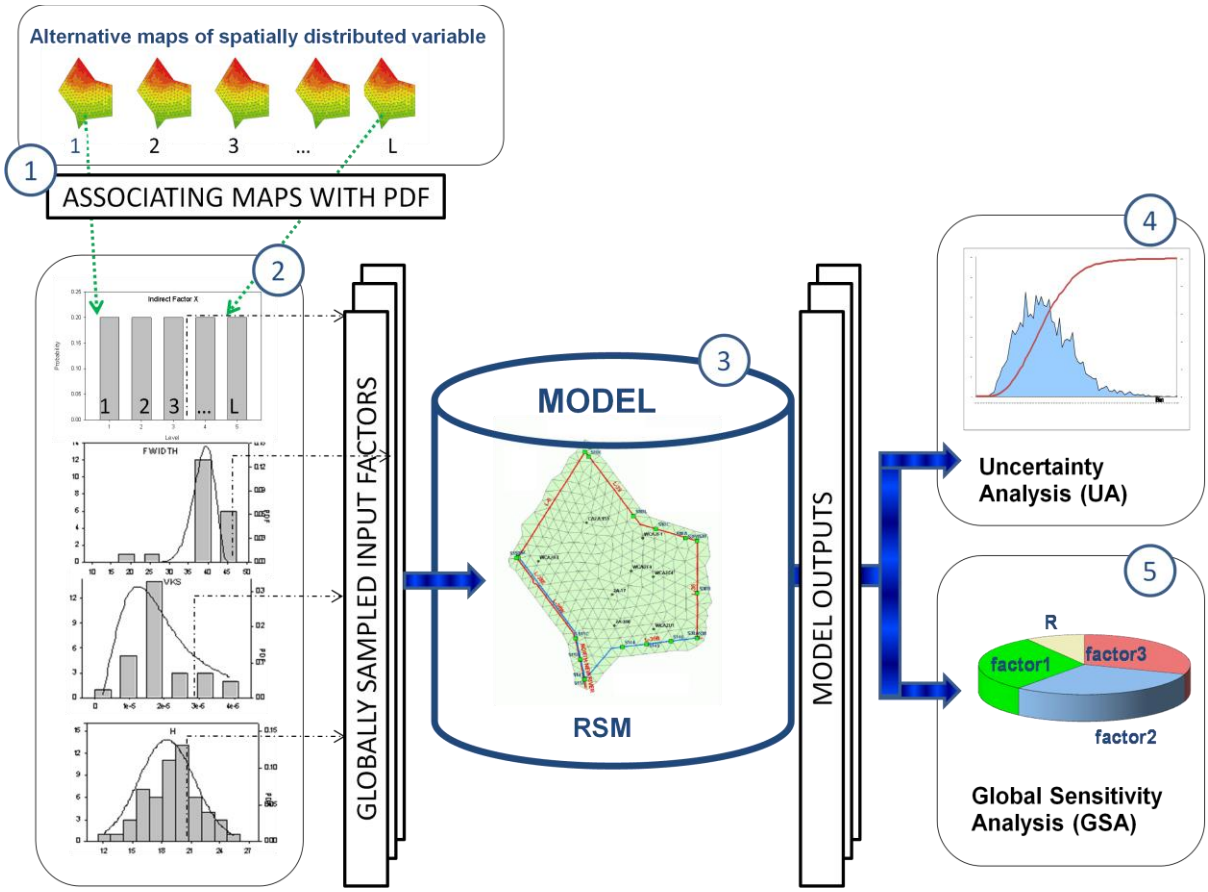


Figure 3-8. General schematic for the global sensitivity and uncertainty analysis of models with incorporation of spatially distributed factors.

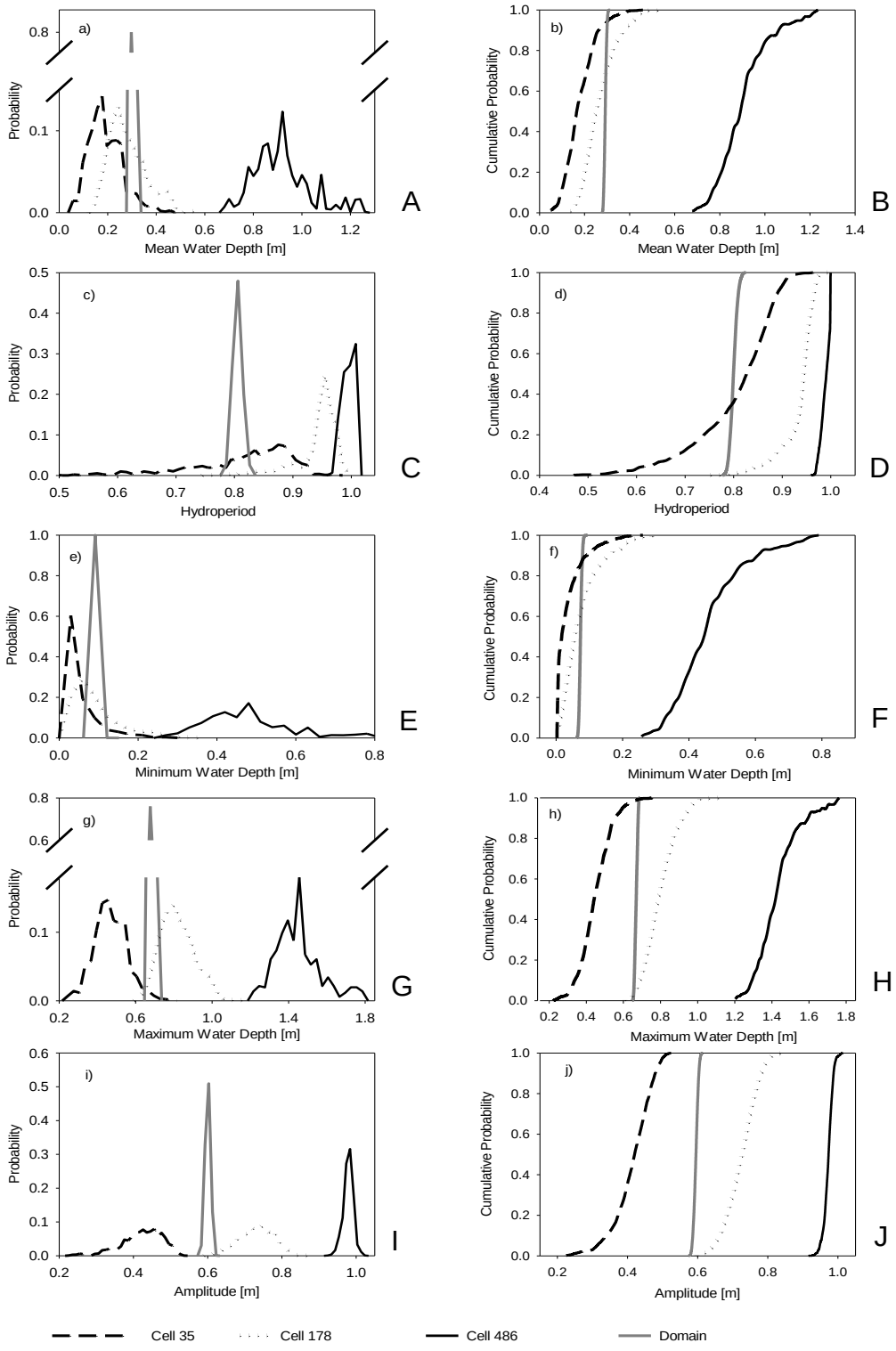
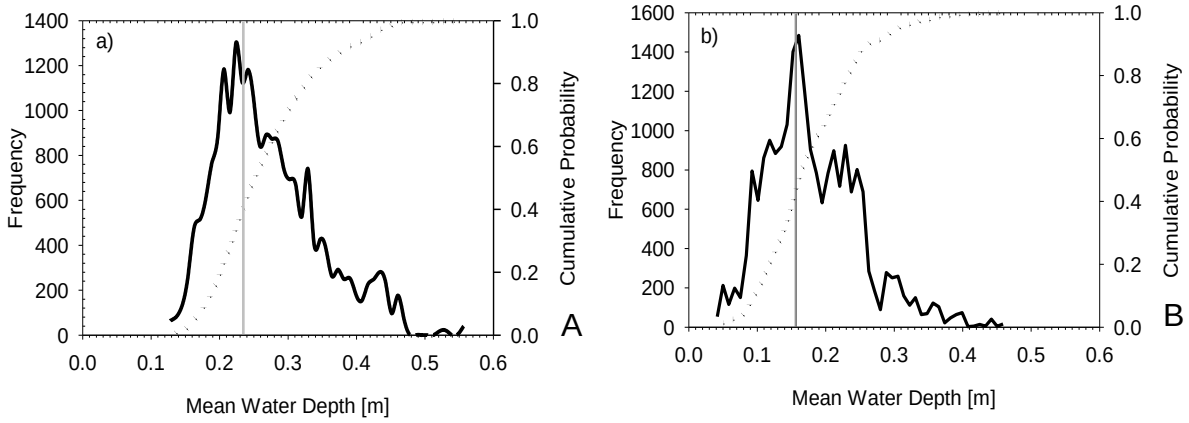


Figure 3-9. Uncertainty analysis results: PDFs (left) and CDFs (right) for domain-based and selected benchmark cell-based results. A), B) mean water depth, C), D) hydroperiod, E), F) minimum water depth, G), H) maximum water depth, I), J) amplitude.



vertical line – model results for base values of input factors
PDF and CDF – model results for 21,504 alternative sets of input factors

Figure 3-10. Comparison of deterministic (vertical line) and probabilistic (PDF and CDF) RSM results for benchmark cells. A) cell 178, B) cell 35.

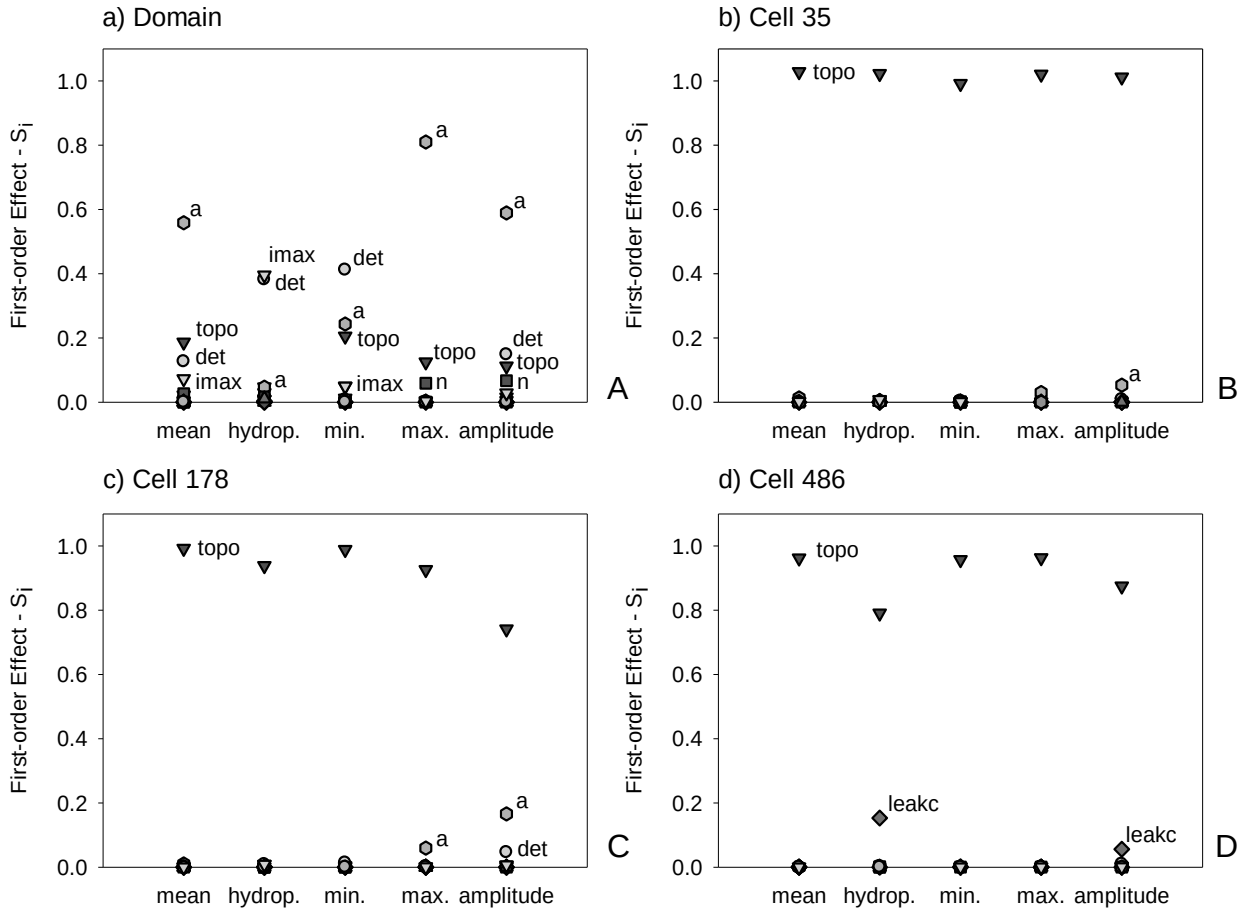


Figure 3-11. Sensitivity analysis results: first-order sensitivity indices (S_i) for domain-based and selected benchmark-cell based outputs. A) domain, B) cell 35, C) cell 178, D) cell 486.

CHAPTER 4
GLOBAL UNCERTAINTY AND SENSITIVITY ANALYSIS FOR SPATIALLY DISTRIBUTED HYDROLOGICAL MODELS, INCORPORATING SPATIAL UNCERTAINTY OF CATEGORICAL MODEL INPUTS.

Introduction

Categorical model inputs are widely used for hydrological and ecological model applications. Categorical model inputs are defined as non-numerical (nominal data) and include inputs like land cover, vegetation type and soil class. The environmental phenomenon is classified into discrete number of classes, which are often used to derive other model parameters. For example, vegetation type may determine the leaf area index or crop coefficient, and the soil type may determine hydraulic conductivity values. The study presented in this chapter aims at the exploration of the effect of potential spatial uncertainty in categorical model inputs on uncertainty of hydrologic model predictions. This study focuses on land cover type as an example of a spatially distributed categorical model input. The effect of land cover type on model uncertainty is evaluated simultaneously with other uncertain model inputs (including spatially uncertain land elevation) within the GUA/SA framework.

Model RSM cells are assumed homogenous in terms of land cover type. However, as it can be observed in Figure 4-1 (and Figure F-1 and F-2 in Appendix F) vegetation patterns may differ at the sub cell scales. Therefore, uncertainty regarding cell classification arises. The uncertainty may be further enlarged by the natural vegetation changes that are not accounted for by long term model simulations (vegetation maps are fixed)

The methodology applied for incorporation of spatial uncertainty of categorical model inputs, proposed in this study, is based on the general framework for

incorporation of spatial uncertainty. The framework incorporates the method of Sobol for the GUA/SA, and sequential simulation for generating alternative maps of model inputs. The difference between approaches for numerical data (described in Chapter 3) and categorical data is that instead of adopting the parametric framework (SGS) for modeling spatial uncertainty, the nonparametric (SIS) framework is used, as described in this chapter.

The spatial uncertainty of categorical data like land cover class was evaluated before (Kyriakidis and Dungan, 2001) using the geostatistical technique of SIS (Goovaerts, 1997). However, studies incorporating this uncertainty into GUA/SA of hydrological models have not been presented in the literature.

SIS of Categorical Variables

Categorical random variable (RV) s(u) can take K mutually exclusive and exhaustive outcomes/states $\{s_k, k=1,\ldots,K\}$ (Goovaerts, 1997). Every sample datum $s(u_\alpha)$ belongs to one and only one of the K classes, with no uncertainty. Within indicator formalism, each category is coded into an indicator variable $(\alpha;s_k)$. Indicator is set to 1 if the category/state s_k is observe at a given location α and to 0 otherwise:

$$i(\alpha;s_k)=\begin{cases}1 & \text{if } s(\alpha)=s_k \\ 0 & \text{otherwise}\end{cases} \qquad (5\text{-}1)$$

For given location α, the distribution (histogram) of categorical data is completely described by a frequency table, which lists K states and their frequency of occurrence (Goovaerts, 1997).

$$f(s_k=\frac{1}{n}\sum_{\alpha=1}^{n} i(\alpha;s_k) \qquad (5\text{-}2)$$

The pattern of continuity (variability) of category s_k, can be characterized by indicator semivariogram, computed as:

$$\gamma_I(\mathbf{h};s_k)=\frac{1}{2N(\mathbf{h})}\sum_{\alpha=1}^{N(\mathbf{h})}[i(u_\alpha;s_k)-i(u_\alpha+\mathbf{h};s_k)]^2 \quad (5\text{-}3)$$

The indicator variogram indicates how often two location a vector **h** apart belong to two different categories (Goovaerts, 1997). The smaller the $\gamma_I(\mathbf{h};s_k)$ the better spatial connectivity for class s_k.

Sequential Indicator Simulation (Gómez-Hernández and Sirvastava, 1990) can be used to model joint uncertainty of the spatial occurrence of categorical class labels e.g. the probability that a specific class prevails at a set of locations. SIS is the most commonly used non-Gaussian simulation technique (Goovaerts, 1997). The SIS procedure consists of generating multiple alternative realizations (maps) of class labels consistent with the available information (i.e. measured data at their locations, global histogram, and models of spatial variability), and determining the probability of class occurrence at more than one location (Goovaerts, 1997). The resulting realizations of class labels provide location dependent models of categorical data variability. Similarly, as in the SGS, the conditional PDF of the indicator RV is assessed by decomposing multivariate Conditional PDF (CPDF) into a product of N one point CPDF (using Bayes axiom) (Kyriakidis and Dungan, 2001). The local CPDF is estimated based on the conditional probability of occurrence of each category s_k, $[p(u_\alpha;s_k|n)]$ based on the conditioning information n (see SIS procedure steps in the methodology). The alternative SIS maps can be used to evaluate spatial variability of categorical data, and can be further used for evaluating model uncertainty and sensitivity due to this spatial uncertainty.

WCA-2A Land Cover

This study focuses on land cover as a spatially distributed model input, therefore the information on the study site that is presented in the previous sections is complemented here by more detailed land cover (vegetation) descriptions. The WCA-2A is a remnant Everglades area, consisting of vegetation communities dominated by sawgrass, with contribution of open marsh, cattail, shrubs and trees and other vegetation communities (Figure 4-2 A, Table F-1 in Appendix F). The vegetation patterns in the WCA-2A are affected by anthropogenic changes related to increased nutrient loads as well as altered water depth, hydroperiod, and flow. The major concern is an expansion of cattail to the areas previously occupied by sawgrass community (Newman et al., 1998), disappearance of tree islands as result of historically higher water depths (Wu et al. 2002), and to a much smaller extent, exotic species expansion (Rutchley et al., 2008).

The current application uses the 2003 baseline land-cover vegetation map of the WCA-2A for deriving input land cover map (Wang, personal communication). This land cover map was produced by the stereoscopic analysis of aerial photographs that allowed identification at species-level resolution for most of the grid cells (Rutchey et al., 2008). A hierarchical classification scheme, created specifically for use in the Comprehensive Everglades Restoration Plan (CERP) vegetation monitoring and assessment project (Rutchey et al., 2008) was utilized to label the grid cells. Each 50x50m grid cell was labeled with the major vegetation category observed within the cell. To verify the spectral signature of vegetation types on the photos with field conditions, a number of ground-truth (reference) sites were selected (Figure 4-2 B).

Constant vegetation pattern changes are reported to take place in the area. The reported rate of yearly spread of cattail is 960.6 ha/year from 1991–1995, and 312.0 ha/year from 1996–2003 (Rutchley et al., 2008). That is equivalent to an area of 8.7 and 2.8 average-size cells (1.1 km^2) per year for the first and second period respectively.

Methodology

The spatial uncertainty of land cover type is incorporated into GUA/SA, together with other input factors presented in Table 4-1. In this analysis land cover maps determine the spatial distribution of evapotranspiration (ET) parameters and the spatial distribution of parameter *a*, used for calculating Manning's n for model cells. ET parameters and parameter *a* maps are generated independently from each other. The two auxiliary input factors used for the GSA are factor LC, associated with land-cover-dependent ET parameters and factor *MZ*, associated with Manning's roughness zones (i.e. parameter *a* zones).

Implementation of Sequential Indicator Simulation

SIS is used for generating alternative class label realizations at the resolution of the land cover map. A realization form the multivariate CPDF is generated by a sequence of drawings from a set of univariate CPDFs. The SIS proceeds with the following actions (Goovaerts, 1997): 1) Transformation of each categorical datum $s(u_\alpha)$ into a vector of hard indicator data, (defined as in the equation 5-2); 2) Definition of random path visiting each undefined node in the domain; 3) At each node: a) Determination of the conditional probability of occurrence of each category s_k, $[p(u';s_k|n)]$ using indicator kriging (IK). The conditional information consists of both hard data and previously simulated nodes within the search radii centered on u'; b) Definition of the ordering of the K categories and constructing the CDF by adding the

corresponding probabilities of occurrence; c) Drawing a random number p uniformly distributed in [0,1]. The simulated category at location u is the one corresponding to the probability interval that contains number p; 4) Adding the simulated value to the conditioned data set and moving to the next model along the random path. In order to generate L realizations the above steps need to be repeated L times, using different random paths.

In the current study the SIS is performed using the class labels based on the reference data for the 2003 WCA-2A vegetation map (Figure 4-2 B). The original vegetation from ground truth data is assigned one of the five land cover types used in the current WCA-2A application, either sawgrass, cattail, cypress, freshwater marsh, and other, following the guidelines from the Vegetation Classification for South Florida Natural Areas (Rutchey et al. 2006). Figure 4-3 presents the frequency of 5 land cover classes, characterizing the global distribution used for SIS. The pattern of continuity of each of the land cover classes is presented using the indicator semivariograms (Figure 4-4). These semivariograms reflect patterns of spatial continuity (autocorrelation) and a range of spatial dependence for each land cover type. The variogram of sawgrass has a long range (approx. 10 km) and a larger scale of spatial variation, whereas variograms for cattail and cypress have short-range structures of spatial continuity. The long-range structure of the variogram for sawgrass is related to the vast extent of this vegetation class for the area. The smaller continuity of other classes can be possibly attributed to local conditions (like phosphorus concentration in case of cattail, tree islands for cypress). The variogram for marsh is very noisy and it appears as a pure nugget effect model (nugget effect is the same as sill). It suggests that the attribute is not spatially

structured. Possibly it is the effect of the inadequacy of classification (this class combines a lot of land cover types like marsh vegetation, shrubs, open water that does not have to be spatially correlated). Also the hard data locations may be a factor. These sites were chosen for referencing classification of satellite image, (i.e. for ambiguous rasters in the map) therefore they do not have to be representative for all of the vegetation classes considered here.

Geostatistical modeling is performed using GSLIB, SISM routines (Deutch and Journel, 1998). SIS is performed using the Simple Indicator Kriging algorithm. It uses 12 measured and 12 previously simulated points, within the search radius of 10 km. A number of 250 alternative land cover maps with 50x50m resolution is produced. The maps honor both the ground truth sites' class labels and indicator variogram models. Two example SIS realizations are shown in Figure 4-5. The simulated land cover maps exhibit patterns that are locally different from the 2003 vegetation map (for comparison see two realization for cell 178 in Figure 4-5 and the corresponding vegetation representation in Figure F-2 in appendix F). These discrepancies between the SIS realizations and the 2003 vegetation map are probably dictated by the fact that only reference data are used for the SIS (without using any image derived information).

The original land cover map, i.e. the map, used as an input for the calibrated RSM is presented in Figure 4-6. It can be seen that one of the 5 land cover classes is assigned to each of the model cells. In order to construct the land cover maps used as inputs for RSM, the 50x50 vegetation maps, produced by the SIS, need to be aggregated to the model scale. For this purpose the model mesh is overlaid over the SIS grid (in ArcMap) and the majority of pixels (class with the largest proportion within a

model cell) falling within a model cell determine which class is assigned to a model cell. The classes are crisp, which means that only one class can be assigned to a model cell for a given realization. Two aggregated maps are presented in Figure 4-7.

Associating RSM parameters with land use maps

The land cover maps are used to derive input values for model simulations. Land cover type can affect RSM outputs by: 1) determination of ET parameters, and 2) determination of parameter *a* (used for calculating Manning's roughness coefficient). Actual ET is calculated by the RSM based on the potential ET provided as input and the crop correction coefficient (*Kc*). The crop correction coefficient is evaluated based on other parameters: *kw*, *rd*, *xd*, *pd*, *kveg* and *imax*. The parameters are defined in Table 5-1 and illustrated in Figure B-1. Manning's roughness coefficient for mesh cells (n_{mesh}) specifies resistance to flow by vegetation for cells in the domain. It depends on the vegetation type (shape and texture of vegetation). Roughness varies greatly with the changes of density, height, flexibility of vegetation, and the relative ratio between flow depth and vegetative elements (Maidment, 1992). Because the geometry of plants is not uniform over the entire height of the plant, the resistance to flow changes with water depth and therefore is calculated for each model time step, depending on the water depth. For the purpose of this study, the Manning map is derived from a land cover map, by assigning each vegetation class a nominal Manning's roughness coefficient. The relationship between the land cover and Manning's roughness n, adopted here is presented in Table 4-2. It is assumed that there is no variation of vegetation density within the class (for example sparse, medium or dense cattail is considered as one type that is cattail). In reality, the density may vary within each land cover class but this is not addressed here and maybe a subject of further study.

ET parameters, as well as parameter *a*, are associated with two sorts of input factors for the GUA/SA. The first kind of input factor represents the uncertainty around the value of parameters for different zones. The first source of uncertainty was modeled in the previous chapters using the level parameter approach. The second kind of factor is related to the uncertainty regarding the spatial uncertainty (uncertainty about spatial distribution of zones within domain). The second source of uncertainty is examined in this chapter, with the use of the auxiliary factor *LC* for ET parameters and factor *MZ* for parameter a (i.e. Manning's roughness).

Implementation of the GUA/SA

A set of alternative maps of class labels (simulated realizations of land cover) can be input into the model and used for propagation of spatial input uncertainty onto model predictions. For each model run, one of the 250 land cover maps is randomly chosen and used as an alternative land cover input that translates into alternative realizations of ET parameters and Manning's n. The effects of alternative realizations are evaluated individually by two independent auxiliary input factors *LC* and *MZ*. Both factors have discrete uniform distributions: DU[1,250], with levels associated with the pre-generated land cover maps.

Four alternative scenarios (input factor sets) are considered for the GUA/SA (Table 4-3): 1) LC_la scenario. 2) MZ_la scenario, 3) VF_5a scenario, and 4) MZ_5a scenario. These scenarios differ in consideration of spatial uncertainty of land cover (*LC* - land cover is spatially variable and affects ET parameters through *LC* factor, *MZ* - land cover is spatially variable and affects spatial distribution of factor a through *MZ* factor, VF – land cover is assumed spatially fixed), and in the approach towards simulating parameter *a* (la- level approach, and 5a-approach based on five independent factors).

The level parameter approach is explained in the previous chapters (see Chapter 2 and Appendix C). Factor a2-a6, representative for zones II-VI are characterized by uniform distribution with ranges equal to ± 20% of base values (Table C-1).In the alternative “5a” approach each Manning’s n zone is represented by an independent factor *a* (*a2-a6*). In this way alternative maps of parameter *a* are no longer just shifted up and down (like in the level approach), but the spatial relationship between parameter values also changes. The GUA/SA results are provided for the domain-based outputs and the selected benchmark cell-based outputs: cell 35 in north, cell 180 in northeast, and cell 486 in south (Figure 2-1).

Results

Uncertainty Analysis Results

The comparative uncertainty results obtained for five input factors’ sets, described in Table 4-3 are presented in Figure 4-8 and Figure 4-9. It is observed that the approach applied for generating alternative values of parameter *a* (level or zone-based) affects uncertainty results for domain-based outputs (Figure 4-8 A). For domain-based mean water depth, maximum water depth and amplitude, the uncertainty is higher when the level approach is applied than for the zone-based approach. However, the differences in the 95%CI are not very high (as generally values for the 95%CI are not high in case of domain-based outputs).

The inclusion of the *LC* factor into UA does not seem to affect uncertainty results, i.e. there is not much difference in the 95% CI for the VF_la and LC_la scenarios. The incorporation of the *MZ* factor seems to increase the uncertainty of the domain-based mean and maximum water depth, compared to the spatially fixed land cover maps. This is observed for both the level and the zone-based approaches for generating alternative

values of parameter *a* (scenarios: VF_la with MZ_la, and scenarios: VF_5a and MZ_5a). The uncertainty results for cells-based outputs indicate that the uncertainty measures are very similar for the four scenarios considered (Figure 4-8 B-D).

Sensitivity Analysis Results

The GSA results show that factor *LC* is not important in respect to the domain-based outputs (Figure 4-10 A, Table 4-4). It indicates that the spatial distribution of ET parameters, conditioned on land cover maps, has negligible effect on the model outputs. ET factors were found to be negligible when they are considered as spatially certain (as presented in Chapter 3). Therefore the lack of importance of spatial variability of ET parameters on output uncertainty is not surprising. The GSA results for the scenario incorporating the *LC* factor are very similar to the previously obtained results for the spatially fixed land cover map (Figure 3-11 A).

The application of the GSA with incorporating factor *MZ* (for the MZ_la set) indicates that the spatial variability of the Manning's n zones have some contribution to the domain-based outputs (Figure 4-10 B). This factor contributes to the variance of mean water depth, maximum water depth and amplitude by 6%, 8%, and 7% respectively (Table 4-5).

Also for the scenario, based on the five individual a parameters for different Manning's n zones (the 5-a approach), factor *MZ* is found important (Figure 4-10 D). It contributes to 13%, 17%, and 9% of mean water depth, maximum water depth, and amplitude respectively (Table 4-7).

Independently form the land cover variability effects, it can also be observed that if the 5-a approach is used instead of the level parameter approach, the influence of this parameter is reduced significantly (compare Figure 3-11 A and Figure 3-11 C). The

reduction of parameter *a* importance is accompanied by the increase of first order sensitivity indices (S_i) for other important factors, for example the factor *MZ*, as described above. Out of the 5 *a* parameters, only *a6* (associated with cattail, Table 4-2) is important for the MZ_6a scenario (no variability of Manning's n maps). In the case when *MZ* is also considered, additionally to the 5 different parameters a (MZ_5a), two factors *a6* and *a5* seem to be of importance, together with factor *MZ*, associated with spatial variability of parameter *a* maps (Table 4-7). Similar to the results presented in Chapter 3, the factor *topo* dominates the uncertainty of all benchmark-cell based outputs. The example for cell 35 and scenario MZ_5a is presented in Figure 4-11.

Discussion

The global uncertainty and sensitivity analysis combined with the sequential indicator simulation enables quantification of the importance of spatial uncertainty of categorical model inputs in terms of model uncertainty and sensitivity. Furthermore, this importance is evaluated relative to the importance of other uncertain model inputs. The application of the GUA/SA with the SIS can indicate how significant the quality of spatial representation of categorical-type information is and therefore how much attention should be paid to preparation (collecting, pre-processing) of such data for modeling purposes. This study evaluates the importance of spatial representation of land cover type for modeling South Florida conditions with the RSM. Model input maps of land cover type are associated with uncertainty due data processing (up-scaling) but also due to the fact that vegetation cover is a dynamic phenomena that changes with time. The temporal variability of vegetation in a domain may introduce error, especially for long term simulations, as land cover maps used for as model inputs cannot account the land cover changes.

The land cover type is an important factor for ecological and hydrological model applications. The relative importance of land cover variability is evaluated in comparison to other factors, including spatial representation of land elevation. Therefore the main controls of the system may be determined.

The analysis of the domain-based indicates that spatial uncertainty of land cover type affects model outputs (domain-based outputs) by specification of Manning's n zones rather than by the ET parameters. Factor *MZ*, representing spatial uncertainty for parameter *a* (and therefore Manning's n zones) contributes significantly to domain-based outputs. While the importance of factor *LC*, associated with spatial representation of ET parameters is negligible. However, factor *MZ* is of smaller importance than some other uncertainty sources like the spatial uncertainty of land elevation that is represented by factor *topo*, or uncertainty about overland parameters' values, represented by factor *a*. The cell-based outputs are dominated by factor *topo* and the spatial representation of land cover type does not affect these outputs at all.

The lack of importance of factor *LC* indicates that the spatial distribution of ET parameters does not affect the selected RSM outputs for the WCA-2A application. Therefore it can be concluded that information requirements regarding the ET parameters can be relaxed, both regarding the value of these parameters and their spatial distribution. If a spatially distributed factor does not affect model uncertainty, there is no need to worry about the spatial structure much. For example in case of *LC* only rudimentary vegetation information would suffice. As long the parameters are within the conservative limits used for the specification of input factors in this study, there should not make much difference for model uncertainty.

The spatial distribution of parameter *a* for calculating Manning's roughness coefficient is somehow important for the domain-based model outputs (especially for the 5-a approach). Factor *a* is also reported as one of the most important factors for the domain-based outputs, especially for the level approach used for generating parameter a values (la). For the level approach, the actual values of factor *a*, assigned to particular zones, are more important than the spatial distribution of zones itself. In the case of the 5-a approach, when all 5 zones are associated with independent factors *a2-a5*, the influence of the spatial distribution of zones is similar to the effect of factors a5, and a6. Therefore, it can be observed that when the uncertainty about factor *a* values is reduced, the spatial distribution of zones becomes more relevant. For the 5a approach all factor *a* values (associated with different zones, i.e. land cover classes) are generated independently. Moreover, the values associated with different zones may overlap, which in some way accounts for similarity of vegetation densities between various classes (like sawgrass – factor a5, and cattail - factor a6). From all parameter *a* zones, only zones associated with sawgrass and cattail are important with respect to domain-based model outputs. This fact is probably related to the highest Manning's roughness coefficient values (the highest flow resistance) associated with these two land cover classes.

The results of this chapter provide an illustration of the significance of specification of uncertainty for factors used in the GUA/SA on the analysis results. In case of zonal factor *a* the level parameter approach seem to inflate the model output variance. The less conservative and probably more realistic approach is based on generating values of parameter *a* for different zones independently. Furthermore, it can be observed that

in the case of reduction of uncertainty of the most important factors, other factors gain importance. Generally, domain-based outputs are controlled to a larger extent by factor *a* (when the level approach is used). However, when the 5a-appraoch is used topography is the main factor controlling model outputs.

The conservative approach is used here for producing alternative land cover maps with the SIS in order to provide the “worst-case” uncertainty of spatial variability. Only ground truth points used for the reference of the source vegetation map (2003 vegetation map) are used for constructing alternative land cover realizations without any regard to the information in the vegetation maps itself. The uncertainty and sensitivity results could be smaller if hard data used for indicator Kriging was supported by soft, image derived information. In spite of this conservative approach land cover variability does not contribute much to model uncertainty. Therefore, it can be assumed that if additional information was used, the uncertainty would be even smaller. However it needs to be considered that the analysis presented in this chapter is of an exploratory nature. It aims at better understanding of model processes affected by land cover input maps.

Conclusions

The framework proposed in this chapter allows for spatial uncertainty of categorical model inputs to be incorporated into global uncertainty and sensitivity analysis (GUA/SA) by combining utilities of the variance-based method of Sobol and geostatistical technique of Sequential Indicator Simulation (SIS). For the purpose of this study it is assumed that land cover maps may affect model outputs by delineation of ET parameter zones, and Manning’s n zones. Five land cover classes, used in the application are externally associated with the corresponding Manning’s roughness

zones (i.e. parameter *a* zones). For both the Manning's n and ET parameters two types of uncertainties are considered independently: spatial uncertainty of parameter zones (related to spatial uncertainty of land cover classes), and uncertainty of parameters assigned to each of the zones. The ET factors, associated with each of the land cover classes, are varied within ranges based on the physical limitations, expert opinion, or ±20% of calibrated value, in case no other information is available. With these assumptions, the results of the analysis show that spatial uncertainty of land cover affects RSM domain-based model outputs through delineation of Manning's roughness zones more than through ET parameters effects. In addition, the spatial representation of land cover has much smaller influence on model uncertainty when compared to other sources of uncertainty like spatial representation of land elevation, or the uncertainty ranges for the parameter *a*.

Table 4-1. Characteristics of input factors, used for GSA/SA.

#	Input Factor	Base Value	Uncertainty Model (PDF)	Source
1	LC	-	DU[3][1,250]	SWFMD, 2001 vegetation map
2	MZ	-	DU[1,250]	SWFMD, 2001 vegetation map
3	$value_{shead}$	3.66[1]	N[3](μ=3.66, σ=0.374)	Jones and Price, 2007
4	topo[2]	-	DU[1,200]	USGS, 2003
5	bottom	0	U[3](0.8, 1)	SFWMD data
6	hc	46.5	Lognormal(μ=4.6, σ=1.2)	SFWMD data
7	sc	0.3	U (0.2, 0.3)	SFWMD expert opinion
8	kmd	0.000026	U (0.000021, 0.000032)	± 20%
9	kms	0.000011	U (0.000009, 0.000013)	± 20%
10	kds	0.0000031	U (0.0000025, 0.0000038)	± 20%
11	n	0.06	Triangular (min.= 0.03, peak=0.10, max.=0.12)	SFWMD expert opinion; USGS, 1996
12	leakc	0.00001	U (0.000002, 0.001)	SFWMD data
13	bankc	0.05	U (0.04, 0.05)	SFWMD data
14	a	0.3	U (0.24, 0.36)	± 20%
15	det	0.03	U (0.03, 0.12)	Mishra et al., 2007
16	kw	1	U (0.8, 1.2)	± 20%
17	rdG	0	U (0, 0.2)	Yeo, 1964,
18	rdC	0	U (0, 1.5)	expert opinion
19	xd	0.9	U (0.7, 1.1)	Mishra et al., 2007
20	pd	1.8	U (1.5, 2.2)	± 20%
21	kveg	0.83	U (0.66, 0.99)	± 20%
22	imax	0	U (0, 0.03)	SFWMD expert opinion

[1] all input factors, except *topo*, have the same PDFs as in screening SA in Chapter 2;
[2] in this chapter factor *topo* is an auxiliary input factor, associated with pre-generated land elevation maps. Unlike in the Chapter 2, where *topo* represents uncertainty of land elevation error, here factor *topo* does not have any physical meaning.
[3] N - normal distribution; DU - discrete uniform distribution; U - uniform distribution;

Table 4-2. Relationship between vegetation type and Manning's n.

Vegetation Type	Manning zone nr	a_{base} [1]	n_{base} [2]
Sawgrass	5	0.70	0.73
Cattail	6	0.90	0.94
Forest	2[3]	0.30	0.31
Freshwater marsh	4	0.50	0.52
Other	1	0.10	0.10

[1] a_{base}, and n_{base} are associated with n zone for the calibrated model;
[2] n_{base} values are calculated for the 0.29m (the median for the domain-based mean water depth distribution);
[3] zone 3 is missing here, it has value of a=0.34 (n=1.99), the value for zone 2 is assigned instead; which is related to the implementation of substituting scripts.

Table 4-3. Input factor scenarios used for the GUA/SA.

Land Cover Effect	Generation of parameter a	
	1 factor - level approach (1a)	5 individual factors (5a)
Land cover affects spatial distribution of ET parameters (*LC* factor)	LC_1a	-
Land cover affects spatial distribution of parameter a (*MZ* factor)	MZ_1a	MZ_5a
Land cove is considered spatially certain (VF)	VF_1a	VF_5a

Table 4-4. First order sensitivity indices for scenario: LC_la.

Input	S_i				
	Mean W.D[1].	Hydroperiod	Min. W.D.	Max. W. D.	Amplitude
$value_{shead}$	-	-	-	-	-
topo	0.19	0.06	0.25	0.15	0.17
bottom	-	-	-	-	-
hc	-	-	-	-	-
sc	-	0.04	-	-	-
kmd	-	-	-	-	-
kms	0.01	0.01	0.01	-	-
kds	0.03	0.04	0.07	-	-
n	0.04	0.02	0.01	0.07	0.06
leakc	-	0.01	-	-	-
bankc	-	-	-	-	-
det	0.13	0.39	0.37	-	0.13
kw	-	-	0.02	-	-
rdG	-	-	-	-	-
rdCY	-	-	-	-	-
xd	-	0.01	-	-	-
pd	-	-	-	-	-
kveg	-	-	-	-	-
imax	0.05	0.31	0.02	-	-
LC	0.01	0.04	0.01	-	-
a	0.54	0.04	0.24	0.78	0.62
Sum S_i	1.00	0.99	1.00	1.00	0.99

[1] W.D. – water depth

Table 4-5. First order sensitivity indices for scenario MZ_la.

Input	S_i				
	Mean W.D.	Hydroperiod	Min. W.D.	Max. W. D.	Amplitude
$value_{shead}$	-	-	-	-	-
topo	0.15	0.04	0.22	0.12	0.15
bottom	-	-	-	-	-
hc	0.01	0.01	0.01	-	-
sc	-	-	-	-	-
kmd	-	-	-	-	-
kms	0.01	0.01	0.01	-	-
kds	0.02	0.03	0.05	-	0.01
n	0.05	0.02	0.01	0.09	0.10
leakc	-	0.01	-	-	-
bankc	-	-	-	-	-
det	0.09	0.33	0.30	-	0.09
kw	-	0.02	0.03	-	-
rdG	-	0.02	-	-	-
rdCY	-	-	-	-	-
xd	-	-	-	-	-
pd	-	-	-	-	-
kveg	-	-	-	-	-
imax	0.09	0.42	0.07	-	0.02
MZ	0.06	0.01	0.04	0.08	0.07
a	0.52	0.04	0.26	0.71	0.56
Sum S_i	1.00	0.98	0.99	1.00	1.00

[1] W.D. – water depth

Table 4-6. First order sensitivity indices for scenario VF_6a

Input	S_i				
	Mean W.D.	Hydroperiod	Min. W.D.	Max. W. D.	Amplitude
$value_{shead}$	-	-	-	-	-
topo	0.33	0.04	0.25	0.36	0.21
bottom	-	-	-	-	-
hc	-	-	-	-	-
sc	-	0.03	-	-	-
kmd	-	-	-	-	-
kms	0.02	0.01	0.02	0.01	-
kds	0.04	0.03	0.06	-	0.03
n	0.05	0.01	-	0.17	0.13
leakc	-	0.01	-	-	-
bankc	-	-	-	0.01	0.01
det	0.22	0.41	0.48	-	0.26
kw	0.03	0.01	0.04	0.01	-0.01
rdG	-	0.02	-	-	-
rdCY	-	-	-	-	-
xd	-	-	-	-	-
pd	-	-	-	-	-
kveg	-	-	-	-	-
imax	0.13	0.41	0.07	0.02	0.05
a2	0.02	-	0.01	0.02	-
a3	0.03	-	0.01	0.04	0.01
a4	0.03	-	0.01	0.06	0.03
a5	0.04	-	0.01	0.04	0.01
a6	0.09	0.01	0.03	0.29	0.15
Sum S_i	0.98	0.99	0.98	0.96	0.94

[1] W.D. – water depth

Table 4-7. First order sensitivity indices for scenario MZ_6a.

Input	S_i				
	Mean W.D.	Hydroperiod	Min. W.D.	Max. W. D.	Amplitude
$value_{shead}$	-	-	-	-	-
topo	0.23	0.05	0.23	0.23	0.19
bottom	0.01	-	-	0.02	0.01
hc	-	-	-	-	-
sc	-	0.03	-	-	-
kmd	-	-	0.01	-	-
kms	0.02	0.01	0.02	-	-
kds	0.05	0.03	0.08	-	0.02
n	0.04	0.01	-	0.14	0.14
leakc	-	0.02	-	-	-
bankc	-	-	-	-	-
det	0.14	0.36	0.37	0.01	0.20
kw	0.02	0.01	0.03	-	0.01
rdG	-	0.02	-	-	-
rdCY	-	-	-	-	-
xd	-	-	-	-	-
pd	-	-	-	-	-
kveg	-	-	-	-	-
imax	0.11	0.43	0.07	0.01	0.04
a2	0.01	-	0.01	0.02	0.01
a3	-	-	-	-	-
a4	0.01	-	0.01	0.01	-
a5	0.18	0.01	0.08	0.22	0.10
a6	0.02	-	-	0.13	0.14
MZ	0.13	0.02	0.07	0.17	0.09
Sum S_i	0.98	1.00	0.98	0.96	0.94

[1] W.D. – water depth

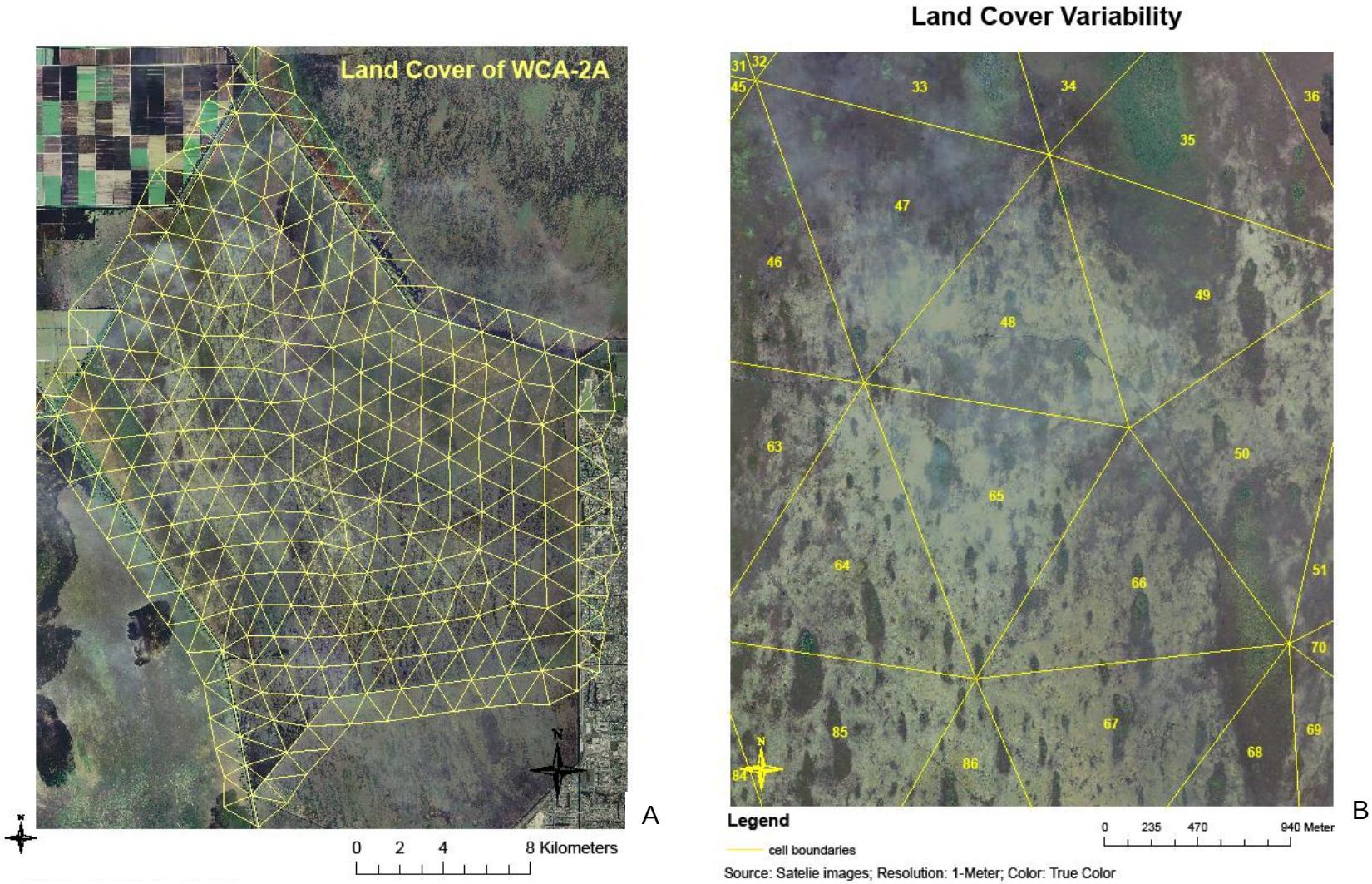


Figure 4-1. Land cover variability for WCA-2A with model mesh cells. A) whole model domain, B) magnified fragment.

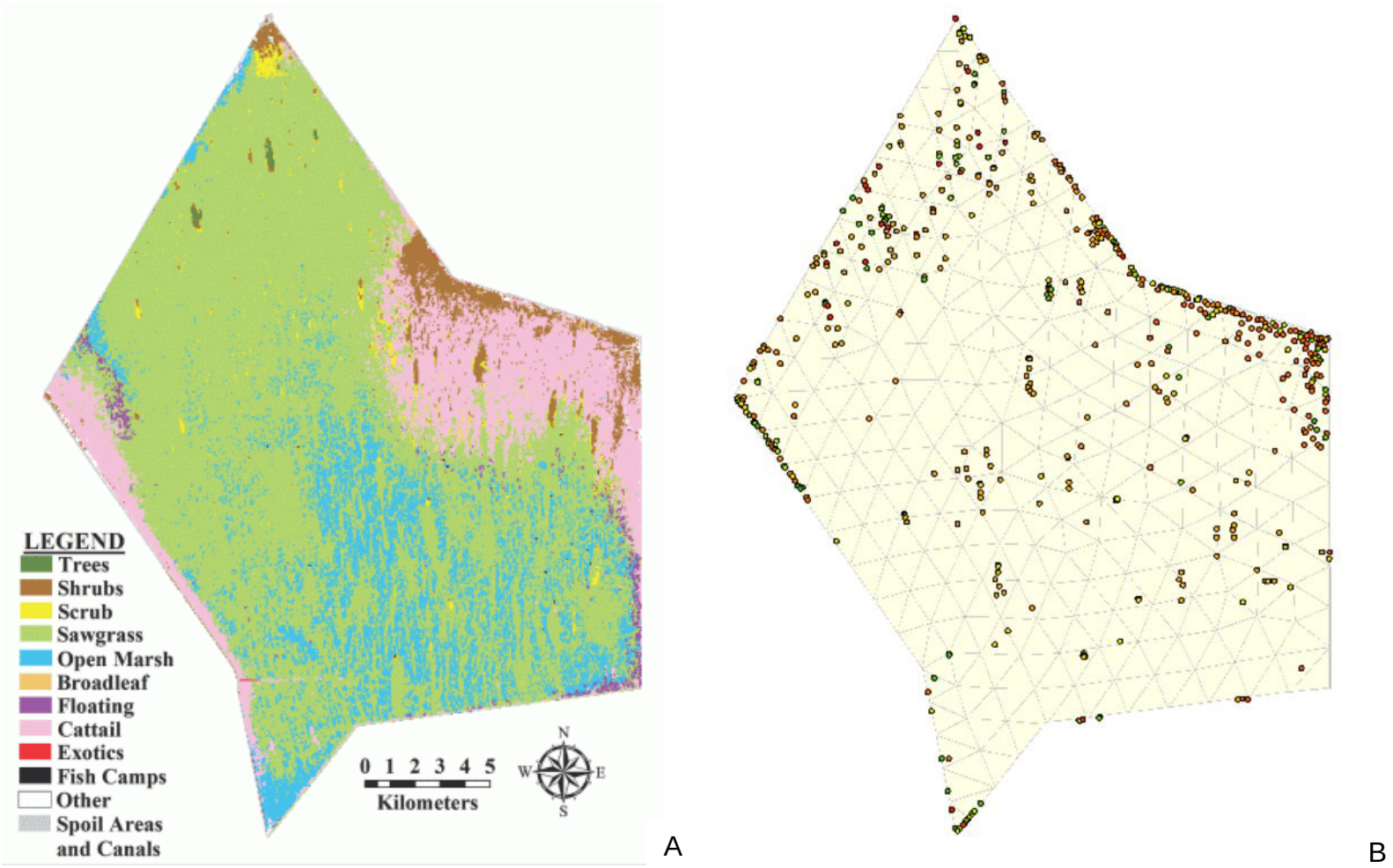


Figure 4-2. Vegetation at WCA-2A. A) Vegetation map (Rutchley, 2008), B) Location of ground truth.

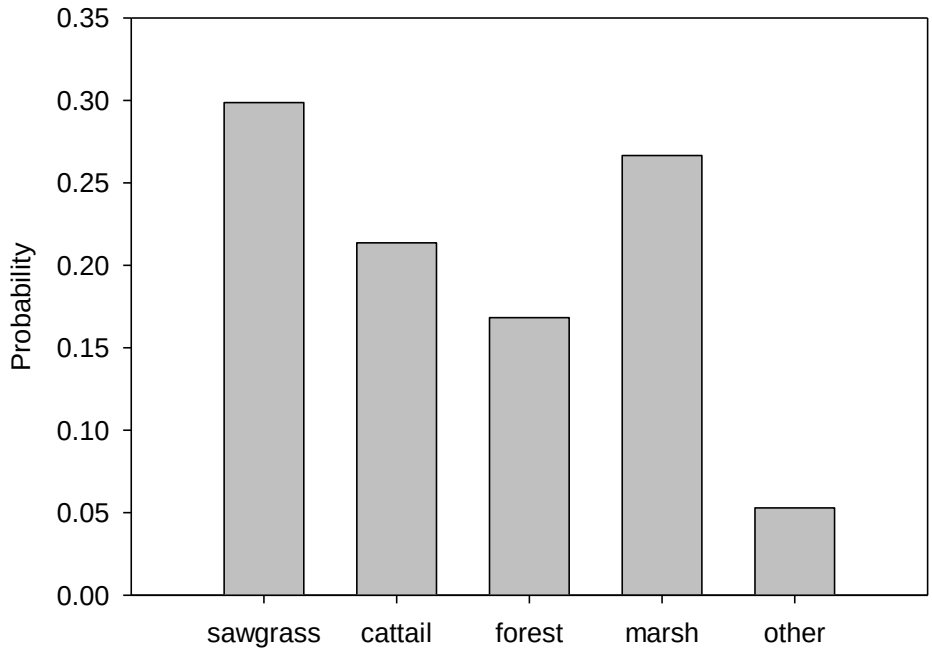


Figure 4-3. Global PDF for land cover types.

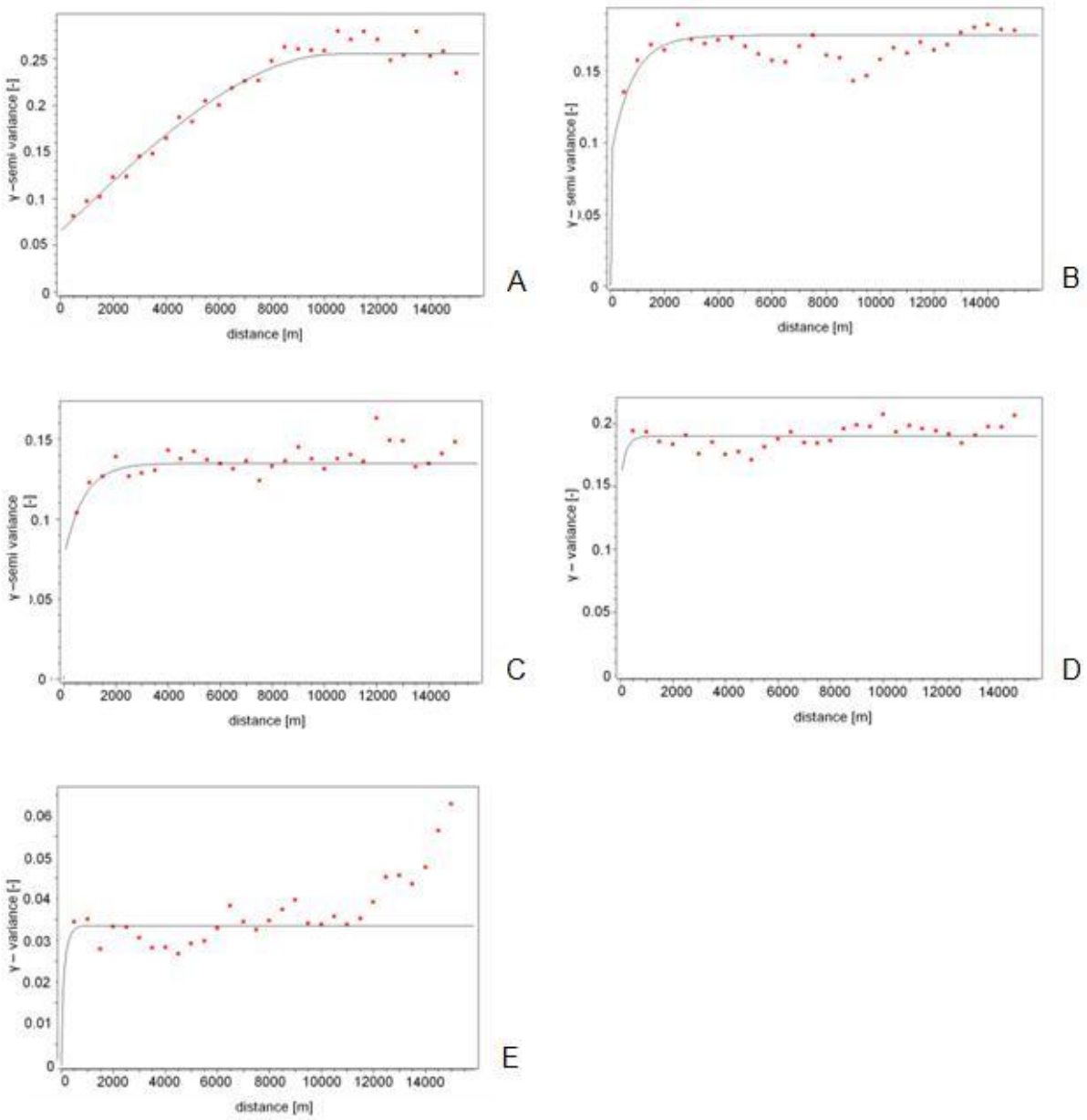


Figure 4-4. Indicator variograms for land elevation datasets. A) sawgrass, B) cattail, C) cypress (trees), D) freshwater marsh, E) other.

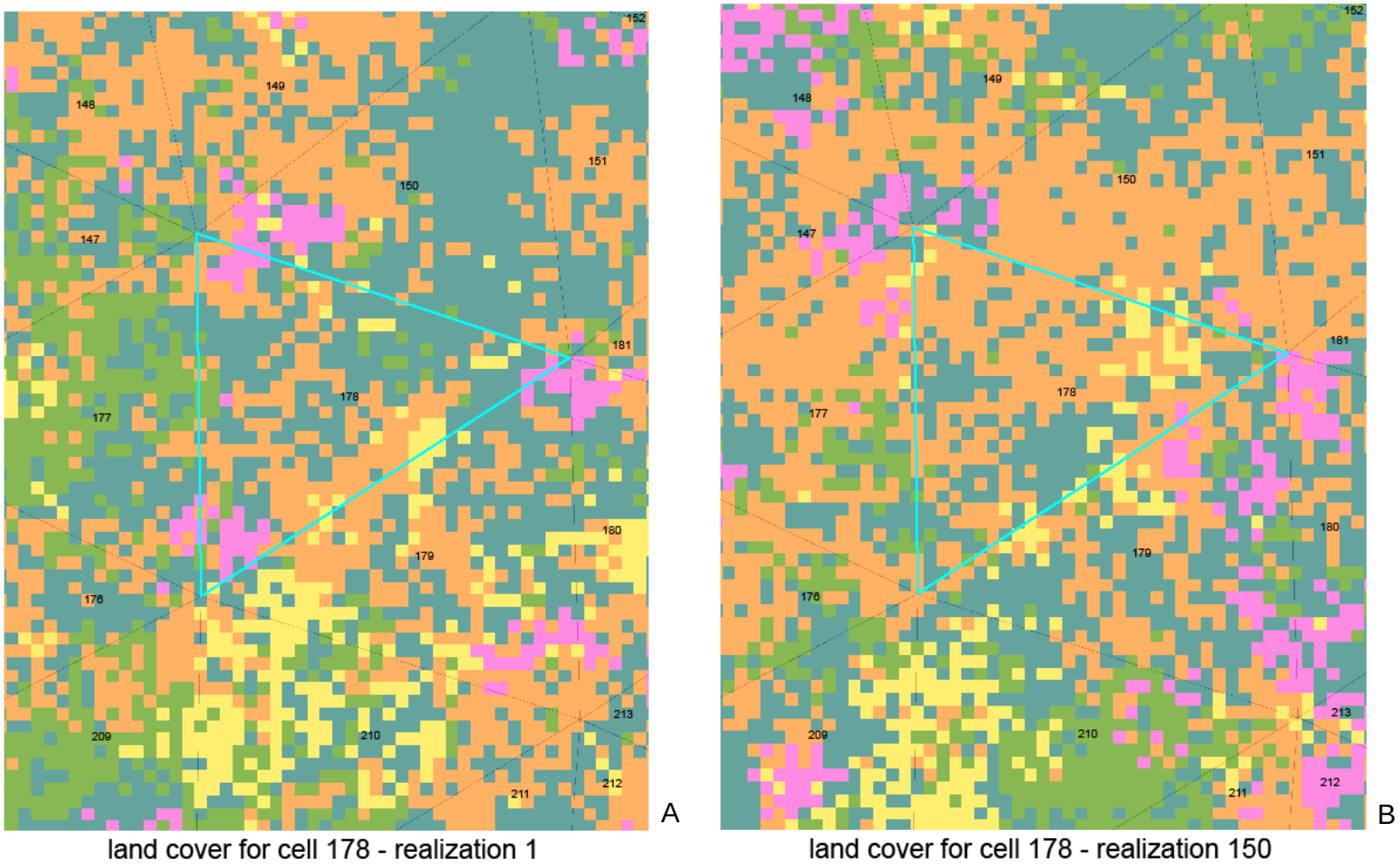


Figure 4-5. Example SIS realizations of land cover for cell 178. A) realization 1, B) realization 150.

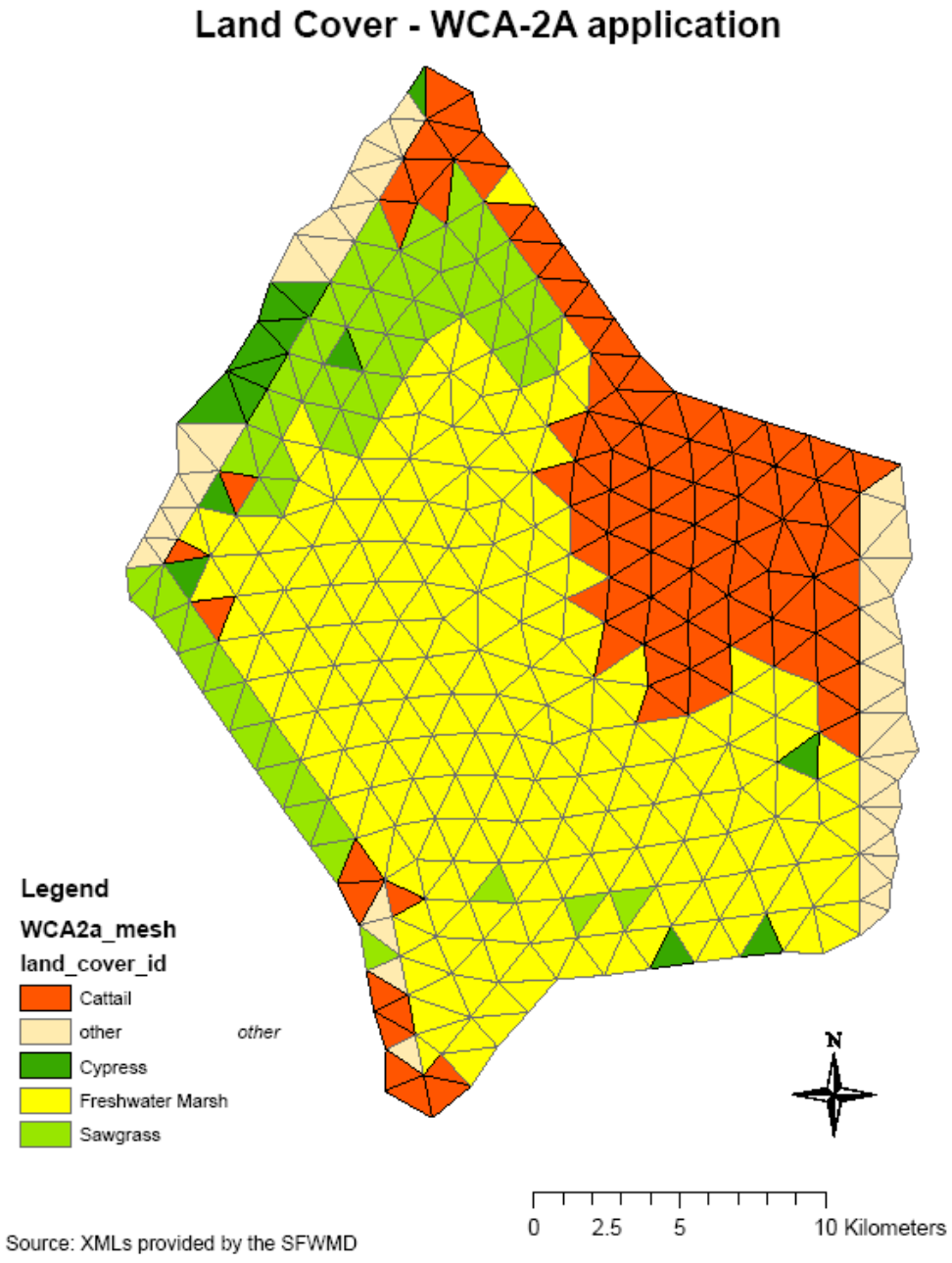


Figure 4-6. Land cover map used originally for WCA-2A application.

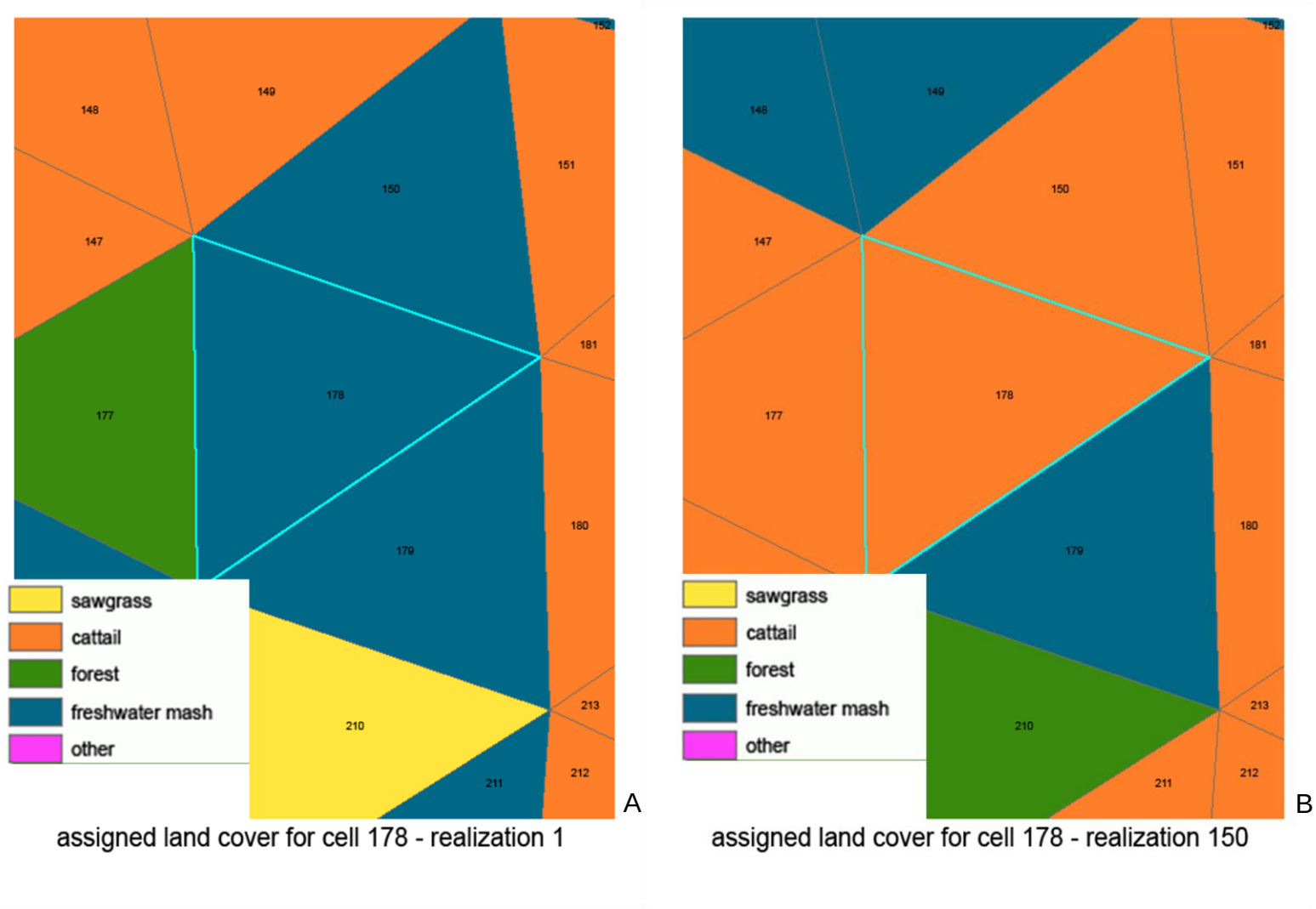


Figure 4-7. Example SIS realizations of land cover for cell 178, aggregated to RSM scale. A) realization 1, B) realization 150.

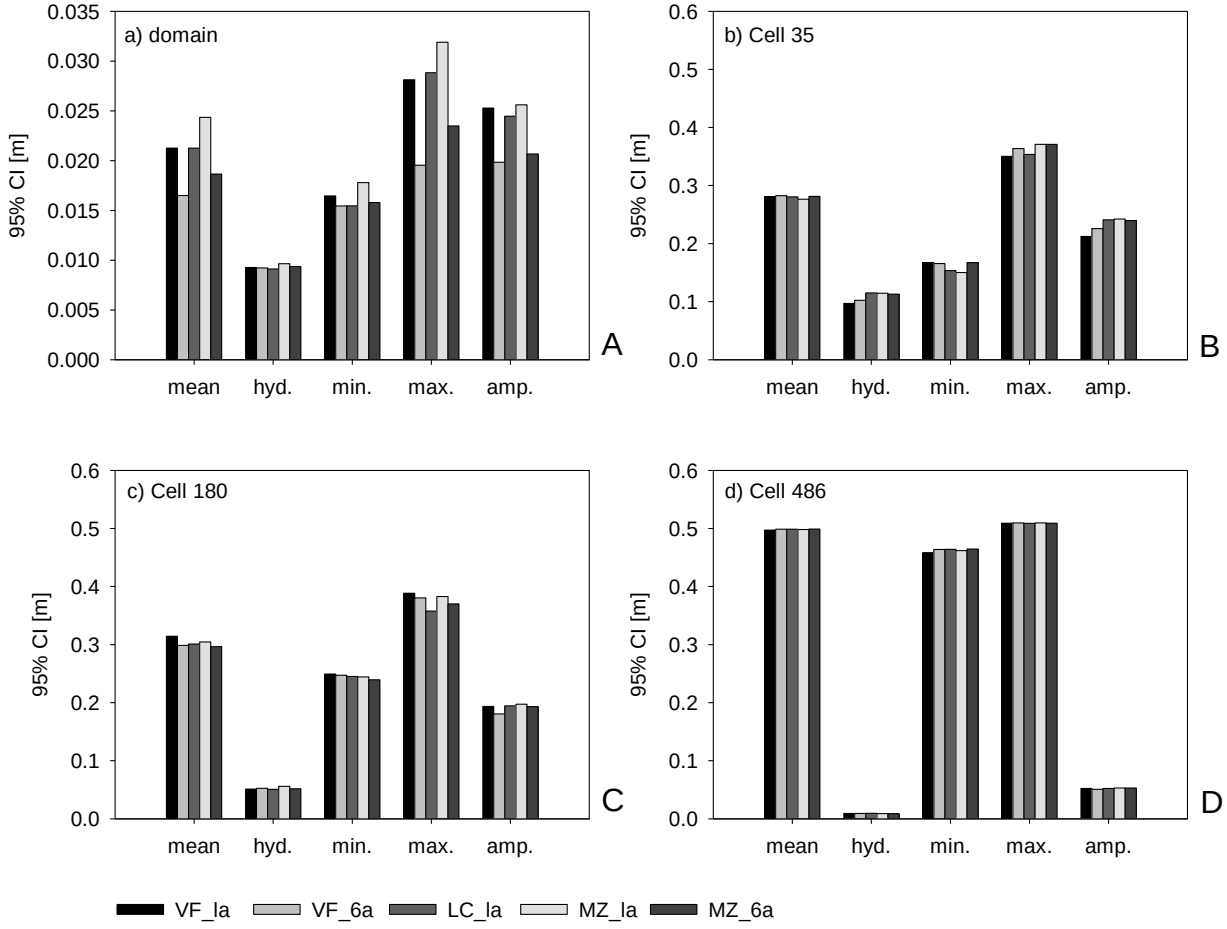


Figure 4-8. GUA results for alternative scenarios from Table 4-3. A) domain-based outputs, B) 35 cell-based outputs, C) 180 cell-based outputs, D) 486 cell-based outputs.

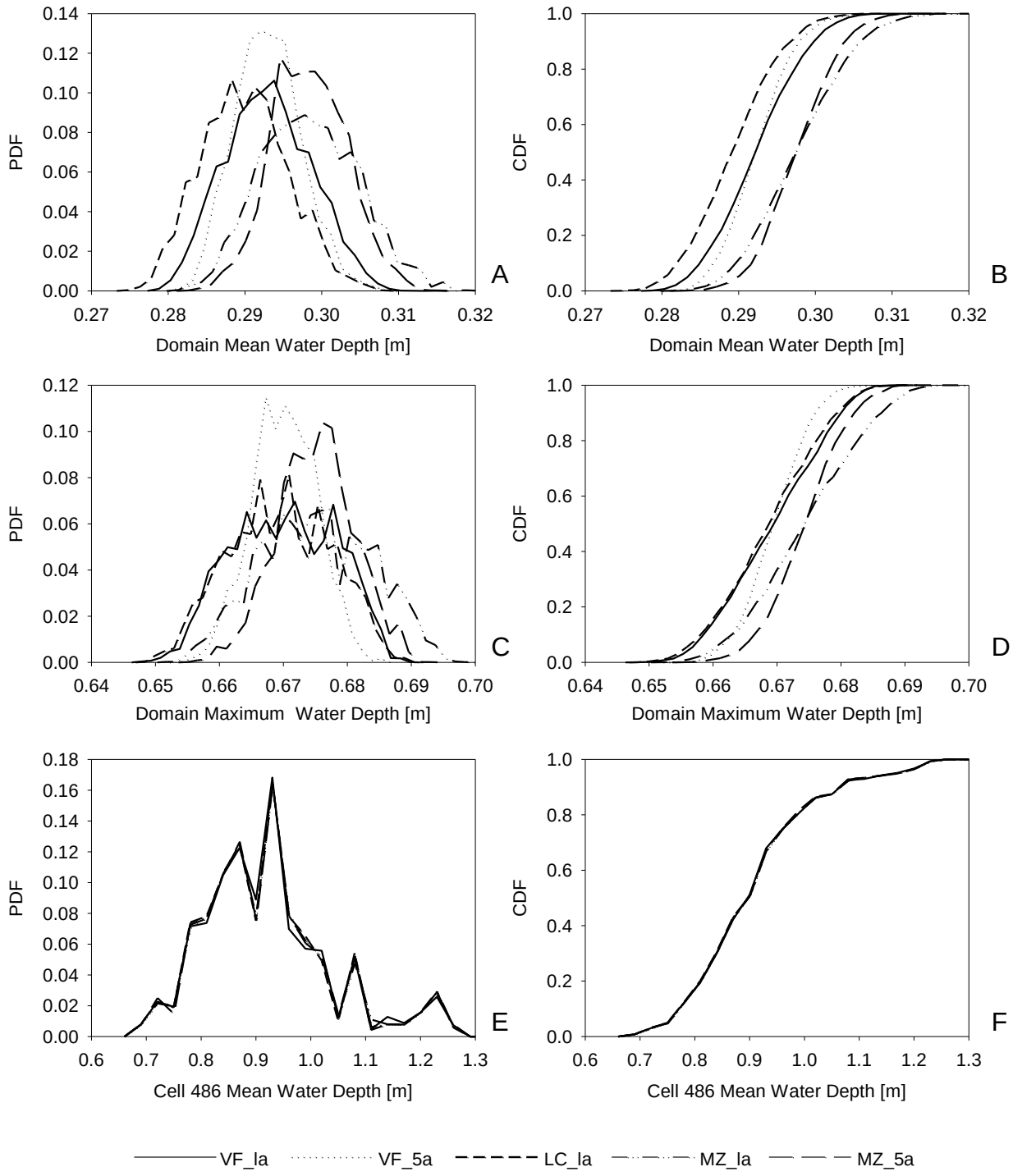


Figure 4-9. GUA results (PDFs – left, CDFs – right) for alternative scenarios from Table 4-3. A), B) domain-based mean water depth, C), D) domain-based maximum water depth, E), F) cell 486-based mean water depth.

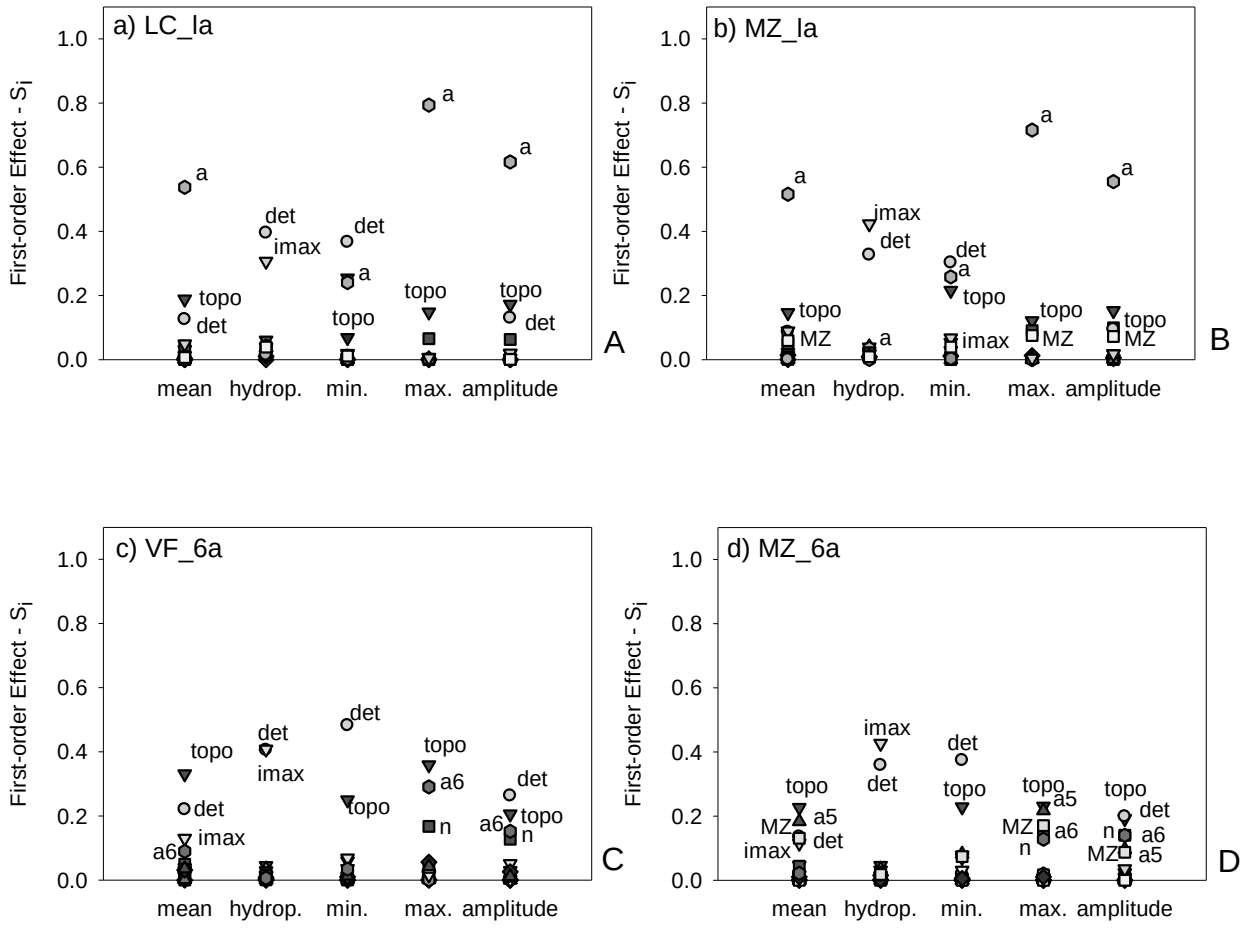


Figure 4-10. GSA results for alternative scenarios. A) LC_la, B) MZ_la, C) VF_5a, D) MZ_5a.

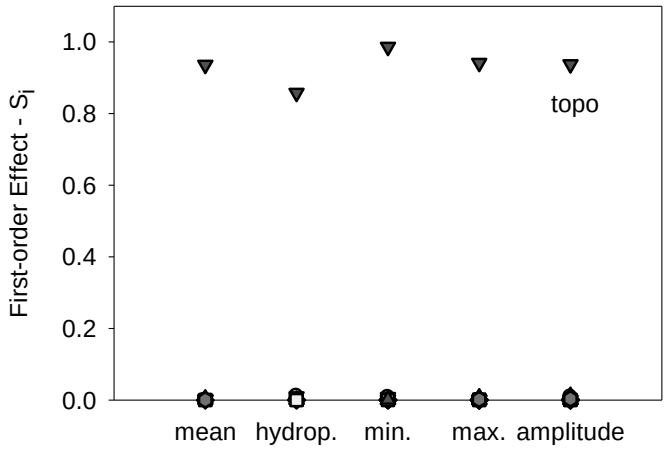


Figure 4-11. Example GSA results for benchmark cell 35, scenario MZ_5a.

CHAPTER 5
UNCERTAINTY AND SENSITIVITY ANALYSIS AS A TOOL FOR OPTIMIZATION OF SPATIAL NUMERICAL DATA COLLECTION, USING LAND ELEVATION EXAMPLE.

Introduction

Despite the fact that the topography is identified as very important input for hydrologic applications very little work has been done to determine the minimum data requirements for this model input. One of the reasons for this is that land elevation uncertainty assessment is complex and challenging, yet it is a mandatory undertaking to the progression of hydrologic science (Wechsler, 2006). The framework used in this study allows for comparing the importance of land elevation maps (or Digital Elevation Models, DEMs) together with other uncertain model inputs. The joint assessment of effects of land elevation uncertainty with other inputs uncertainty has not been addressed so far (Fisher and Tate, 2006) since studies presented in the literature considered either DEM uncertainty on its own or focused on other hydrological model inputs. Simultaneous comparison of land elevation uncertainty and uncertainty from other inputs (spatially lumped or distributed) allows for evaluating the importance of DEM for a particular model application.

The procedure of evaluation of hydrological model uncertainty due to sampling density of land elevation data is a two-step process. At first, land elevation data density translates into spatial uncertainty of land elevation maps used as model inputs. The spatial uncertainty of these maps is assessed by the geostatistical technique of SGS (described in Chapter 3). Secondly, the model of spatial uncertainty, evaluated by SGS, is used for GUA/SA analysis and the corresponding hydrological model uncertainty is evaluated. The approach presented in this Chapter can be used as guidance for spatial data collection for hydrological model applications as it may indicate optimal spatial

density of numerical model inputs in terms of model uncertainty. The analysis presented in this chapter focuses on evaluation of model uncertainty due to alternative land elevation sampling densities.

Spatial Input Data Resolution and Spatial Uncertainty

Spatial density of model inputs is one of the factors affecting spatial uncertainty of input parameters and consequently model predictive quality. Spatial data collection is the most expensive part of distributed modeling (Crosetto and Tarantola, 2001), therefore its optimization can lead to significant improvements in allocation of resources. In case of field data, the optimization of data collection could be obtained by specification of minimum data density (or resolution) that would allow model predictions to meet quality requirements (accuracy and precision).

The effect of data resolution (i.e. soil, meteorological, and land elevation data) on hydrological model output uncertainty was explored in the literature (Inskeep et al., 1996; Wagenet and Hutson, 1996; Wilson et al., 1996; Zhu and Mackay, 2000). These studies show that, in general, model predictions based on input data sets with low spatial resolution were linked with higher model uncertainty. However, it was not always the case. For example, a study presented in Watson et al. (1998) showed that despite more realistic terrain representation of high resolution DEM data, simulation of runoff did not produce better results than using the coarser DEM resolution. This was explained by the fact that the model could not make use of the additional terrain information in the detailed data. This indicates that the input data resolution - model predictive quality relationship is more complex than simple “more data - less uncertainty” concept. As stated by Fisher and Tate (2006): “Whilst there is an increasing tendency to collect

larger volumes of elevation data with seemingly ever-improved precision and accuracy, we have no evidence that this improvement and the associated costs are worthwhile".

Figure 5-1, proposed by Grayson and Blosch (2001), illustrates a conceptual relationship between model complexity, data availability (understood as both the amount and the quality of data) and predictive performance of a model. Grayson and Blosch (2001) stated that: "For a given model complexity, increasing data availability leads to better performance up to a point, after which the data contains no more "information" to improve predictions; i.e. we have reached the best a particular model can do and more data does not help to improve performance". A similar graph (Figure 5-2) presents a conceptual relation between model output uncertainty and data density, used as a hypothetical relationship between model uncertainty and data resolution in this work. The uncertainty decreases with an increase of sampling density but only until a threshold value of data sampling density is reached. Above this threshold value the change of sampling density does not influence the uncertainty. If a threshold value (i.e. optimal data density in Figure 5-2) illustrated in these graphs can be identified for specific model output and spatially distributed model input, this could be considered as an indication of minimum data quality requirements in terms of model output uncertainty. By specifying the optimal data density for a given model and model application, rather than utilizing "one size fits all" approach (i.e. using the same input data densities for various models and applications), the resources spent on data collection may be allocated efficiently.

The Influence of Land Elevation Uncertainty on Hydrological Model Uncertainty

Topography is an important factor for hydrological models (Wilson and Atkinson, 2005, Wechsler 2006). Land elevation affects surface flow routing as it is used to derive

terrain characteristics (like slope and aspect, i.e. direction in which a slope faces) for hydrological applications. Land elevation is usually represented in a form of digital elevation models (DEMs). A DEM is a numerical representation of surface elevation over a region of terrain (Cho and Lee, 2001). DEM is just a model (abstraction) of reality that inherently contains deviations from the true values, or errors. As the “true land elevation” is not known, the error cannot be calculated and uncertainty arises. Despite the DEM uncertainty and its potential importance for hydrologic applications, DEM data are often used for hydrological simulations without quantification of DEM uncertainty and its propagation. Uncertainty regarding land elevation should inform the uncertainty of topographic parameters (like slope) and further propagate into uncertainty of hydrological outputs. The DEM error/uncertainty is especially important in areas of relatively flat terrain, since small variations in such areas significantly affect hydrological flow paths (Burrough and McDonnel 1998). In such conditions, even a small degree of uncertainty in elevation may have a relatively large effect of model predictions.

Uncertainties associated with land elevation for hydrologic applications have been studied with different approaches (Fisher and Tate, 2006; Wechsler, 2006). DEM accuracy is usually reported as a global statistic Root Mean Square Error (RMSE), obtained based on comparison with more accurate land elevation data. However, this is just one value for the map and it has been suggested that the assessment of DEM uncertainty requires more information on spatial structure of the error not possible by RMSE (Wechsler, 2006). Kyriakidis el al. (1999) suggests using maps of local probabilities for over or underestimation of the unknown reference elevation values from those reported in the DEM, and joint probability values attached to different spatial

features. There is still little known about spatial structure of DEM error (Liu and Jezek, 1999), and it is currently often difficult, if not impossible, to recreate the spatial structure of error for a particular DEM, as higher accuracy data – usually non available - is required. In fact, the uncertainty of DEM is related to the following factors: a) source data (accuracy, density and distribution); b) characteristics of the terrain surface; c) method used for construction of the DEM surface (interpolation and processing) (Gong et al, 2000).

Two approaches towards simulating DEM uncertainty for uncertainty assessment and error propagation are usually applied (Wechsler, 2006): 1) derivation of error analytically, and 2) stochastic simulation of error (unconditional, conditional).The example of the first approach was presented by Hunter and Goodchild (1995). For every pixel (single point in DEM grid), error was assumed to follow the normal distribution around the estimated elevation value and the global RMSE was assumed as a local error variance around this estimate. DEM errors are not spatially correlated and spatial structure of error is not considered; DEM error is normally distributed with mean zero and standard deviation approximated by the RMSE. For the second approach for simulating error, the spatial structure of error is considered; the information on spatial structure of the error is obtained by comparison with more detailed DEM (Endreny and Wood, 2001) or ground measurements (Canters et al.2002), or both (Enderny et al., 2000).

Propagation of DEM Uncertainty due to DEM Resolution

Among all the factors affecting DEM uncertainty, this study focuses on the density of source measured data. The spatial resolution of DEM affects the accuracy of the terrain. For the case of raster or regular grid DEMs, a sampling interval is constant and

it is referred as resolution. Similarly, for field measurements distributed on a grid the sampling density is equivalent to DEM resolution. Irrespective of the source of the data used for DEM construction (field surveys, topographic maps, stereo aerial photographs or satellite images), the error in a DEM can be influenced by the density and distribution of the measured point source data. Gong et al. (2000) found that the sampling interval is the most important factor affecting accuracy of DEM for a given type of terrain and that the relationship between DEM accuracy and sampling interval was linear and negative, more pronounced, for hilly areas than for flat ones (Gong et al., 2000). The influence of DEM resolution on the DEM accuracy was also examined by Li (1992) that concluded that smaller sampling interval was more accurate, especially for complex terrains. Similarly, Östman (1987) observed that an increased point density reduced the RMSE, while Gao (1997) showed that RMSE increased with a decrease of resolution from 10 to 60m (and this relation was linear) when producing DEM from contour maps because larger sample size captured the terrain better (Gao, 1997). In summary, smaller grid cell size allows for better representation of complex topography and high resolution DEMs are better able to depict characteristics of complex topography.

DEM resolution was also reported to affect terrain attributes (Carter, 1992; Chang and Tsai, 1991; Kenzle, 2004). Chang and Tsai (1991) reported that slope and aspect were less accurate if generated from DEM of lower resolution.

As a result of affecting DEM uncertainty and terrain characteristics uncertainties, DEM resolution was shown to directly impact hydrologic model predictions for spatially distributed models like TOPMODEL (Band and Moore, 1995; Quinn et al., 1995; Wolock

and Price, 1994; Zhang and Montgomery, 1994), the SWAT model (Chaubey et al., 2005; Chaplot, 2005), and AGNPS (Perlitsh, 1994; Vieux and Needham, 1993).

Based on the hypothesis presented in Figure 5-2, despite the generally reported trends between increased DEM resolution and derived terrain characteristics accuracy, increase of land elevation source data resolution does not always produce better hydrological models predictions. For land elevation maps used as model inputs, constant increasing data resolution will inevitably lead to some redundancy. For example, Zhang and Montgomery (1994) concluded that a 10 m grid size provides a substantial improvement over 30 and 90 m data, but 2 or 4 m data provide only marginal additional improvement for the performance of physically based models of runoff generation and surface processes.

What resolution of land elevation should be used to construct a DEM used as inputs for model simulations? Two aspects of modeling need to be considered for answering this question, these are the financial cost of obtaining land elevation data and, accuracy requirements that need to be met by model predictions. The identification of the optimal data density for modeling requires answering two questions: 1) to what extent is the source data resolution a factor in the propagation of errors from DEMs to model outputs, and 2) how does this uncertainty relate to other model input uncertainties associated with a given model and its application, i.e. is land elevation uncertainty important when compared with uncertainties of other model inputs? In order to answer these questions the GUA/SA needs to be performed using land elevation maps obtained from alternative data resolutions (sampling densities). The methodology, proposed in the previous chapter, based on the combination of the SGS and method of

Sobol, allows for evaluation of spatial uncertainties related to different land elevation data densities. Moreover, the uncertainty of DEM is evaluated simultaneously with the uncertainties of other model inputs and relative uncertainty of land elevation can be evaluated.

The objectives of the study presented in this chapter are to: a) evaluate the effect of spatial sampling resolution of a distributed model input data (specifically source land elevation data) on output uncertainty and parameter sensitivities of a complex hydrological model (RSM); b) estimate the optimal spatial resolution of source land elevation data in terms of tradeoffs between costs associated with higher spatial resolution of data collection and reduction of uncertainty of model outputs.

Methodology

Subsets from the original WCA-2A, AHF land elevation survey are extracted and used as alternative data sources for construction of DEMs. The methodology presented in the study is based on two steps: geostatistical technique of sequential Gaussian simulation (SGS) for assessment of land elevation spatial uncertainty, and on the method of Sobol, global uncertainty and sensitivity analysis, for propagation of the input uncertainty onto the model outputs. As described in Chapter 3, the synergistic combination of these two methodologies results in a global spatial uncertainty and sensitivity analysis that has the ability to account for spatial autocorrelation of input variables and is independent of model behavior. Detailed description of the procedure, together with its assumptions, is provided in (Chapter 3).

Description of Land Elevation Data Subsets

As described in Chapter 3, a total of 1,645 land elevation data points are available for WCA-2A (USGS, 2003) (see Table 3-1). Data is regularly spaced, on a 400 x 400 m

grid. Land elevation measurements were obtained using the Airborne Height Finder (AHF), a helicopter-based instrument developed specifically for South Florida conditions (vast extent, very flat topography, impenetrable vegetation). The vertical accuracy of data is at least +/- 15 cm (USGS, 2003).

To investigate the effect of sample data density, the original land elevation data set (400x400 m spacing) is reduced to subsets of 1/2, 1/4, 1/8, 1/16, 1/32 and 1/64 of original data. All 7 data sets are approximately regularly distributed (example data sets are presented in Figure 5-3). The descriptive statistics and histograms for each data set are presented in Table 5-1 and Figure 5-4. These datasets consisting of different densities of measured point data are used individually to produce alternative land elevation maps for RSM simulations.

Estimation of Spatial Uncertainty of Land Elevation

The method of Sequential Gaussian Simulation (SGS) is used for estimation of spatial uncertainty for land elevation maps, produced based on the 7 datasets. For each dataset of land elevation values, SGS reproduces the measured data, data histogram and variogram. “The remaining “space” of spatial uncertainty beyond these data constrains is explored via a random number generator (Kyriakidis, 2001). For each of the datasets, L=200 equiprobable maps of land elevation are generated by SGS. Alternative land elevation realizations, taken together, constitute spatial uncertainty of land elevation. The procedural steps presented in Figure 3-5 and described in Chapter 3 are followed for each land elevation dataset individually:

1) land elevation data are de-trended using a trend fitted for the original data;
2) normal score transform is performed for the measured values;
3) SGS is performed for the nscore space;

4) simulated grid values are back-transformed into residuals space;

5) the trend is added to simulated residuals.

The nscores of residuals are interpolated into elevation matrices with a Simple Kriging (SK) algorithm. The same interpolation grid is used for all data densities, that is 200x200m grid. After SGS, each of the alternative realizations (maps) is aggregated to the RSM mesh scale by overlaying the model mesh over the 200x200m grid. Values for SGS nodes corresponding with centroids of the RSM triangular cells are extracted and used as effective land elevation values for model cells. Since the centroids' values are conditioned on the measured data and SGS simulated values within the search radii the continuity between land elevation values for neighboring RSM cells is maintained. Cell-by-cell comparison of 200 aggregated maps of land elevation provides a PDF of land elevation values for each model cell, from which estimation variance, confidence intervals, and other desired statistics can be derived. The estimation variance is calculated for each of model cells, based on the PDF, constructed from 200 aggregated values. Then, for each of the datasets, the average estimation variance is calculated as a global measure representing map variability.

Two alternative approaches are considered for the SGS in this study: 1) SGS is performed using the same "true" histogram and variogram model for all datasets; 2) SGS is performed using experimental variograms and histograms, constructed for each dataset separately, based on the data in the given dataset.

For the first approach, it is assumed that the 'true' global distribution (histogram) of data in a domain is known and that it is approximated by the histogram of the original data (density 1), and that the 'true' model of spatial variability is approximated by the

variogram for the same densest dataset (density 1). In this case, the only factor changing between different datasets is the density of measured data, while the histogram and the variogram are the same. This assumption allows filtering out effects related to various sample sizes and histograms of the considered datasets. The variogram model for the original land elevation data, used for the SGS of all datasets, is presented in Figure 3-7. It has a nugget of 0.59 (dimensionless) and two structures: exponential with sill contribution of 0.25 and range of 5.3 km; and Gaussian with sill contribution of 0.16 and range of 12 km.

For the second approach it is assumed the only information available for generation of plausible land elevation realizations is the actual dataset, so different measured data sets, histograms, and variograms are used for each data density. The histograms for datasets with different densities are presented in Figure 5-4. The variogram models, fitted to experimental variograms for each dataset are presented in Figure 5-5, and parameters for these exponential variogram models are summarized in Table 5-2. It can be seen that these variograms are very similar. Unlike, variogram for the density of 1, these are one-structure variograms.

This first approach allows for examination of effect of various data densities on the spatial uncertainty of land elevation realizations, and consequently, its propagation to hydrological model outputs. Therefore, this first approach is going to be presented in this Chapter. The SGS results for the second approach are presented in Appendix E.

Global Uncertainty and Sensitivity Analysis

In this study the GUA/SA analysis is performed for each of the7datasets separately. As presented in Chapter 3, the 200 maps, embodying the spatial uncertainty

are used in the GUA/SA using the method of Sobol through the auxiliary input factor associated with alternative land elevation realizations.

The RSM outputs chosen as metrics for GUA/SA for this study are: mean water depth, hydroperiod, and maximum water depth for domain and 3 benchmark cells: 35, 215, and 486 (Figure 2-1). These cell-based performance measures reflect the hydrological variability across the domain. Raw model results are post processed using the approach described in the previous chapters. Model simulations are performed for period of 1983-2000 with first year used for model warm-up.

Results

Sequential Gaussian Simulation Results

Maps presenting estimation variances for selected data densities are presented in Figure 5-6. The general increase of spatial uncertainty is visually observed (by visual analysis) in the maps produced from smaller data densities. Furthermore, it can be observed that for a given map, there is no spatial pattern in estimation variances within the domain. As specified in the SGS theory section in Chapter 3, for sufficiently large number of realizations, at a given SGS grid node, the estimation variance should be similar to the SK interpolation variance. The SK variance is a function of distance from measured data and data distribution. Since for each dataset, measured data are regularly distributed in the domain, the variances of kriged nscore values and back-transformed values should not exhibit spatial patterns.

As seen in Figure 5-7, the average estimation variance decreases with the increase of source data density. The decrease accelerates at the inflection points 1/8 of original data density. The average estimation variance decreases rapidly from

0.0121 m^2 for density 1/64 to 0.0106 m^2 for density 1/8, and then decreases slowly to 0.0097 m^2 for density 1.

Global Uncertainty and Sensitivity Analysis Results

The relationship between output uncertainty (expressed as the 95% Confidence Interval) and land elevation data density for the domain outputs is illustrated in Figure 5-8. The trends for mean and maximum water depth (Figure 5-8 A and C) are similar to the trend observed for the average estimation variance. There is not much change in output uncertainties for greater than 1/4, while uncertainty increases sharply with reduction of data density below 1/4 to 1/8 of initial data density. In contrast, the uncertainty for hydroperiod does not seem to be affected by change of land elevation data density (Figure 5-8 B).

The relationship between benchmark cells outputs and land elevation data density is presented in Figure 5-9. In case of benchmark cell-based outputs, no general pattern between uncertainty and data density is observed. Mean and maximum water depth for cell 215 show pattern similar to patterns observed for the corresponding domain-cased outputs. On the other hand, the outputs for benchmark cells 35 and 486 do not seem to display any relation between uncertainty and land elevation data density.

The sensitivity analysis (SA) results for domain-based outputs exhibit similar trends as the uncertainty results (Figure 5-10). The SA results indicate that the importance of factor *topo* (S_{topo}) increases with a reduction of land elevation data density for mean and maximum water depth (Figure 5-10 A and C), while it is unchanged for hydroperiod (Figure 5-10 B). There seems to be not much difference in S_{topo} for densities between 1 and 1/4, and the contribution of this factor increases significantly below the density of 1/8. For example for mean water depth variance, the

first-order sensitivity index S_{topo} contributes to about 20% for the density of 1, below the density of 1/8 its influence increases and eventually reaches over 40% for the density of 1/64. A similar trend is exhibited by the first order sensitivity index for *topo* in case of domain's maximum water depth. The factor *topo* does not seem to influence uncertainty of domain-based hydroperiod in large extent. It contributes to the variability of this output from 5% (density 1) to 10% (density 1/64). As seen in Figure 5-10, the decreased contribution of factor *topo* to the output variance is accompanied by the increase of importance of a spatially certain factor *a.* This factor, together with factor *det*, also plotted in the figure, is one of the most important factors contributing to the output variances for the original land elevation density (as presented in Chapter 3). The sum of first order sensitivity indices is close to one for domain-based outputs when the original land elevation density is used for the analysis (Figure 5-10, A and C). Therefore increase of topo contribution, observed for smaller data densities, needs to be accompanied by decrease of importance of other factors. No interactions between factors are observed (the total order effects are similar to the first order effects) but it seems that factors *topo* and *det* are somehow interconnected as they switch the importance in affecting model output, while other important factor, parameter *a*, remains unaffected.

GSA first order sensitivity indices results for the benchmark cell-based outputs indicate that the responses of the benchmark cells are completely dominated by the land elevation spatial variability. Figure 5-10 illustrates the example of S_i results for cell 35.

Discussion

The results of this study show that the domain-based outputs follow the hypothetical trend for the model uncertainty and spatial density of model input data presented in Figure 5-2. This nonlinear, negative trend, with inflection point, is observed for domain-based mean water depth and maximum water depth. These two outputs are affected by land elevation uncertainty as indicated by the GSA results (i.e. have high values of S_{topo}). The domain-based hydroperiod, that is not affected by factor *topo* to a great extent does not display any trend. The trend observed for model outputs seems to be reflection of the pattern for spatial land elevation uncertainty and data density what is related to the fact that the variability of land elevation maps is transferred into uncertainties of model predictions.

Both relations (spatial uncertainty and model uncertainty vs. data density) are characterized by the inflection point around data density of 1/4 to 1/8 (Figure 5-7, Figure 5-8). These densities correspond to average measured data spacing of 800 m and 1131 m respectively (Table 5-1), that is in the range of model cell size (on average 1.1 km^2). The general increase of spatial uncertainty can be explained by the fact that with smaller resolution of the data, there is a larger uncertainty due to spatial structure of the land elevation maps (larger interpolation variance). The kriging estimation variance depends on the number and proximity of supporting data points and degree of spatial dependence as quantified by a semivariogram (Robertson, 1987). It is directly proportional to the distance of an interpolated value from an input observation. Therefore the less dense datasets are associated with higher interpolation variance. Since SGS realizations are aggregated to the RSM scale, the estimation variance for cell values is also affected by the aggregation method (in this case the centroids

approach). Other aggregation method, for example spatial averaging of SGS values within model cell, would probably result in different estimation variance.

The question that comes into mind is which factors determine the value of inflection density for the spatial uncertainty vs. density relationship. In this study the inflection density coincides with the average cell size. Since spatial uncertainty is estimated as the average of variances for selected SGS grids (i.e. grids that contain mesh centroids), it seems that the observed pattern is related to interpolation method rather than the aggregation method (i.e. spacing of cells centroids related to cell size). Besides, aggregation method is constant for all data densities, so it should not affect the relative results for the datasets.

The lack of clear pattern presented in Figure 5-2 is observed for the benchmark cell-based outputs and land elevation density. This may be related to the mismatch of scales between cell-based outputs and model inputs changing on the domain-scale. In case of the WCA-2A application, the general direction of flow (from north to south) is maintained irrespectively of land elevation data density. Therefore the uncertainty of this cell is not affected by land elevation density used for generation of land elevation maps, as no matter what topography-conditioned path will be selected for model simulations, the water will eventually end up in this cell. Cell 35 located in the north of domain, does not exhibit clear trend, because of the similar reasons. This cell is located at the generally higher and drier part of the domain. Therefore irrespective of the data density used for generating topography maps, this cell will always be higher and drier than cells located southwards in a domain. However, the uncertainty of mean and maximum water depth for this cell increases for the smallest two densities 1/32 and 1/64 of original data

density, suggesting that these densities are associated with spatial uncertainty that affects northern cells outputs. The SA results of benchmark cell outputs are dominated by factor *topo.* As reported in the previous chapter, this factor associated with land elevation spatial uncertainty dominates cell-based outputs even for the original data density (i.e. density associated with the smallest spatial uncertainty); therefore further increase of land elevation with decrease of land elevation density importance is not possible.

This study provides findings that are specific to the examined model and its application. By examining the uncertainty and sensitivity results obtained for different land elevation datasets, it is possible to isolate model uncertainty solely due to land elevation data resolution. Furthermore, it is possible to determine land elevation data density threshold, below which the model uncertainty increases significantly. For the current RSM application to the WCA-2A, one could accept the domain-based outputs uncertainty increase from density 1 to density 1/4, as a tradeoff for smaller spatial data requirements. Such information could be helpful in designing data collection efforts for areas similar to WCA-2A (possibly other wetland areas in extensive South Florida region). It is important to remember that the currents results are obtained using several assumptions. Spatial uncertainty models for the alterative datasets are constructed based on the assumption that the “true” global probability distribution (histogram) and model of spatial variation (variogram) are known. In this way the influence of other effects (like variability of sampled data in a given dataset) is eliminated from the experiment.

The more general (model and application independent) findings of this study are related to the corroboration of patterns illustrated in Figure 5-1and Figure 5-2. This study illustrated that the relationship between model uncertainty and input data quality can be defined, and that the inflection point can be identified. Possibly similar patterns can be identified for other hydrological models and applications in order to further explore general factors affecting model outputs uncertainty.

As noted by Crosetto and Tarantola (2001), such an approach would be especially useful at the beginning of a large-scale modeling project, when it needs to be decided how to allocate of resources for data collection, and what should be the minimum data requirements for model inputs. The analysis based on the SGS and method of Sobol could be applied for the small area, representative of the modeling domain, before larger data collection efforts are undertaken.

Conclusions

Spatial data collection efforts can be optimized by specification of minimum data requirements for a given model application. In this chapter, a hypothetical negative, nonlinear relationship between model uncertainty and source data density is developed and tested. The GUA/SA with incorporation of spatial uncertainty is applied for identification of minimum spatial data requirements (data density) for land elevation. Source data density is found to affect spatial uncertainty of topography maps, used as alternative model inputs, and consequently the hydrological model outputs. Comparative GUA/SA results for the 7 land elevation densities show that domain-based outputs (mean water depth and maximum water depth) are impacted by the density of land elevation data. The results corroborate the hypothetical relationship between model uncertainty and source data density. The inflection point in the curve is identified

for the data density between 1/4 and 1/8 of original data density. It is postulated that the inflection point is related to the characteristics of the spatial dataset (variogram) and the aggregation technique (model grid size). Sensitivity analysis results indicate that contribution of land elevation to the domain-based outputs variability (mean water depth and maximum water depth) shows similar pattern as the uncertainty results. In case of benchmark cell-based outputs, generally no clear trend is observed between output uncertainty and data density. Based on the comparative results for the considered land elevation densities, it is concluded that the reduced data density (up to 1/8 of original land elevation data points) could be used for simulating the WCA-2A application with RSM, without significantly compromising the certainty of model predictions and the subsequent decision making process. The results of this chapter illustrate how quantification of model uncertainty related to alternative spatial data resolutions allows for more informed decisions regarding planning of data collection campaigns.

Table 5-1. Summary of descriptive statistics for land elevation datasets.

Sample statistics	Sampled data density						
	1	1/2	1/4	1/8	1/16	1/32	1/64
Sample Size	2643	1320	663	332	162	81	40
Interval [m]	400	565	800	1131	1600	2262	3200
Range [m]	3.51	2.54	2.54	2.23	1.54	1.31	1.22
Mean [m]	3.04	3.04	3.05	3.05	3.04	3.05	3.05
Variance [m^2]	0.10	0.09	0.09	0.10	0.09	0.09	0.10
Minimum [m]	0.77	1.74	1.74	2.05	2.07	2.25	2.34
Maximum [m]	4.28	4.28	4.28	4.28	3.61	3.56	3.56

Table 5-2. Summary of nscore variogram parameters for data subsets.

variogram parameter	variogram type	Sampled data density					
		1/2	1/4	1/8	1/16	1/32	1/64
nugget effect	Exp.	0.58	0.64	0.62	0.60	0.62	0.62
sill contribution	Exp.	0.42	0.37	0.34	0.40	0.38	0.38
range [m]	Exp.	10000	11180	8100	10400	9450	9450

Exp. – exponential model

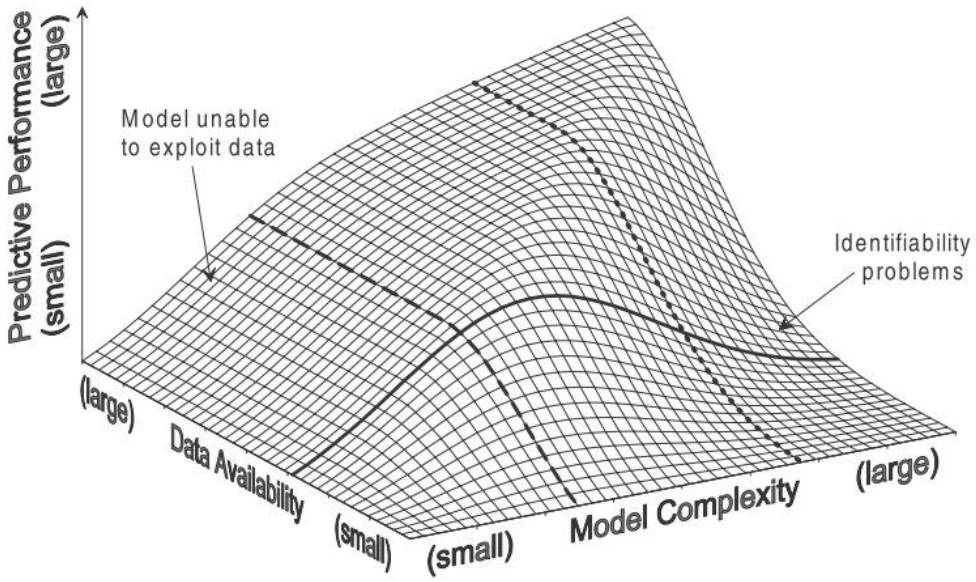


Figure 5-1. Schematic diagram of the relationship between model complexity, data availability and predictive performance (after Grayson and Bloschl, 2001).

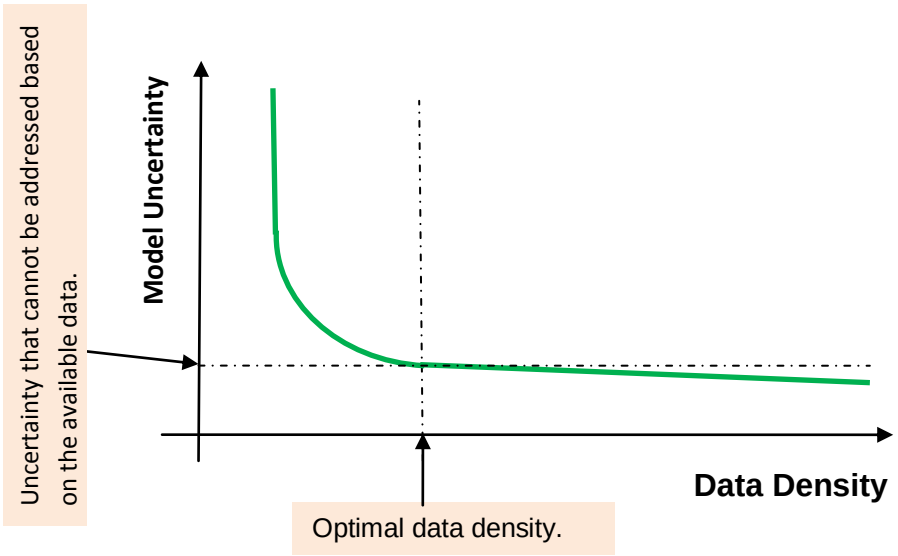


Figure 5-2. Hypothetical relation between data density and variance of the model output.

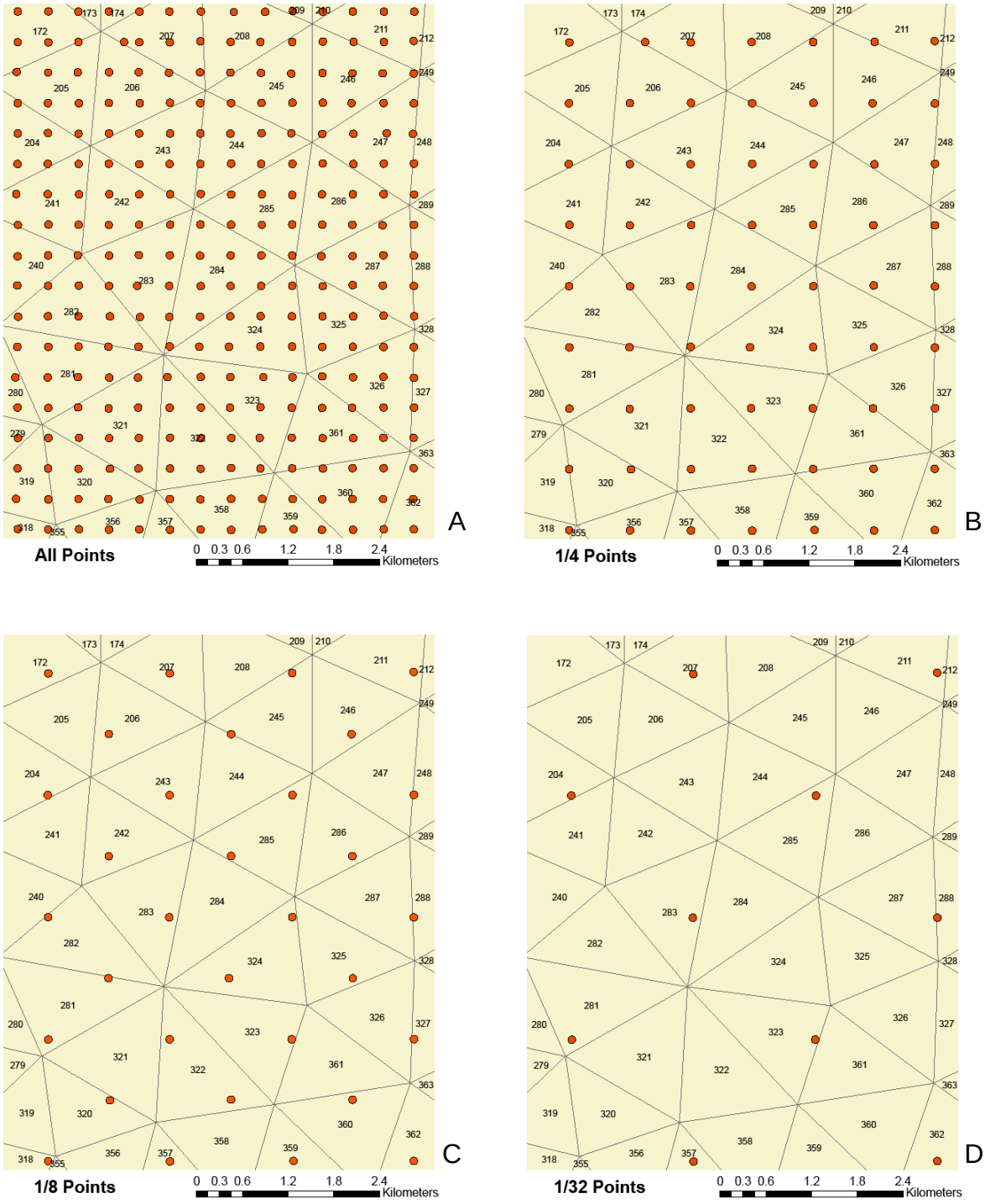


Figure 5-3. Selected datasets used for the analysis. A) original data points, density of 1, B) density of 1/4, C), density of 1/8, D) density of 1/32.

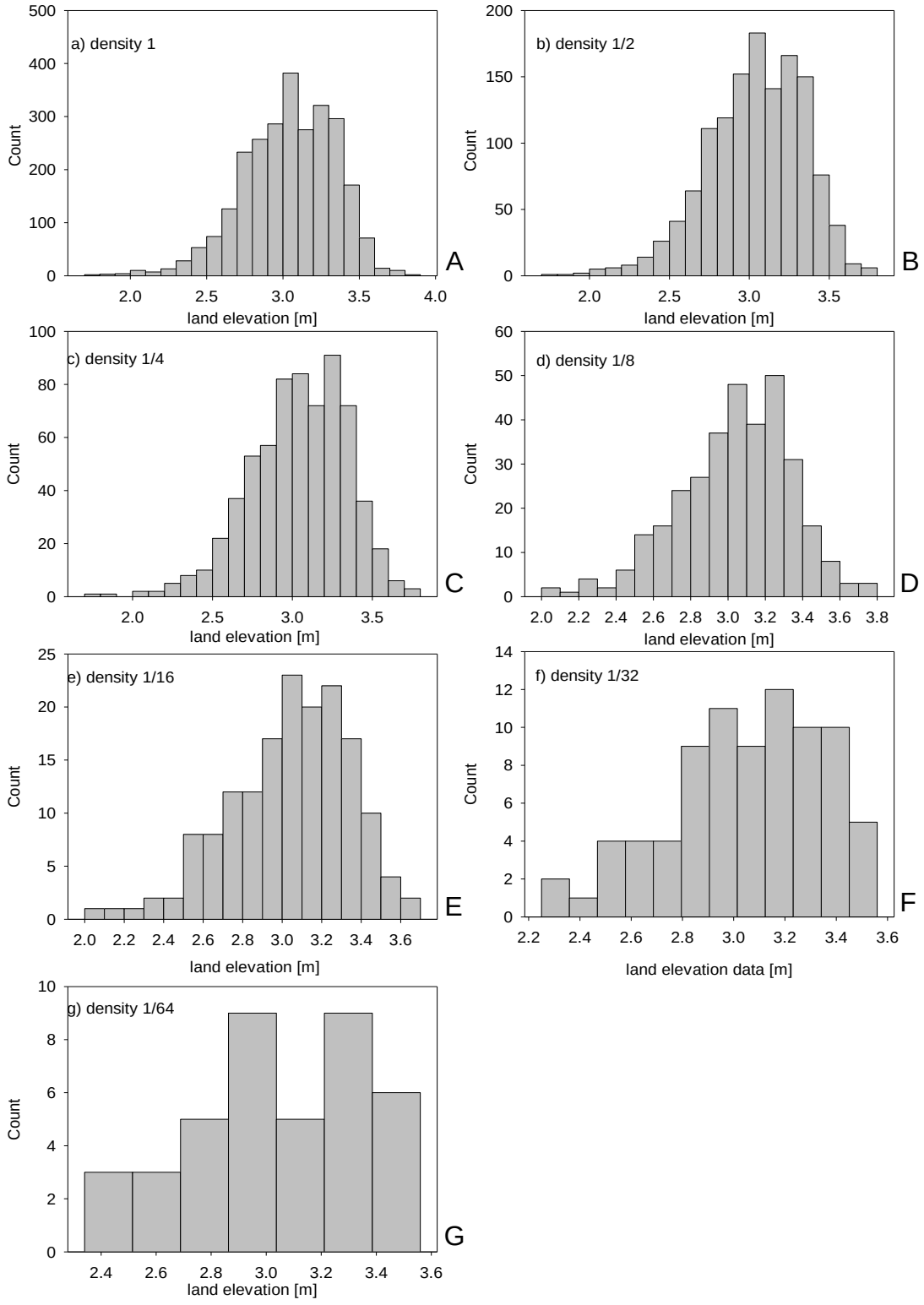


Figure 5-4. Histograms for land elevation datasets. A) density 1, B) density 1/2, C) density 1/4, D) density 1/8, E) density 1/16, F) density 1/32, G) density 1/64.

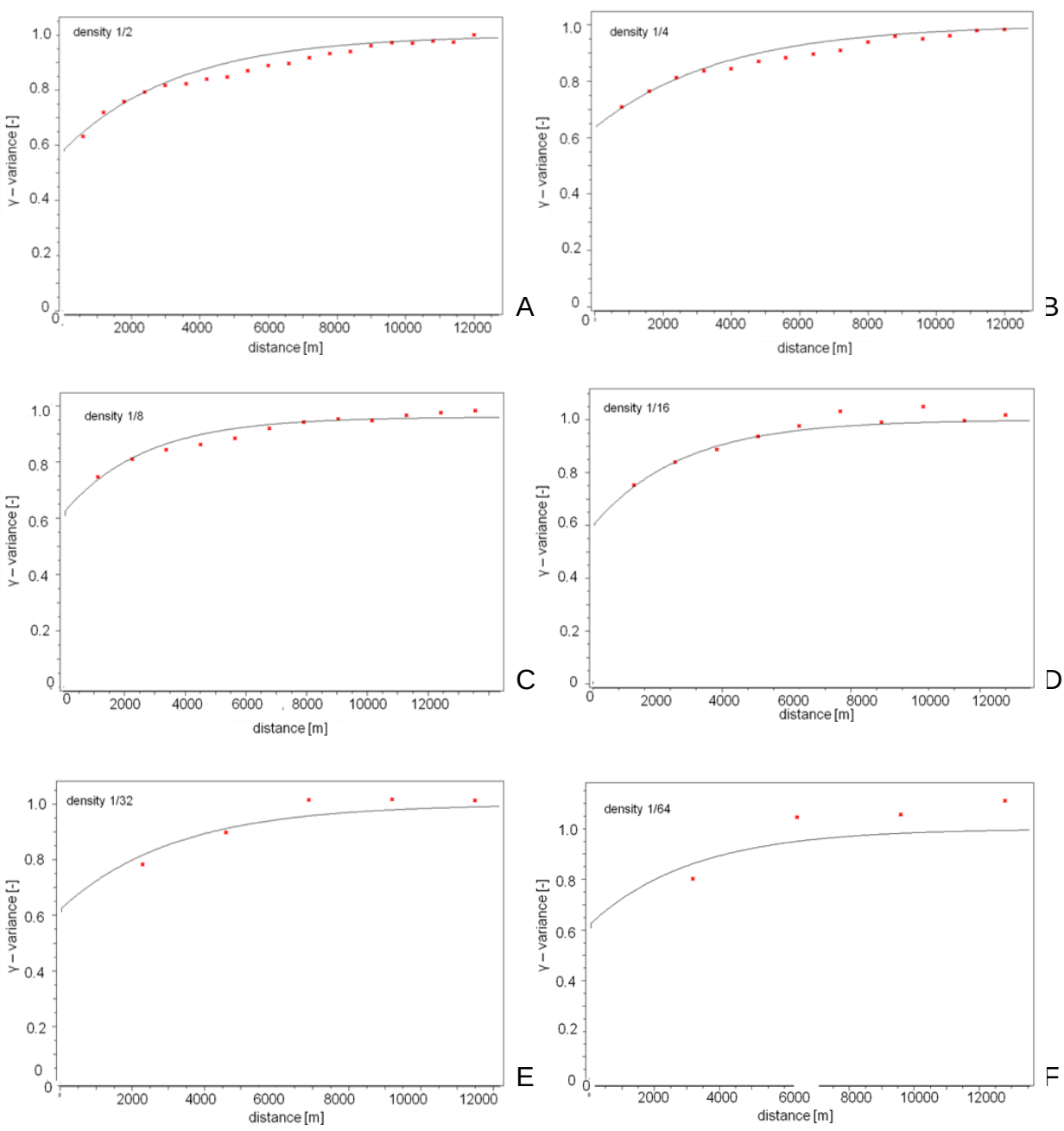


Figure 5-5. Nscore variograms for land elevation datasets. A) density 1/2, B) density 1/4, C) density 1/8, D) density 1/16, E) density 1/32, F) density 1/64.

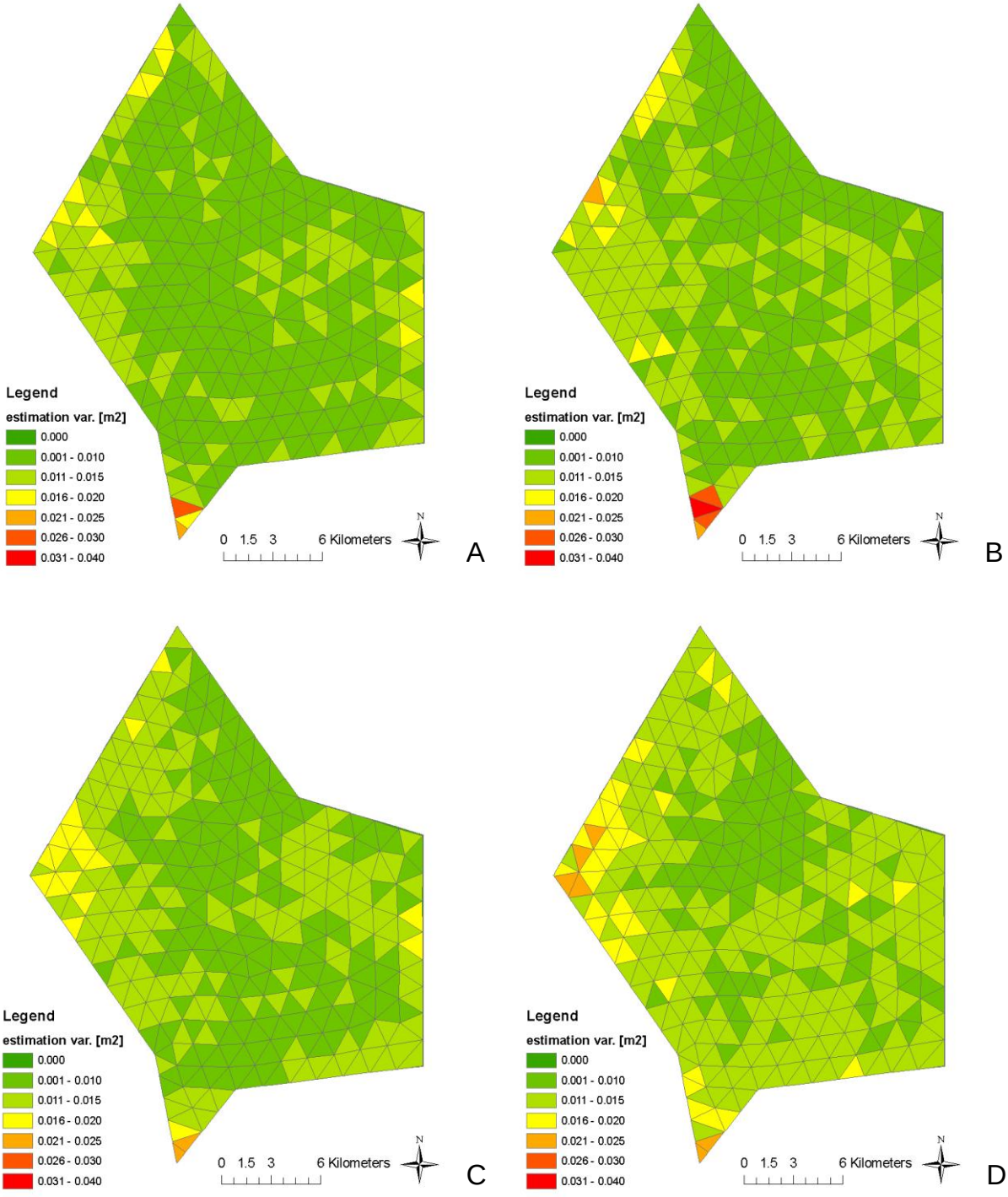


Figure 5-6. Example maps of estimation variances. A) density 1, B) density 1/4, C) density 1/8 D) density 1/32

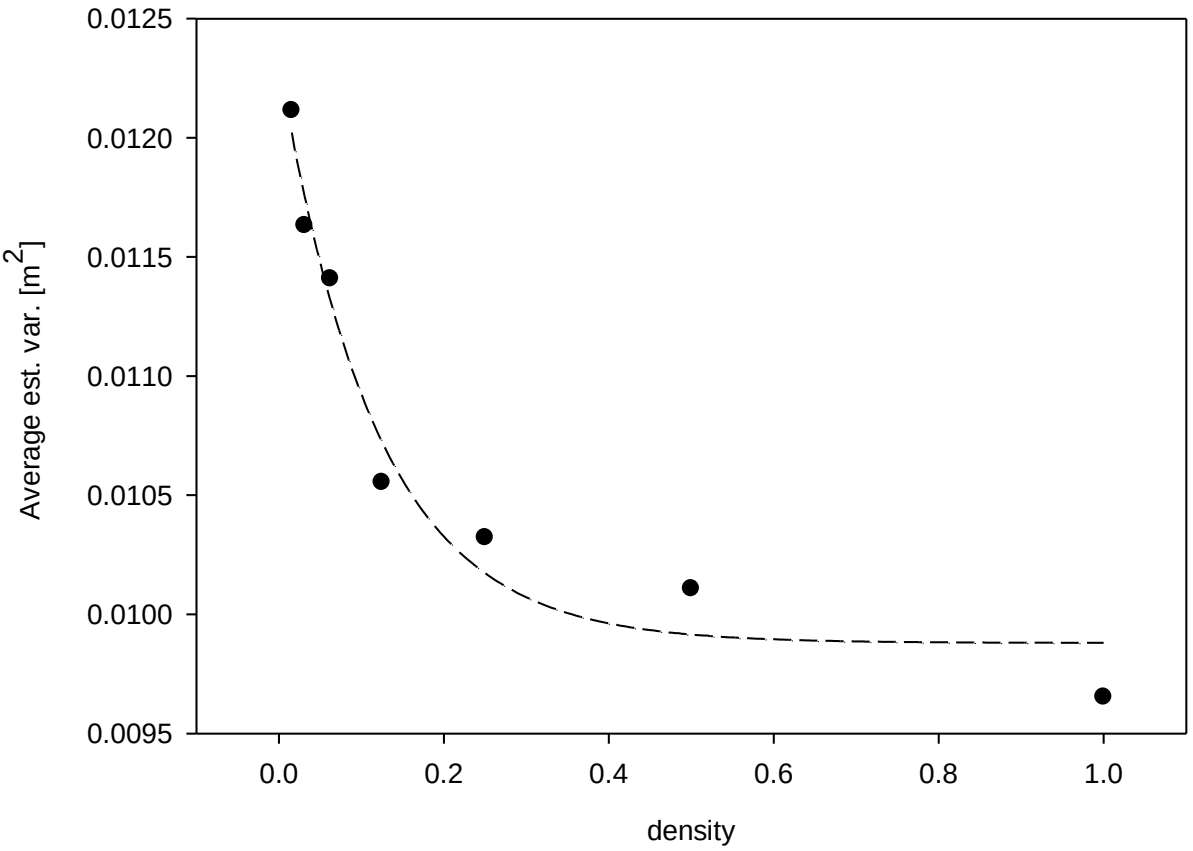


Figure 5-7. Average estimation variance (based on 200maps) for cells vs data density

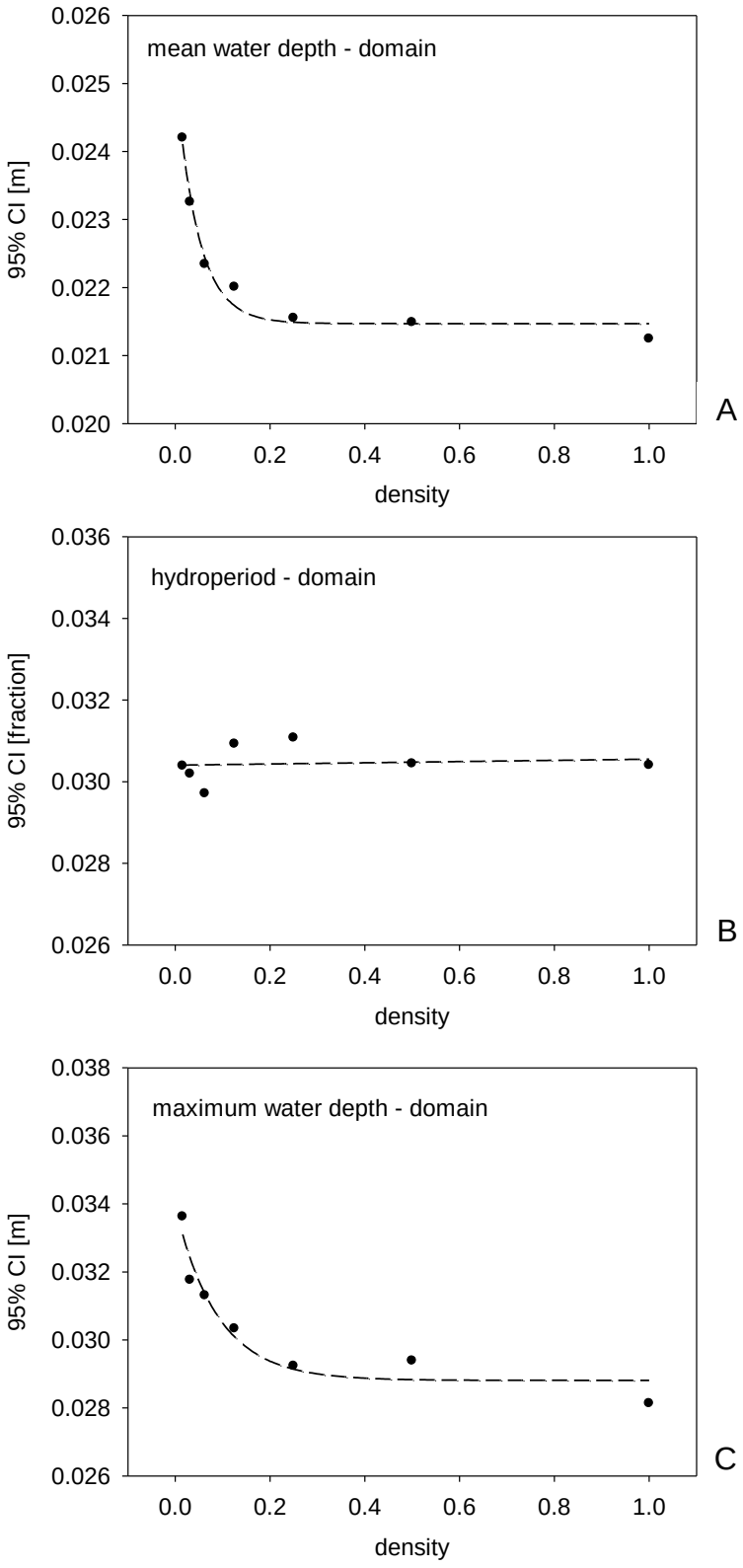


Figure 5-8. Uncertainty results for domain-based outputs. A) mean water depth, B) hydroperiod, C) maximum water depth.

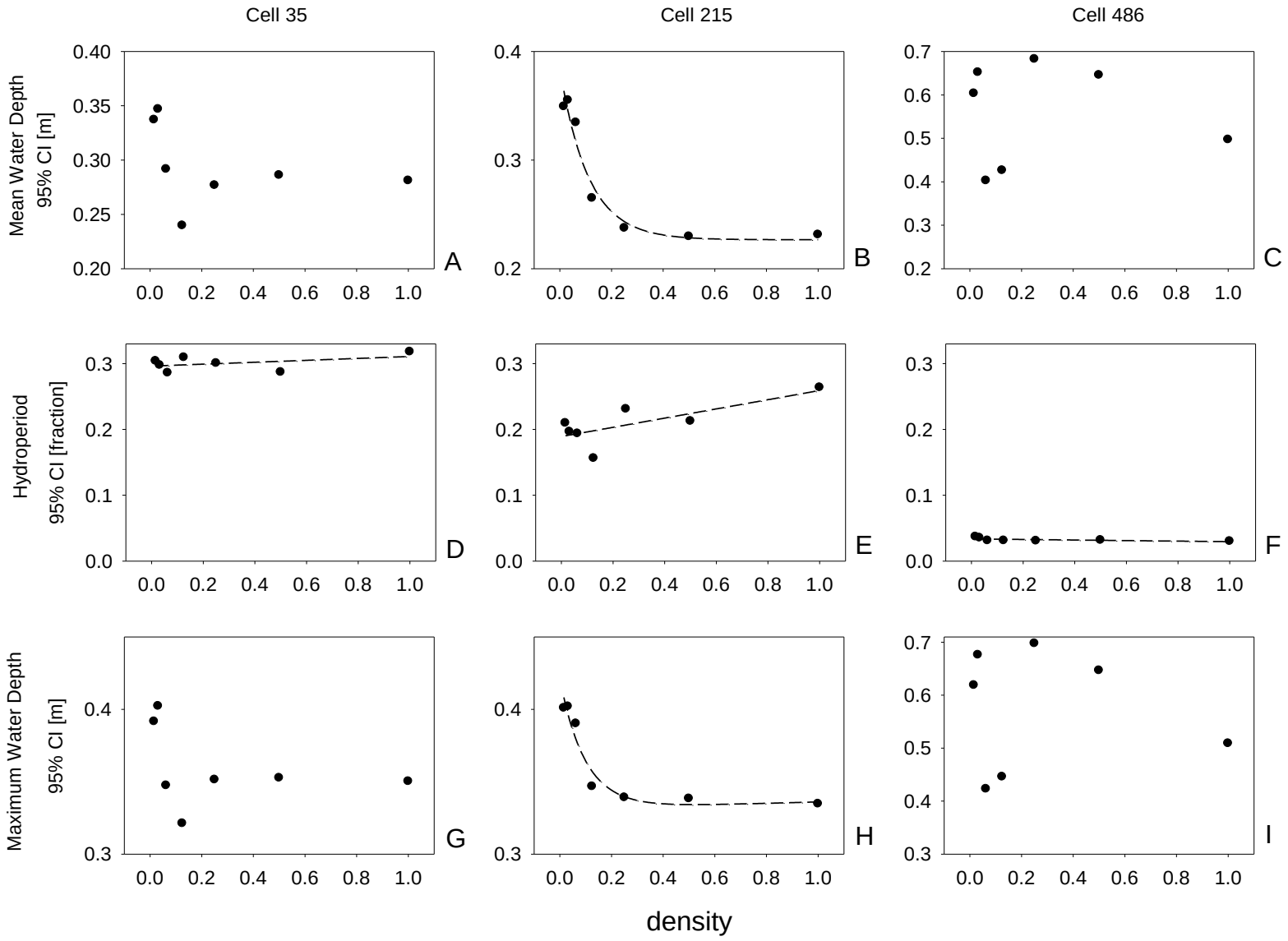


Figure 5-9. Uncertainty results for selected cell-based outputs. A), B), C) mean water depth, D), E), F) hydroperiod G), H), I) maximum water depth.

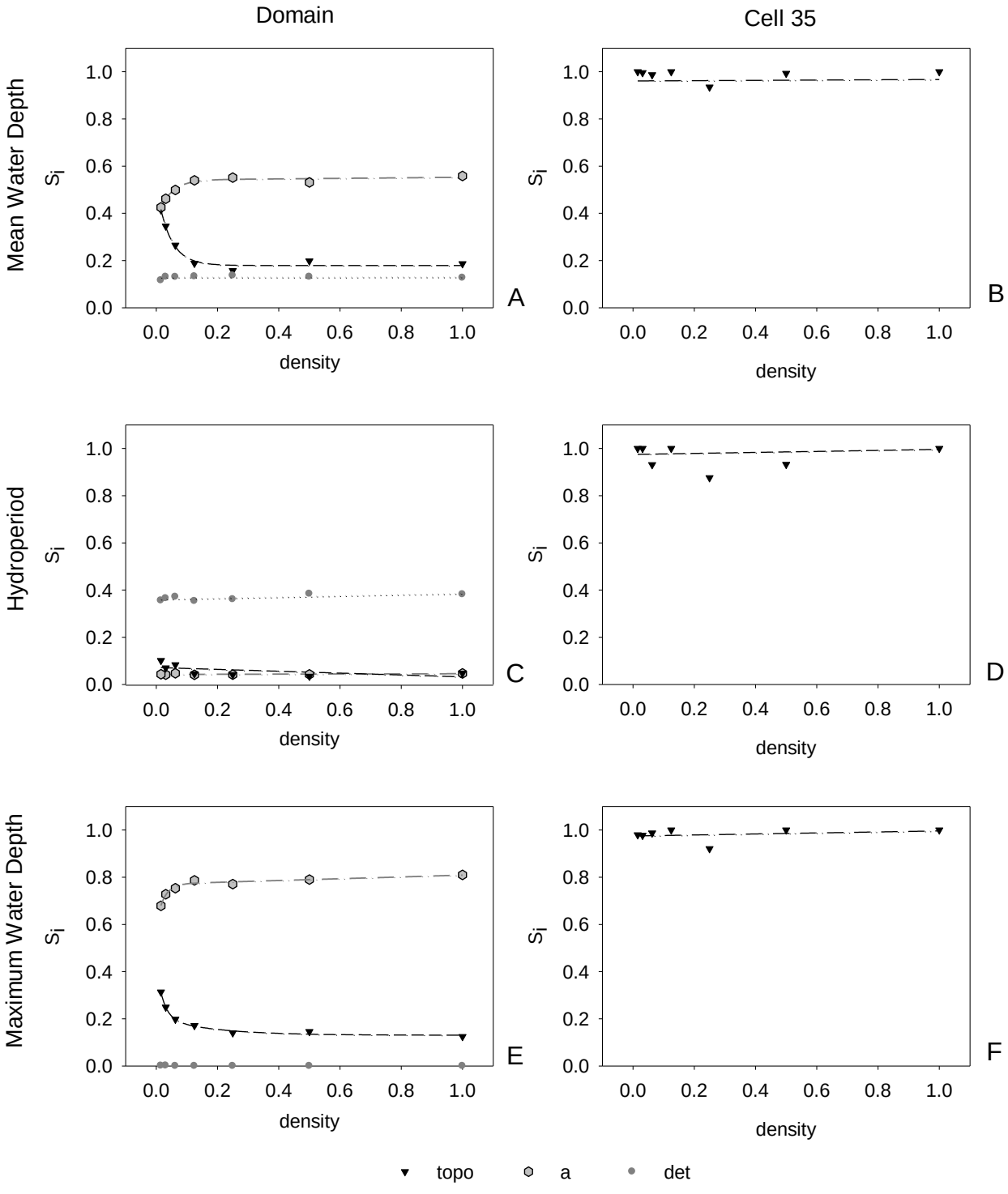


Figure 5-10. Sensitivity results for domain-based outputs (left) and benchmark cell - based outputs (right). A), B) mean water depth, C), D) hydroperiod, E), F) maximum water depth.

CHAPTER 6
SUMMARY

Application of spatially distributed environmental models is currently expanding due to the increased availability of spatial data and improved computational resources. With spatially distributed models, the effect of spatial uncertainty of the model inputs is one of the least understood contributors to output uncertainty and can be a substantial source of errors that propagate through the model. The application of the global uncertainty and sensitivity (GUA/SA) methods for formal evaluation of models is still uncommon in spite of its importance. Even for the infrequent cases where the GUA/SA is performed for evaluation of a model application, the spatial uncertainty of model inputs is disregarded due to lack of appropriate tools.

The central question related to specification of data quality for a modeling process is whether the uncertainty present in model inputs is significant in terms of uncertainty and sensitivity of model outputs. The global uncertainty and sensitivity analysis (GUA/SA) framework can quantify the contribution of uncertain model inputs to uncertainty of model predictions and identify critical regions in the input space (i.e. model inputs that need to be measured or evaluated more accurately), and determine minimum data standards in order for model quality requirements to be met. Furthermore GUA/SA can corroborate model structure, and establish priorities in updating the model, including model simplifications.

The uncertainty regarding spatial structure of model inputs can affect hydrological model predictions and therefore its influence should be evaluated formally in the context of uncertainty deriving from other non-spatial inputs. The framework proposed in this dissertation allows for incorporation of spatial uncertainty of model inputs into GUA/SA.

The proposed framework is based on the combination of variance-based method of Sobol and geostatistical technique of Sequential Simulation (SS). The SS is used for estimation and simulation of spatial variability of input factors. Alternative realizations of inputs are realistic and preserve spatial autocorrelation, since they are conditioned on measured data, global CDF (histogram) and variogram model. Both continuous (land elevation) and categorical (land cover) model inputs are considered. Sequential Gaussian Simulation is used for producing alternative realizations of continuous data, while Sequential Indicator Simulation is applied for categorical inputs. The method of Sobol allows for incorporation of alternative maps into GUA/SA through an auxiliary input factor sampled from the distributed uniform distribution.

The Regional Simulation Model (RSM) and its application to WCA-2A in the South Florida Everglades is used as test bed of the methods developed in this dissertation. RSM simulates physical processes in the hydrologic system, including major processes of water storage and conveyance driven by rainfall, potential evapotranspiration, and boundary and initial conditions. The model domain is spatially represented in a form of triangular elements (cells), which are assumed homogenous in terms of model inputs. The simulations of the RSM are used for support of complex water management and ecosystem restoration decisions in South Florida. The RSM outputs chosen as metrics for GUA/SA for this study are key performance measures generally adopted in the Everglades restoration studies: hydroperiod, water depth amplitude, mean, minimum and maximum. The GUA/SA results for two types of outputs: domain-based approach (spatially averaged over domain), and benchmark cell-based approach are compared. The two kinds of objective function may be used to support various-purpose

management decisions. For example, RSM domain-based results can be more adequate to support decisions of regional scale, like regional water budget assessment. Benchmark cell-based results provide information on local hydrological conditions and they may be used for supporting decisions on ecological restoration (for example restoration of sawgrass communities) in particular locations of WCA-2A.

The general steps in this work include: 1) an initial GUA/SA screening analysis, without consideration of spatial uncertainty of model inputs (Chapter 2), 2) GUA/SA analysis with incorporation of spatial uncertainty of numerical model input (land elevation) (Chapter 3), 3) incorporation of spatial uncertainty of categorical model input (land cover) into the GUA/SA (Chapter 4), and 4) application of the GUA/SA methodology for specification of the optimal data density for the land elevation (Chapter 5).

As the first step in this study (Chapter 2) the traditional GUA/SA is applied to RSM and WCA-2A application, using spatially fixed model inputs. The results of this screening analysis are used as a reference for more advanced methodology, i.e. incorporating spatially distributed inputs, developed in this dissertation. The screening is applied using the modified method of Morris. This method is characterized by a relatively small computational cost and it is applied for identification of important and negligible model inputs. The qualitative screening results indicate that, out of the 20 original model inputs, 8 inputs are important for the model outputs considered. Input factor *topo*, characterizing land elevation uncertainty (for the screening analysis, expressed as vertical shift of land elevation values) is identified as the most important factor in respect to most of the outputs (both domain-based and benchmark cell-based).

Other important factors include: factors *a* and *det* (conveyance parameters), factor *imax* (precipitation interception parameter), factor *kds* (levee hydraulic conductivity), and factor *leakc* (leakage coefficient for canals). Small interactions between parameters are observed, indicating that the model is of additive nature. Since land elevation is identified as one of the most important model inputs this model input is used as an example of spatially distributed numerical model input.

The incorporation of spatial uncertainty of a numerical model input (land elevation) into GUA/SA (Chapter 3) shows that the choice of objective functions used for GUA/SA has significant impact on analysis results. The domain-based outputs are characterized with smaller uncertainty (95% Confidence Interval PDF) than their cell-based counterparts. For example, for the domain-based mean water depth the 95%CI is 0.02 m whereas the 95%CI for the mean water depth for benchmark cells ranges from 0.28 m to 0.5 m depending on the cell location in the domain. The uncertainty regarding hydrological outputs for specific cells is large enough to induce incorrect conclusions and decision, regarding small-scale projects, as it is discussed in Chapter 3. The uncertainty of the domain-based outputs, although small compared to cell-based results may be still important factor affecting decision making process on regional-scale projects, given the very smooth relief in the area. The smaller variation of the domain-based model response can be explained by two factors: spatial averaging of raw model outputs calculated for each cell over the entire domain, and because WCA-2A is confined within levees, and inflows and outflows are controlled and considered as deterministic for all model runs. On the other hand, the higher uncertainty for benchmark cell-based outputs is related to different water distribution patterns between

model simulations, affected by different land elevation scenarios. Uncertainty results for benchmark cells depend on the location of the cell in the area. For example uncertainty of mean water depth is much larger for the cell 486, located in the southern (inundated) part of the domain, than for cell 35, located in the northern (drier) part.

GSA results for the majority of domain-based outputs indicate that the most important factors are factor *a*, used for calculating Manning's roughness coefficient for mesh cells, factor *topo*, representing spatial uncertainty of land elevation and factor *det*, specifying detention depth. The results confirm that spatial uncertainty of model inputs (land elevation) can indeed propagate through spatially distributed hydrological models and can be an important factor, affecting model predictions. The GSA results for benchmark cells show that uncertainty of benchmark cell-based outputs is attributed to the variability of land elevation maps, represented by the factor *topo*. Similarly, to the screening analysis results, no interactions are observed, confirming the additive nature of the RSM for this application.

The procedure for incorporation of spatial uncertainty of categorical model inputs into GUA/SA is proposed in Chapter 4. For the purpose of this study it is assumed that land cover maps may affect model outputs by delineation of ET parameter zones, and Manning's n zones. Five land cover classes, used in the application are externally associated with the corresponding Manning's roughness zones (i.e. parameter *a* zones). For both the Manning's n and ET parameters two types of uncertainties are considered independently: spatial uncertainty of parameter zones (related to spatial uncertainty of land cover classes), and uncertainty of parameters assigned to each of the zones. The ET factors, associated with each of the land cover classes, are varied

within ranges based on the physical limitations, expert opinion, or ±20% of calibrated value, in case no other information is available. With these assumptions, the results of the analysis show that spatial uncertainty of land cover affects RSM domain-based model outputs through delineation of Manning's roughness zones more than through ET parameters effects. In addition, the spatial representation of land cover has much smaller influence on model uncertainty when compared to other sources of uncertainty like spatial representation of land elevation, or the uncertainty ranges for the parameter *a*.

Spatial data collection efforts can be optimized by specification of minimum data requirements for a given model application. In Chapter 5, a hypothetical negative, nonlinear relationship between model uncertainty and source data density is developed and tested. The GUA/SA with incorporation of spatial uncertainty is applied for identification of minimum spatial data requirements (data density) for land elevation. Source data density is found to affect spatial uncertainty of topography maps, used as alternative model inputs, and consequently the hydrological model outputs. Comparative GUA/SA results for the 7 land elevation densities show that domain-based outputs (mean water depth and maximum water depth) are impacted by the density of land elevation data. The results corroborate the hypothetical relationship between model uncertainty and source data density. The inflection point in the curve is identified for the data density between 1/4 and 1/8 of original data density. It is postulated that the inflection point is related to the characteristics of the spatial dataset (variogram) and the aggregation technique (model grid size). Sensitivity analysis results indicate that contribution of land elevation to the domain-based outputs variability (mean water depth

and maximum water depth) shows similar pattern as the uncertainty results. In case of benchmark cell-based outputs, generally no clear trend is observed between output uncertainty and data density. Based on the comparative results for the considered land elevation densities, it is concluded that the reduced data density (up to 1/8 of original land elevation data points) could be used for simulating the WCA-2A application with RSM, without significantly compromising the certainty of model predictions and the subsequent decision making process. The results of this chapter illustrate how quantification of model uncertainty related to alternative spatial data resolutions allows for more informed decisions regarding planning of data collection campaigns.

In general, results for this dissertation show that the main controls of the system identified as important by the GUA/SA (like land elevation and conveyance parameters) are justifiable from the conceptual perspective. This constitutes further corroboration of the RSM behavior.

Limitations

The GUA/SA results are based on the set of assumptions, on the specification of uncertainty models for model input factors, and the interpolation and aggregation methods used for spatial data, as well as the nature of the selected outputs (domain vs. cell-based). Furthermore the GUA/SA techniques have high computational cost and abundant spatial data is required for construction of variograms.

Future Research

Since the framework proposed in this dissertation could be applied to any spatially distributed model and input, as it is independent from model assumptions, the general relationship between spatial model uncertainty and spatial data quality could be further

examined by application of the GUA/SA with Sequential Simulation for other spatial models and applications. Specific focus should be given to the identification of a functional relationship for optimal data density for a given model resolution (grid size) using spatial input semivariogram characteristics. In addition, the effect of model resolution (cell size) and aggregation methods could be further explored.

APPENDIX A
RSM GOVERNING EQUATIONS

The finite volume method is built around governing equations in integral form (SFWMD, 2005a). The Reynolds transport theorem is at the core of the RSM model. Reynolds transport theorem is generally used to describe physical laws written for fluid systems applied to control volumes fixed in space. More recently, it has been used as a first step in the derivation of many conservative laws in partial differential equation form (Chow et al., 1988). The Reynolds transport theorem is expressed for an arbitrary control volume (Figure A-1) as:

$$\frac{D\mathbf{N}}{Dt} = \frac{\partial}{\partial t}\int_{cv} \eta\rho dV + \int_{cv} \eta\rho\left(\mathbf{E}\times\mathbf{n}\right)dA \qquad \text{(A-1)}$$

where: N = an arbitrary extensive property such as the total mass; η = arbitrary intensive property, or property per unit mass such as concentration; E = flux vector; n = unit normal vector; dV = volume element; dA = area element; cv = control volume; and cs = control surface. Variables N and η can be vectors or scalars. This representation of Reynolds transport theorem can be used to write any conservation law with the application of different assumptions. For example, in the case of mass balance, η = 1, and in the case of momentum, η = ux + vy in Cartesian coordinates in which u and v are the velocity components in x and y directions (SFWMD, 2005a).

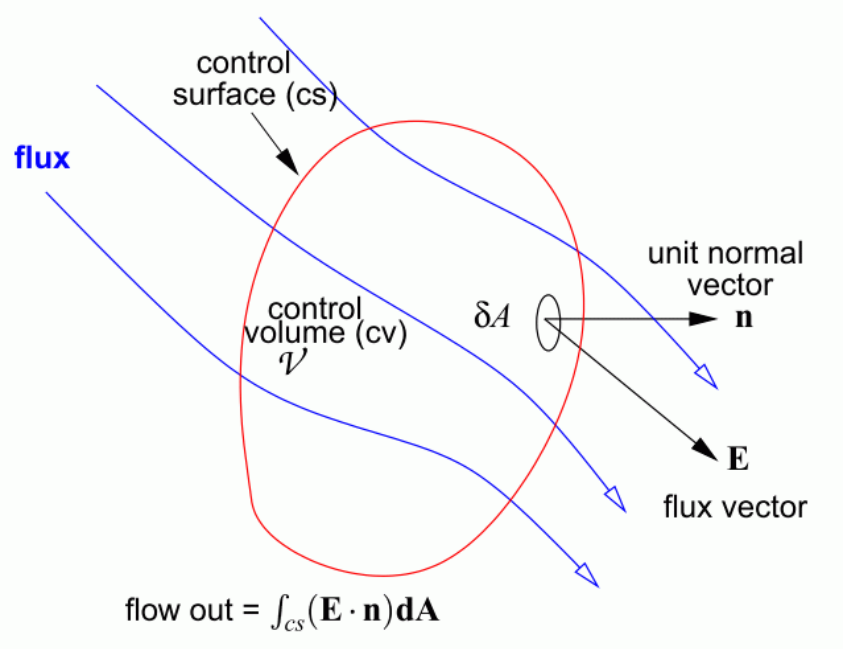


Figure A-1. An arbitrary control volume, after RSM Theory Manual (SFWMD, 2005a)

APPENDIX B
INPUT FACTOTS FOR THE GUA/SA

RSM inputs include dynamic data such as historical rainfall, estimated evapotranspiration, and boundary conditions as well as static data such as topography, land cover, and aquifer thickness. Input parameters include groundwater parameters such as hydraulic conductivity, storage coefficient, seepage parameters, and surface water parameters such as Manning's coefficient. All model inputs, considered as uncertainly sources in this analysis are presented in Table 2-1 in Chapter 2.

All model inputs required for running RSM-HSE are provided in XML files specified in the DTD (document type definition) file. The purpose of a DTD is to define the legal building blocks and structure of an XML document. The RSM-HSE input factors for the WCA-2A application are organized into logical groups represented by the XML main elements under <root>, that are <control>, <mesh>, <watermovers> defined in Table B-1, below. Location of all model inputs, considered in the GUA/SA is provided in Table B-2.

A brief description of these inputs is provided below:

topo - represents land elevation map. Unique land elevation values are assigned on the cell basis. The elevation values are assigned to each cell in the file containing a list of values. Different approaches for modeling the uncertainty of this factor are considered in this dissertation. In the screening analysis in Chapter 2, the topography from the original XML file is modified during the simulations by a Linux batch script. The parameter *topo* characterizes error around land elevation values; it is generated in Simlab from the Gaussian distribution and added to the original topography values (the

same value of error is added to all cells). In the GUA/SA analysis with incorporation of SGS, the facto *topo* is an auxiliary factor, associated with maps generated by the SGS.

bottom - specifies the elevations of aquifer bottom; it is assigned to each cell individually in the file containing a vector of values. The uniform distribution with range ± 20% of the base value (value for a cell from the calibrated model application) is used due to lack of information on the bottom uncertainty in the WCA-2A. For analysis simplicity, the unit multiplier: $mult_{BOTTOM}$ is used as an actual parameter in the Simlab analysis.

$value_{shead}$ - specifies the initial head of water in the domain. This is a lumped parameter with normal distribution with μ = base value from the calibrated model and σ = 0.374 ft . The variance of water depth measurements, applied here, is derived from the USGS report: Initial Everglades Depth Estimation Network (EDEN) digital elevation model research and development (Jones and Price 2007).

a - a parameter used for calculating the Manning's n for model cells. The RSM-HSE defines Manning's n using the following equation:

$$n = ad^{b} \quad \text{(B-1)}$$

where: *d* - water depth, and, *a*, *b* - empirical constants, *b* is fixed to -0.77.

det - represents the detention storage for a cell and defines the minimum depth of surface ponding required in order to produce overland flow. The detention storage accounts for the micro-topography not represented by the topography defined by the

scale of the cells. The detention storage basically acts as a switch. When the ponding is less than the detention storage then the overland flow is set to zero. When the ponded water exceeds the detention storage overland flow occurs.

kveg - specifies the vegetation crop coefficient. The crop coefficient defines plants maximum capability to transpire water. The coefficient is not directly measurable and can only be determined through calibration. The same value of *kveg* is used for all year. This parameter, similarly to other ET parameters is presented in Figure B-1.

xd - defines the extinction depth, i.e. the water table depth at which ET ceases to remove water from the water table and vadose zone. The ET crop correction factor (Figure B-1) linearly approaches zero starting from the root depth at which point the ET factor is defined as *kveg*. In the HSE formulation the extinction depth accounts for the dwindling number of roots at depth by further reducing the ET factor and thus the ET rate for the cell. This is a calibration parameter. There is no direct measurement of the extinction depth. In the current analysis *xd* is treated as regional variable, associated with land cover type, and the level approach is used: a level parameter (*xd* value for cattail) is used to derive *xd* values for other land cover types.

kw - specifies the maximum crop coefficient for open water, the same for all land cover types.

pd - describes the open water ponding depth. In the current analysis the level approach is used for 4 different *pd* parameters, associated with different land use types: cypress, freshwater marsh, sawgrass, and cattail; *pd* for cattail is used as the level parameter.

imax - characterizes the maximum interception. In the current analysis the same range of *imax* is assigned for all land uses.

rd - defines the shallow root zone depth . Currently two different distributions are assigned to low vegetation areas (cattail, sawgrass, marsh) and to cypress tree areas: rd_G (for grasses) and rd_{CY} (for cypress).

hc - specifies the aquifer hydraulic conductivity. Hydraulic conductivity values are assigned to each cell individually in the file containing a vector of values. The hydraulic conductivity is assumed to be spatially independent due to large variability at the cell scale. The lognormal distribution is fitted to all non-boundary cell values reported in the domain.

sc - represents the storage converter. Stage-volume converters have been developed to allow a more accurate representation of the volume of water stored at different water levels. Depending on the area under water, wetlands can store variable amounts of water at various depths. A flat ground with a designated storage coefficient below ground level and the assumption of open water above ground level is generally a poor

representation of wetland storage conditions. However, this has been the standard method used to conceptualize water storage above and below ground.

n - Manning's Roughness Coefficient for canals

leakc - defines the leakage coefficient, and is used for computing flow between the aquifer and the canal (leakc=k/δ) using the following equation.

$$q = \text{leakc} \times p(H - h) \tag{B-2}$$

where: q = seepage flow per unit length of the canal, k = hydraulic conductivity of bottom sediment, δ = thickness of the sediment layer, p = wetted perimeter of the canal h = water level in the canal segment, H = water level in the cell.

bankc - used for calculating overland flow between canal segment and a cell . The overland flow is modeled as a weir flow over a "lip" along the edge of the canal segment. The overland flow is calculated from equation:

$$Q = CL\sqrt{gh}^{1.5} \tag{B-3}$$

where: C = bankc - weir coefficient, L - length of overlap between the segment and the cell, h – difference between canal head and leap height.

kmd - specifies the levee seepage, i.e. levee hydraulic conductivity from a marsh cell to a dry cell. There are 4 different values of *kmd* assigned to different canals in the

application (L35B, L36, L6, and L38E), the parameter *kmd* for L38 is used as a level parameter.

kds - specifies the levee seepage, i.e. levee hydraulic conductivity from a dry cell to a segment. There are 4 different values of *kds* assigned to different canals in the application (L35B, L36, L6, and L38E), the parameter *kds* for L38 is used as a level parameter.

kms - specifies the levee seepage, i.e. levee hydraulic conductivity from a mash cell to a segment. There are 4 different values of *kms* assigned to different canals in the application (L35B, L36, L6, and L38E), level the parameter *kms* for L38 is used as a level parameter.

Table B-1: Main XML elements in the WCA-2A application.

XML element	Description
<control>	All the program control parameters such as time step size, beginning time, ending time, etc. are defined using this XML element.
<mesh>	Information regarding the 2-D mesh, land input factors
<network>	Information regarding the canal network
<watermovers>	Water movers such as structures are defined here; levee seepage

Table B-2: Location of inputs in XML input structure

#	Model Input	XML Structure Location
1	$value_{shead}$	<mesh><shead>
2	topo	<mesh><surface>
3	bottom	<mesh><bottom>
4	hc	<mesh><transmissivity>
5	sc	<mesh><svconverter>
6	kmd	<watermovers><leveeSeepage>
7	kms	<watermovers><leveeSeepage>
8	kds	<watermovers><leveeSeepage>
9	n	<network><arcs><arcflow>
10	leakc	<network> <arcseepage>
11	bankc	<network><arcoverbank>
12	a	<mesh><conveyance>
13	det	<mesh><conveyance>
14	kw	<mesh><hpm><layer1nsm>
15	rdG	<mesh><hpm><layer1nsm>
16	rdC	<mesh><hpm><layer1nsm>
17	xd	<mesh><hpm><layer1nsm>
18	pd	<mesh><hpm><layer1nsm>
19	kveg	<mesh><hpm><layer1nsm>
20	imax	<mesh><hpm><layer1nsm>

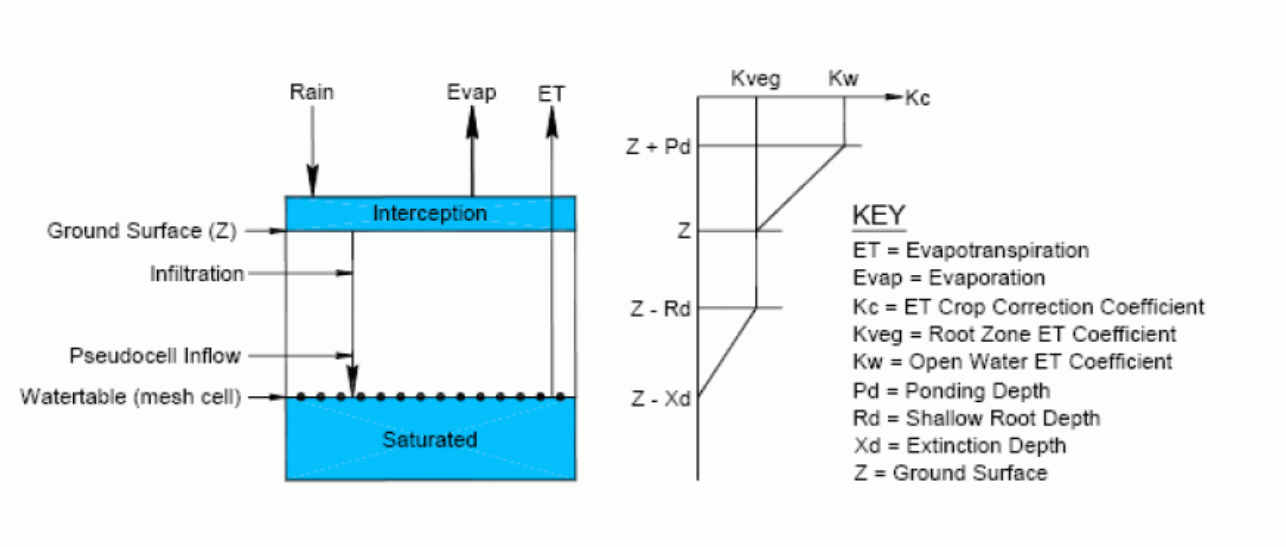


Figure B-1: Parameters used for modeling ET in RSM (RSM-HSE User Manual, 2005b).

APPENDIX C
SPATIAL STRUCTURE OF MODEL INPUTS

The spatial representation of model inputs may range from spatially lumped, through regionalized to fully distributed. Some of the factors are spatially lumped, i.e. only one value of the factor is assigned for the whole domain, and in such case the generated values of input factors are substituted for the model parameter and used for model simulations. Other factors, like parameter *a*, are regionalized. In such case, the value of the parameter varies between zones in the domain. The so called "level parameter" approach is used for the zonal parameters in order to reduce the number of input factors used for the analysis. In this approach values for a parameter in one zone are generated from the assigned PDF, and the parameter values in other zones are obtained from the initial ratio of parameter values in different zones. Another group of factors are fully spatially distributed (e.g. hydraulic conductivity), the sample level approach is applied for these factors, with a parameter for one cell being generated. The values for other cells are obtained by preserving the initial ratio with the selected cell. The spatial representation of model input factors (lumped, regional or fully distributed) is conditioned on the structure of input files associated with model inputs.

An example of the level parameter approach is provided for the regionally varied parameter *a* for calculating Manning's *n*. Six regions (zones) are delineated, each of the zones characterized by different value of the parameter (Figure 2-2 A, Table C-1). Parameter *a* for each zone could be considered as a separate input factor in the GUA/SA, however this approach would increase the overall number of input factors and the computational requirements for the analysis (especially if applied to all regionalized model inputs). In order to make the GUA/SA more efficient, all zones for parameter *a*

are represented by the same input factor (in this case factor *a* for zone 2). Value of parameter *a* for all other zones are obtained from the MC realizations generated for parameter *a* in zone 2, by preserving the original relationship between parameters (i.e. relationship from the calibrated model).

The original XML file for the WCA-2A application with the values of parameter *a* for 6 Manning's roughness zones is presented in Figure C-1. The input factor *a* is assigned a uniform PDF with ± 20% (around the base value of *a* for zone II), and values of *a* for other zones II-VI are obtained by preserving the original relationship of base values (Table C-1). The values of parameter *a* for zones II-VI (a2-a6) are substituted in the input file using AWK script shown in Figure C-4. Figure C-2 presents XML file that is used for substituting the values, generated by the MC simulations. The indexed file, with the format presented in Figure C-3 is used to specify which Manning's roughness zone is assigned to each cell. Similar level approach is used for other zonal parameters (ET parameters: *kveg*, *kw*, *rd*, levee seepage parameters: *kmd*, *kms*, *kds*) and for fully distributed hydraulic conductivity (*hc*).

Table C-1: Ranges of parameter *a*, assigned to different vegetation density zones in the WCA-2A in the calibrated model.

Zone	Base value *a*	# of cells
I	0.1[1]	125
II	0.3	50
III	0.33786	62
IV	0.5	63
V	0.7	103
VI	0.9	106

[1] The values for zone I - the boundary cells - are fixed in the GUA/SA analysis.

```
<conveyance>
  <indexed file="./input/zone_wca2_10-29-2007.xml">
   <entry id="1" label="Zone I"> <!-- Boundary Cells -->
     <mannings a="0.1" b="-0.77" detent="0.11"></mannings>
   </entry>
   <entry id="2" label="Zone II">
     <mannings a="3.0000E-01" b="-0.77" detent="0.1"></mannings>
   </entry>
   <entry id="3" label="Zone III">
     <mannings a="3.3786E-01" b="-0.77" detent="0.1"></mannings>
   </entry>
   <entry id="4" label="Zone IV">
     <mannings a="5.0000E-01" b="-0.77" detent="0.1"></mannings>
   </entry>
   <entry id="5" label="Zone V">
     <mannings a="7.0000E-01" b="-0.77" detent="0.1"></mannings>
   </entry>
   <entry id="6" label="Zone VI">
     <mannings a="9.0000E-01" b="-0.77" detent="0.1"></mannings>
   </entry>
  </indexed>
</conveyance>
```

Figure C-1. Example of original input file for specification of parameter *a* for calculating Manning's *n*

```
<conveyance>
  <indexed file="./input/zone_wca2_10-29-2007.xml">
   <entry id="1" label="Zone I"> <!-- Boundary Cells -->
     <mannings a="0.1" b="-0.77" detent="0.11"></mannings>
   </entry>
   <entry id="2" label="Zone II">
     <mannings a=" a2 " b="-0.77" detent=" det "></mannings>
   </entry>
   <entry id="3" label="Zone III">
     <mannings a=" a3 " b="-0.77" detent=" det "></mannings>
   </entry>
   <entry id="4" label="Zone IV">
     <mannings a=" a4 " b="-0.77" detent=" det "></mannings>
   </entry>
   <entry id="5" label="Zone V">
     <mannings a=" a5 " b="-0.77" detent=" det "></mannings>
   </entry>
   <entry id="6" label="Zone VI">
     <mannings a=" a6 " b="-0.77" detent=" det "></mannings>
   </entry>
  </indexed>
</conveyance>
```

Figure C-2. Example of modified input file for specification of parameter *a* for calculating Manning's *n*

```
OBJTYPE 'mesh2d'
BEGSCL
ND 510
NAME 'zone_wca2_10-29-2007.xml'
TS 0 0
1
1
1
1
1
5
1
1
1
1
1
4

.
.
.

1
1
ENDDS
```

Figure C-3. Structure of the indexed file specifying which Manning's n zone is assigned to each model cell.

```
# create the table of substitutions for this run to be used by
"a_subst" script based on command-line parameters and labels.txt

    exec 3>&1                 #save current stdout as &3
    exec > substitute.tab     #echo to substitute.tab file
    exec < ../labels.txt      #read from labels.txt file

    sample=$1
    shift

    for par in $*
    do
      read lbl
      echo $lbl $par
      case $lbl in
        "a2")
          echo a3 `python -c "print $par * 1.1262"`
          echo a4 `python -c "print $par * 1.666"`
          echo a5 `python -c "print $par * 2.333"`
          echo a6 `python -c "print $par * 3"`
          ;;
        "xdCA")
          echo xdCY `python -c "print $par * 3"`
          echo xdM `python -c "print $par * 0.4"`
          echo xdS `python -c "print $par * 1.5"`
          ;;
        "pdCA")
          echo pdCY `python -c "print $par * 1.666666667"`
          echo pdM `python -c "print $par * 0.666666667"`
          echo pdS `python -c "print $par * 1.166666667"`
          ;;
        "kmdL38E")
          echo kmdL35B `python -c "print $par * 2.210526316"`
          echo kmdL36 `python -c "print $par * 0.442105263"`
          echo kmdL6 `python -c "print $par * .178947368"`
          ;;
        "kmsL38E")
          echo kmsL35B `python -c "print $par * 0.859388646"`
          echo kmsL36 `python -c "print $par * 1"`
          echo kmsL6 `python -c "print $par * 2.082969432"`
          ;;
        "kdsL38E")
          echo kdsL35B `python -c "print $par * 3.443786982"`
          echo kdsL36 `python -c "print $par * 1"`
          echo kdsL6 `python -c "print $par * 9.097633136"`
          ;;
```

```
        "hc333")
          ../../common/doMath.sh input/hyd_con.xml "*$par" >
hyd_con.xml
          ;;
        "topo")
          cp ../topomaps/200/1/$par.txt topo_wca2.xml
          ;;

      esac
    done

    exec 1>&3  #echoing to default stdout (screen)

    # Substitute parameters into the XML input files for this
simulation
    ../../common/a_subs ../run_wca2_gms.xml > run_wca2_gms.xml
    ../../common/a_subs input/canal_index.xml > canal_index.xml
    ../../common/a_subs input/mann_wca2_10-29-2007.xml >
mann_wca2_10-29-2007.xml
    ../../common/a_subs input/evap_prop_hpm.xml >
evap_prop_hpm.xml
    ../../common/a_subs input/levee_seep_123.xml >
levee_seep_123.xml

    #run hse for this sample combination
    /apps/rsm/2961/src/hse run_wca2_gms.xml > /dev/null

    # check line count in output
    linecnt=`wc -l wca2_pond.gms | awk '{print $1}'`
    echo "$sample" "$linecnt" >> linecnt.txt
    if [ "$linecnt" -lt 3359830 ]
    then
      # log error
      echo "$sample" "$linecnt" >> errors.txt
      mv wca2_pond.gms wca2_pond"$sample".gms
    else
      # process and save the model output
      echo -n "$sample " >> sensitivityMulti.out
      echo -n "$sample " >> sensitivityDomain.out
      ../../common/doOutputMulti.sh wca2_pond.gms >>
sensitivityMulti.out
      ../../common/doOutputDomain.sh wca2_pond.gms >>
sensitivityDomain.out
    fi
```

Figure C-4. AWK script used to substitute parameters in model input files.

APPENDIX D
POST-PROCESSING MODEL OUTPUTS

Output provided by the HSE-RSM (water depth) is generated on a daily time step basis for each model cell. The raw model outputs are aggregated into performance measures, selected in this study. The model outputs chosen as metrics for the sensitivity and uncertainty analysis are the performance measures generally adopted in the Everglades restoration studies (SFWMD, 2007): 1) hydroperiod (here defined as a percent of time a given area is inundated); 2) seasonal water depths (mean, maximum and minimum), and 3) seasonal amplitude (the difference between average annual maximum depth and average annual minimum depth over period of simulation).

Raw outputs are post-processed using scripts in AWK programming language. For the domain-based outputs the following steps are performed using the script presented in Figure D-1: 1) raw output values (daily water depth reported for each cell) is averaged over the domain's space; 2) annual mean, minimum, maximum and amplitude are calculated from the spatially averaged daily values, 3) seasonal (simulation period) averages are calculated from the annual values. For benchmark-cell based outputs – processed using the script presented in Figure D-2 – the first step is omitted, therefore the raw results are reported for each cell (i.e. they are averaged only over simulation time).

```
awk '
# step - day of year
# count - total no of days from start
# cell - base + current cell no
# base - starting index used in min,max,... arrays
# leap=4 means a leap year
# period - number of days in year
```

```
BEGIN {
  step = 0; count = 0; base = 0; leap = 1; period = 365; sum =
0; above = 0;
  }

# skip first year
NR <= 186520 {next; }

$1 == "TS" {
  if (step++ == period) {
    #print "step " step-1;
    step = 1;
    base = cell;
    if (leap++ == 4) {
      leap = 1;
      period = 366;
      }
    else
      period = 365;
    }
  cell = base;
  next;
  }

step == 0 {next; }
{sum += $1; cell++; count++; }
$1 > 0 {above++; }
step == 1 {min[cell] = $1; max[cell] = $1; next; }
$1 < min[cell] {min[cell] = $1; }
$1 > max[cell] {max[cell] = $1; }

END {
  summin = 0;
  summax = 0;
  for (i=1; i<=cell; i++ ) {
    summin += min[i];
    summax += max[i];
    }
  #if (cell == 0 || count == 0) {print cell count >
"error.txt"};
  print sum/count " " above*100/count " " summin/cell " "
summax/cell " " summax/cell-summin/cell;
  }
' "$@"
```

Figure D-1. AWK script used to calculate domain-based outputs.

```
awk '
# step - day of year
# count - total no of days from start
# year - total no of years from start
# cell - base + current cell no
# base - starting index used in min,max,... arrays
# leap=4 means a leap year
# period - number of days in year

BEGIN {
step = 0; count = 0; year = 1; base = 0; leap = 1; period = 365;
benchCells[1] = 35;
benchCells[2] = 48;
benchCells[3] = 147;
benchCells[4] = 180;
benchCells[5] = 215;
benchCells[6] = 355;
benchCells[7] = 120;
benchCells[8] = 178;
benchCells[9] = 224;
benchCells[10] = 244;
benchCells[11] = 279;
benchCells[12] = 288;
benchCells[13] = 447;
benchCells[14] = 486;
}

# skip first year
NR <= 186520 {next; }

$1 == "TS" {
  if (step++ == period) {
    #print "step " step-1;
    year++;
    step = 1;
    base = cell;
    if (leap++ == 4) {
      leap = 1;
      period = 366;
      }
    else
      period = 365;
    }
  count++;
  cell = base;
  next;
  }
```

```
step == 0 {next;}

# check if benchmark cell
{ cc = ++cell - base;
  notBc = 1;
  for (b in benchCells)
    if (cc == benchCells[b])
      notBc = 0;
  }
notBc == 1 {next; }

step == 1 {min[cell] = $1; max[cell] = $1; sum[cell] = 0;
above[cell] = 0; }
{sum[cell] += $1; }
$1 > 0 {above[cell]++; }
$1 < min[cell] {min[cell] = $1; }
$1 > max[cell] {max[cell] = $1; }

END {
for (b=1; b<=14; b++) {
  bc = benchCells[b];
  sumsum[bc] = 0;
  sumabove[bc] = 0;
  summin[bc] = 0;
  summax[bc] = 0;

  for (i=0; i<year; i++ ) {
    cc = i*510 + bc;
    sumsum[bc] += sum[cc];
    sumabove[bc] += above[cc];
    summin[bc] += min[cc];
    summax[bc] += max[cc];
    }

  #printf "%s",bc " ";
  printf "%s",sumsum[bc]/count " " sumabove[bc]/count " "
summin[bc]/year " " summax[bc]/year " " summax[bc]/year-
summin[bc]/year " ";
  }
print "";
}
' "$@"
```

Figure D-2. AWK script used to calculate benchmark-cell based outputs.

APPENDIX E
ALTERNATIVE RESULTS FOR SGS

This appendix presents alternative results for Chapter 4. The alternative results were obtained in the case when land elevation maps are generated using the Sequential Gaussian Simulation (SGS) with histograms and variograms specific for given data set (density). No general trend is observed for the relationship between average estimation variance and data density. This is attributed to the fact that apart from data density, other factors like different variability of sampled data within datasets affect the spatial uncertainty of generated land elevation realizations.

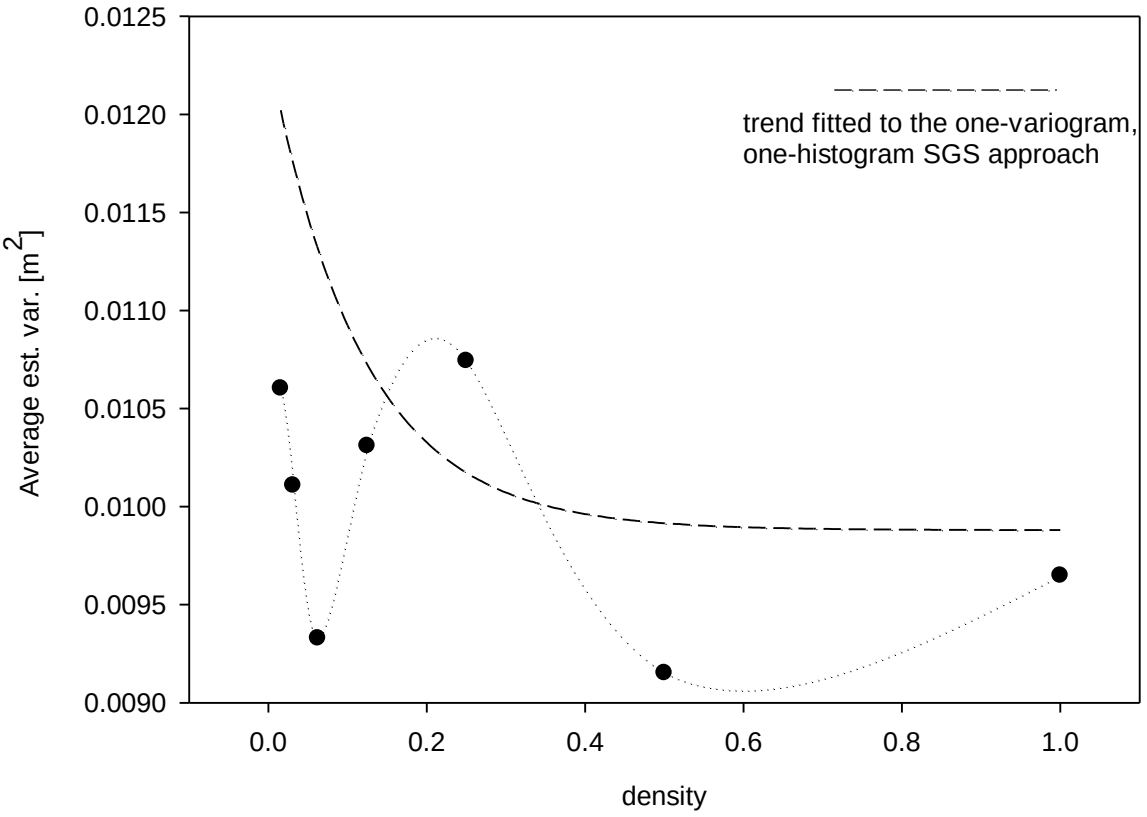
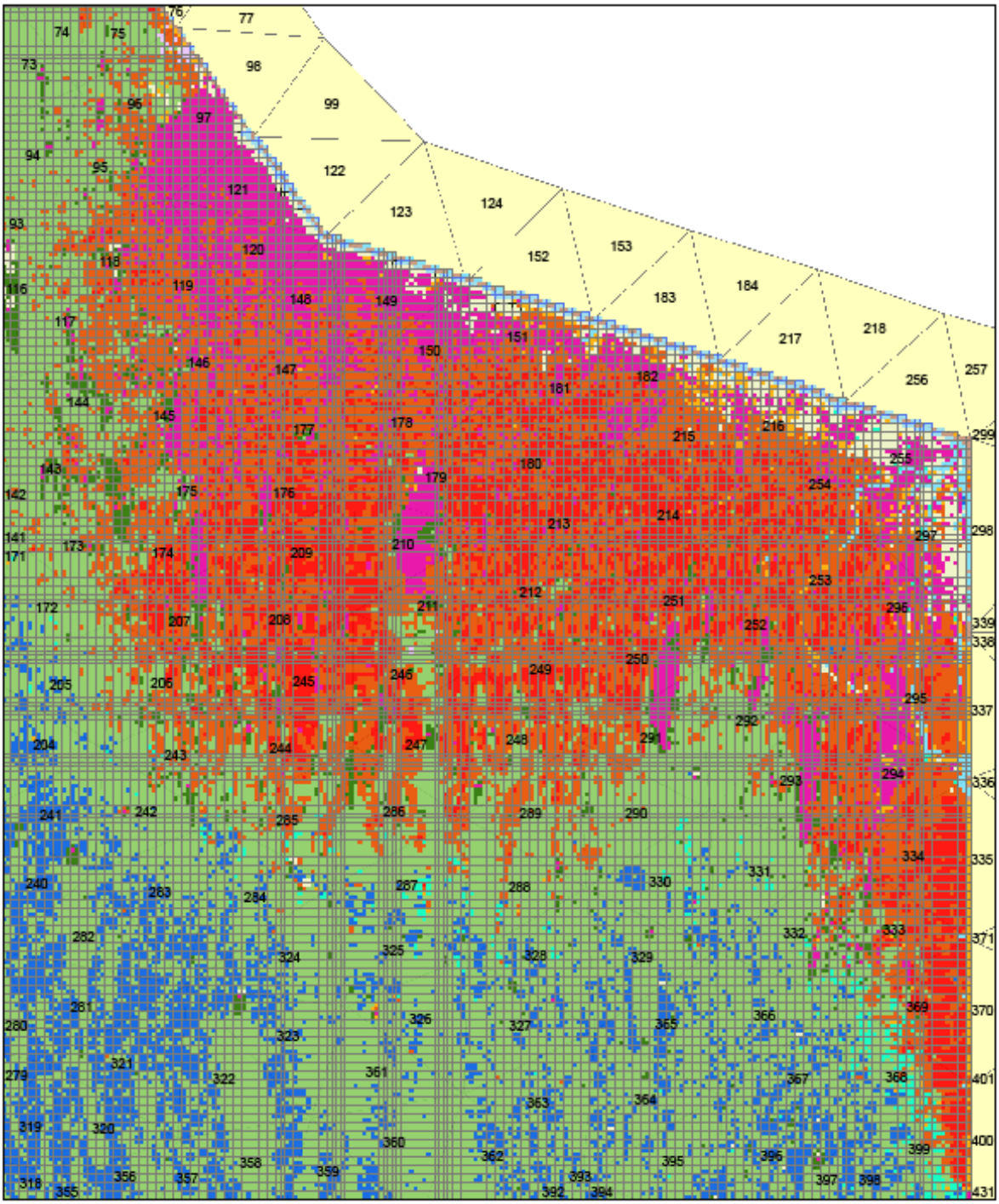


Figure E-1. Average estimation variance versus data density for alternative approach towards SGS.

APPENDIX F
SUPPLEMENTARY VEGETATION INFORMATION

Table F-1. Distribution of vegetation categories for the 2003 WCA-2A vegetation map (after Rutchey et al., 2008).

Grid Category	Area (ha)	Percentage
Trees	51	< 1%
Shrubs	1,400	3%
Scrub	619	1%
Sawgrass	27,638	65%
Open Marsh	5,700	14%
Broadleaf	47	< 1%
Floating	386	1%
Cattail	6,039	14%
Exotics	28	< 1%
Fish Camps	11	< 1%
Spoil Areas and Canals	451	1%
Other	187	< 1%
Total	42,635	100%



vegetation cover E, red - cattail, pink-willow, green -sawgrass,
blue - wet praire slough

Figure F-1. Subsection of the 2003 vegetation map for NE of WCA-2A (cattail invaded areas),

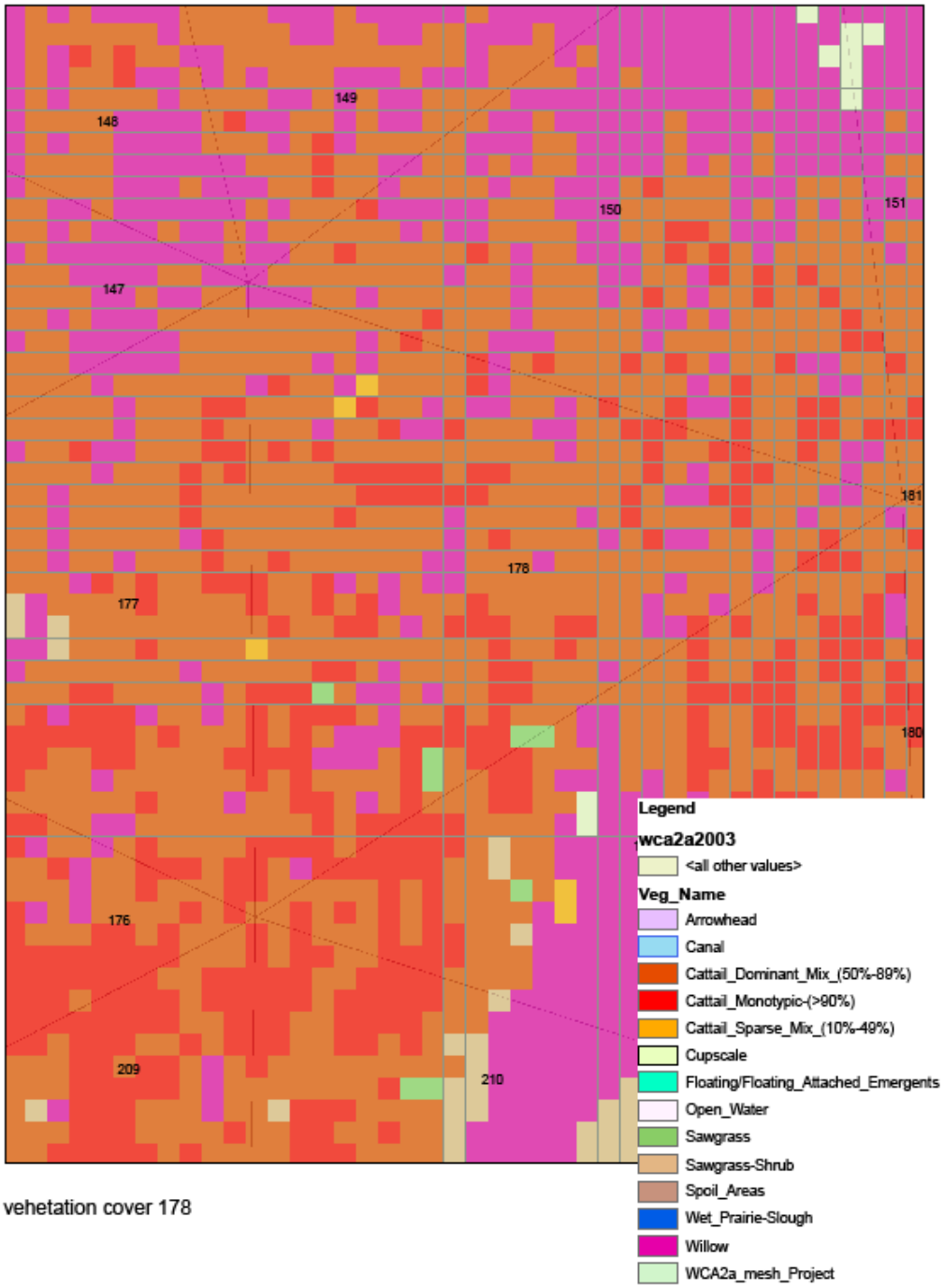


Figure F-2. Subsection of the 2003 vegetation map for cell 178 in the NE of WCA-2A.

LIST OF REFERENCES

Bell V.A., Moore R.J., 2000. The sensitivity of catchment runoff models to rainfall data at different spatial scales. Hydrology and Earth System Sciences 4 (4), 653-667.

Beven K., 2006. On undermining the science? Hydrol.Process. 20 (14), 3141-3146.

Beven K., 1989. Changing ideas in hydrology — The case of physically-based models. Journal of Hydrology 105 (1-2), 157-172.

Burrough P.A., McDonnell R., 1998. Principles of geographical information systems. Oxford University Press, Oxford, New York.

Cacuci D.G., Navon I.M., Ionescu-Bujor M., 2005. Sensitivity and Uncertainty Analysis, Volume II: Applications to Large-Scale Systems. Chapman & Hall/CRC Press, Boca Raton.

Cacuci D.G., Ionescu Bujor M., Navon I.M., 2003-. Sensitivity and uncertainty analysis. Chapman & Hall/CRC Press, Boca Raton.

Campolongo F., Cariboni J., WIM S., 2005. Enhancing the Morris Method.

Campolongo F., Saltelli A., Jensen N.R., Wilson J., Hjorth J., 1999. The Role of Multiphase Chemistry in the Oxidation of Dimethylsulphide (DMS). A Latitude Dependent Analysis. J.Atmos.Chem. 32 (3), 327-356.

Campolongo F., Cariboni J., Saltelli A., 2007. An effective screening design for sensitivity analysis of large models. Environ.Model.Softw. 22 (10), 1509-1518.

Campolongo F., Saltelli A., 1997. Sensitivity analysis of an environmental model: an application of different analysis methods. Reliab.Eng.Syst.Saf. 57 (1), 49-69.

Chaubey I., Cotter A.S., Costello T.A., Soerens T.S., 2005. Effect of DEM data resolution on SWAT output uncertainty. Hydrol.Process. 19 (3), 621-628.

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

CHO Sung-Min, LEE M., 2001. Sensitivity considerations when modeling hydrologic processes with digital elevation model. 37(4).

Chu-Agor M.L., Muñoz-Carpena R., Kiker G., Emanuelsson A., Linkov I., Chu-Agor, M.L., Muñoz-Carpena, R., Kiker, G., Emanuelsson, A. and Linkov, I. Exploring sea level rise vulnerability of coastal habitats through

global sensitivity and uncertainty analysis. Environ. Modell. Soft. .

Cowell P.J., Zeng T.Q., 2003. Integrating Uncertainty Theories with GIS for Modeling Coastal Hazards of Climate Change. Mar.Geod. 26 (1), 5.

Crosetto M., Tarantola S., 2001. Uncertainty and sensitivity analysis: tools for GIS-based model implementation. Int.J.Geogr.Inf.Sci. 15 (5), 415.

Crosetto M., Tarantola S., Saltelli A., 2000. Sensitivity and uncertainty analysis in spatial modelling based on GIS. Agric., Ecosyst.Environ. 81 (1), 71-79.

Cukier R.I., Fortuin C.M., Schuler K.E., Petschek A.G., Schaibly J.H., 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. Part I: Theory. Journal of Chemical Physics 59, 3873-3878.

David P., 1996. Changes in plant communities relative to hydrologic conditions in the Florida Everglades. Wetlands 16 (1), 15-23.

Delbari M., Afrasiab P., Loiskandl W., 2009. Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. Catena 79 (2), 163-169.

DEP, 1999. Southeast District Assessment and Monitoring Program. Ecosummary. Water Conservation Area 2A. Southeast District Assessment and Monitoring Program .

Deutsch C.V., Journel A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, Inc., .

Doherty J., 2004. PEST Model-Independent Parameter Estimation User Manual. 5th Edition. Watermark Numerical Computing .

Endreny T.A., Wood E.F., 2001. Representing elevation uncertainty in runoff modelling and flowpath mapping. Hydrol.Process. 15, 2223-2236.

Fisher P.F., Tate N.J., 2006. Causes and consequences of error in digital elevation models. Prog.Phys.Geogr. 30 (4), 467-489.

Francos A., Elorza F.J., Bouraoui F., Bidoglio G., Galbiati L., 2003. Sensitivity analysis of distributed environmental simulation models: understanding the model behaviour in hydrological studies at the catchment scale. Reliab.Eng.Syst.Saf. 79 (2), 205-218.

Goovaerts P., 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103 (1-2), 3-26.

Goovaerts P., 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103 (1-2), 3-26.

Goovaerts P., 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103 (1-2), 3-26.

Goovaerts P., 1997. Geostatistics for natural resources evaluation. Oxford University Press, New York.

Grace J.B., 1989. Effects of Water Depth on Typha latifolia and Typha domingensis. Am.J.Bot. 76 (5), 762-768.

Grace J.B., 1989. Effects of Water Depth on Typha latifolia and Typha domingensis. Am.J.Bot. 76 (5), 762-768.

Grayson R., Blöschl G., 2001. Spatial Modelling of Catchment Dynamics. In: Grayson R., Blöschl G. (Eds.), Spatial patterns in catchment hydrology : observations and modelling. Cambridge University Press, Cambridge, New York, pp. 51-81.

Haan C.T., 1989. Parametric uncertainty in hydrologic modeling. Trans. ASAE 32 (1), 137-146.

Haan C.T., Allred B., Storm D.E., Sabbagh G.J., Prabhu S., 1995. Statistical procedure for evaluating hydrologic/water quality models. Trans. of ASAE 38 (3), 725-733.

Haan C.T., Storm D.E., Al-Issa T., Prabhu S., Sabbagh G.J., Edwards D.R., 1998. Effect of parameter distributions on uncertainty analysis of hydrologic models. Trans. of ASAE 41 (1), 65-70.

Hall J.W., Tarantola S., Bates P.D., Horritt M.S., 2005. Distributed Sensitivity Analysis of Flood Inundation Model Calibration. J.Hydr.Engrg. 131 (2), 117-126.

I.M. S., A. S., 1995. About the use of rank transformation in sensitivity analysis of model output. Reliability Engineering and System Safety 50, 225-239(15).

Jaime Gómez-Hernández J., Mohan Srivastava R., 1990. ISIM3D: An ANSI-C three-dimensional multiple indicator conditional simulation program. Comput.Geosci. 16 (4), 395-440.

Kenward T., Lettenmaier D.P., Wood E.F., Fielding E., 2000. Effects of Digital Elevation Model Accuracy on Hydrologic Predictions. Remote Sens.Environ. 74 (3), 432-444.

Kyriakidis P.C., 2001. Geostatistical models of uncertainty for spatial data. In: Hunsaker C.T., Hunsaker C.T. (Eds.), Spatial uncertainty in ecology : implications for remote sensing and GIS applications. Springer, New York, .

Kyriakidis P.C., Dungan J.L., 2001. A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions. Environ.Ecol.Stat. 8 (4), 311-330.

Le Coz M., Delclaux F., Genthon P., Favreau G., 2009. Assessment of Digital Elevation Model (DEM) aggregation methods for hydrological modeling: Lake Chad basin, Africa. Comput.Geosci. 35 (8), 1661-1670.

Lilburne L., Tarantola S., 2009. Sensitivity analysis of spatial models. Int.J.Geogr.Inf.Sci. 23 (2), 151.

Luis S.J., McLaughlin D., 1992. A stochastic approach to model validation. Adv.Water Resour. 15 (1), 15-32.

Maidment D. (Eds.), 1992. Handbook of hydrology. .

McKay M.D., 1995. Evaluating prediction uncertainty. NUREG/CR-6311, LA-12915-MS.

Mckay M.D., Beckman R.J., Conover W.J., 2000. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 42 (1), 55-61.

Moore I.D., Grayson R.B., Ladson A.R., 1991. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrol.Process. 5 (1), 3-30.

Morgan, M.G., and M. Henrion, 1992. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, Cambridge (UK).

Morris M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161-174.

Neumann L.N., Western A.W., Argent R.M., 2010. The sensitivity of simulated flow and water quality response to spatial heterogeneity on a hillslope in the Tarrawarra catchment, Australia. Hydrol.Process. 24 (1), 76-86.

Newman S., Grace J.B., Koebel J.W., 1996. Effects of Nutrients and Hydroperiod on Typha, Cladium, and Eleocharis: Implications for Everglades Restoration. Ecol.Appl. 6 (3), 774-783.

Newman S., Schuette J., Grace J.B., Rutchey K., Fontaine T., Reddy K.R., 1998. Factors influencing cattail abundance in the northern Everglades. Aquat.Bot. 60 (3), 265-280.

Nowak M., Verly G., 2005. The Practice of Sequential Gaussian Simulation. Geostatistics Banff 2004 .

Pappenberger F., Beven K.J., Ratto M., Matgen P., 2008. Multi-method global sensitivity analysis of flood inundation models. Adv.Water Resour. 31 (1), 1-14.

Phillips D.L., Marks D.G., 1996. Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. Ecol.Model. 91 (1-3), 213-229.

Romanowicz E.A., Richardson C.J., 2008. Geologic Settings and Hydrology Gradients in the Everglades. Everglades Experiments .

Rossi R.E., Borth P.W., Jon J. Tollefson, 1993. Stochastic Simulation for Characterizing Ecological Spatial Patterns and Appraising Risk. Ecol.Appl. 3 (4), 719-735.

Rutchey K, Schall T.N., Doren R.F., Atkinson A., Ross M.S., Jones D.T., Madden M., Vilchek L., Bradley K.A., Snyder J.R., Burch J.N., Pernas T., Witcher B., Pyne M., White R., Smith T.J. III, Sadle J., Smith C.S., Patterson M.E., Gann G.D., 2006. Vegetation Classification for South Florida Natural Areas. USGS.

Rutchey K., Schall T., Sklar F., 2008. Development of Vegetation Maps for Assessing Everglades Restoration Progress. Wetlands 28 (3), 806-816.

Saltelli A., Ratto M., Andres T., Campolongo F., Cariboni J., Gatelli D., 2008. Global Sensitivity Analysis: The Primer. John Wiley & Sons Ltd, .

Saltelli A., 2004. Sensitivity analysis in practice : a guide to assessing scientific models. Wiley, Hoboken, NJ.

Saltelli A., 2004. Sensitivity analysis in practice : a guide to assessing scientific models. Wiley, Hoboken, NJ.

Saltelli A., Chan K., Scott E.M. (Eds.), 2000. Sensitivity Analysis: Gauging the Worth of Scientific Models. Wiley, Chichester.

Saltelli A., Tarantola S., Chan K.P.-., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41 (1), 39-56.

Saltelli A., Ratto M., Tarantola S., Campolongo F., 2005. Sensitivity Analysis for Chemical Models. Chem.Rev. 105 (7), 2811-2828.

SFWMD, 2005a. Regional Simulation Model (RSM). Theory Manual.

SFWMD, 2005b. Regional Simulation Model (RSM). Hydrologic Simulation Engine (HSE) User's Manual.

SFWMD, 2007. Natural Systems Regional Simulation Model v2.0 Results and Evaluation.

Sobol I.M., 1993. Sensitivity analysis for non-linear mathematical models. Math. Modell. Comput. Exp. 1, 407-414.

Sobol I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics 7, 86-112.

Tang Y., Reed P., van Werkhoven K., Wagener T., 2007. Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. Water Resour.Res. 43 (6), W06415.

Tang Y., Reed P., Wagener T., van Werkhoven K., 2007. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrology and Earth System Sciences 11 (2), 793-817.

Tarantola S., Gatelli D., Mara T.A., 2006. Random balance designs for the estimation of first order global sensitivity indices. Reliab.Eng.Syst.Saf. 91 (6), 717-727.

Urban N.H., Davis S.M., Aumen N.G., 1993. Fluctuations in sawgrass and cattail densities in Everglades Water Conservation Area 2A under varying nutrient, hydrologic and fire regimes. Aquat.Bot. 46 (3-4), 203-223.

USGS, 2003. Measuring and Mapping the Topography of the Florida Everglades for Ecosystem Restoration. USGS Fact Sheet 021-03 .

USGS, 1996. Vegetation Affects Water Movement in the Florida Everglades. FS-147-96.

Wagener T., McIntyre N., Lees M.J., Wheater H.S., Gupta H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrol.Process. 17 (2), 455-476.

Wallach D., Makowski D., Jones J.W., 2006. Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization and Application. Elsevier, Amsterdam, The Netherlands.

Wang M., Hjelmfelt A.T., Garbrecht J., 2000. DEM AGGREGATION FOR WATERSHED MODELING1. J.Am.Water Resour.Assoc. 36 (3), 579-584.

Wechsler S.P., 2007. Uncertainties associated with digital elevation models for hydrologic applications: a review. Hydrology and Earth System Sciences 11 (4), 1481-1500.

Widayati A., Lusiana B., Suyamto D., Verbist B.Uncertainty and effects of resolution of digital elevation model and its derived features: case study of Sumberjaya. Sumatera, Indonesia, Int.Arch.Photogrammetry Remote Sensing 35, 2004.

Wilson M.D., Atkinson P.M., 2003. Prediction uncertainty in elevation and its effect on flood inundation modelling. .

Wolock D.M., Price C.V., 1994. Effects of digital elevation model map scale and data resolution on a topography-based watershed model. Water Resour.Res. 30 (11), 3041-3052.

Wu Y., Rutchey K., Guan W., Vilchek L., Sklar F.H., 2002. Spatial simulations of tree islands for Everglades restoration. In: Sklar F.H., van der Valk A. (Eds.), Tree Islands of the Everglades. Kluwer Academic Publishers, Boston, MA, USA, pp. 469-498.

Yeo R.R., 1964. Life history of common cattail. Weeds 12 (4), 284-288.

Zanon S., Leuangthong O., 2005. Implementation Aspects of Sequential Simulation. Geostatistics Banff 2004 .

Zerger A., 2002. Examining GIS decision utility for natural hazard risk modelling. Environmental Modelling & Software 17 (3), 287-294.

Zhang J., Zhang J., Yao N., 2009. Geostatistics for spatial uncertainty characterization. Geo-Spatial Information Science 12 (1), 7-12.

Zhang W., Montgomery D.R., 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. Water Resour.Res. 30 (4), 1019-1028.

Zhu A.X., Scott Mackay D., 2001. Effects of spatial detail of soil information on watershed modeling. Journal of Hydrology 248 (1-4), 54-77.

BIOGRAPHICAL SKETCH

Zuzanna Zajac obtained her M.Sc. degree in Applied Ecology at University of Lodz, Poland. Since 2005 she worked as a Research Assistant at the Department of Agricultural and Biological Engineering at University of Florida. In 2010 she obtained a Ph.D. degree in Agricultural and Biological Engineering.