Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

TOWARDS AN ANALYTICAL FRAMEWORK FOR EXPERIMENTAL DESIGN IN EXPLORATORY MODELING

Enayat A. Moallemi
Sondoss Elsawah
Michael J. Ryan

Capability Systems Centre
School of Engineering and Information Technology
The University of New South Wales
Campbell ACT 2612, AUSTRALIA

ABSTRACT

Exploratory modeling—as an approach for modeling under uncertainty—is based on the analysis of computational experiments representing many possible model responses in the face of uncertainty. Experiments are generated based on various design factors, such as the way uncertainties are defined and the techniques by which value sets are sampled from these uncertainties. The choice of the design of experiments can impact on the computational cost of experiments as well as has an effect on the results and the conclusions drawn from those results. Despite this significance, experimental design has not been adequately discussed in the exploratory modeling literature. This article investigates which dimensions and what methods should be considered for an appropriate design of experiments in exploratory modeling. We conclude that there is a need to develop an analytical framework which can assist modelers to design experiments appropriately and to consider a wide range of model responses.

1 INTRODUCTION

Modeling, as a widely used approach for the formalization and development of a hypothesis and for problem solving and decision making (Axelrod 1997; Yücel 2010), is challenged by uncertainties in real-world applications (Bankes 1993). *Exploratory modeling* emerged as a group of computer-assisted approaches to address this challenge (Bankes 1993; Bankes et al. 2001; Lempert et al. 2003). Exploratory modeling argues that a conclusion based on one superior modeling technique, a single model structure, a set of input parameters, and a unique definition of desirability is not reliable under uncertainty conditions. Instead, it advocates building confidence in modeling results by considering a broad range of assumptions about the model and the values of its parameters, resulted from 'deep or severe uncertainty' (Lempert et al. 2003; Ben-Haim 2006), and by exploring the implications of these assumptions using computational experimentation (Bankes et al. 2001; Davis 2000). Exploratory modeling uses a model to generate an extensive database of computational experiments, in the format of randomly sampled set of values from input uncertainties and the corresponding model response. Exploratory modeling then applies a range of statistical, machine learning, and optimization techniques (Bryant and Lempert 2010; Hamarat et al. 2014; Watson and Kasprzyk 2017) to analyze the behavior of generated experiments. See Walker et al. (2013a) for a detailed explanation of the exploratory modeling process.

The primary focus of the exploratory modeling literature so far has been on the development of analytical approaches for robust decision making, adaptive planning, and scenario planning (Haasnoot et al. 2013; Kasprzyk et al. 2013; Guivarch et al. 2016; Kwakkel et al. 2016a; Trutnevyte et al. 2016; Guivarch et al. 2017; Moallemi et al. 2017; Walker et al. 2013a). While these approaches for the analysis of computational experiments have been developed extensively, the design of experiments, as the prerequisite

978-1-5386-6572-5/18/$31.00 ©2018 IEEE

step in all these approaches, has not been adequately discussed. Kwakkel and Pruyt (2015) have mentioned this gap and invited further research on efficient ways for finding the boundary of the uncertainty space and sampling techniques in exploratory modeling. Various outcomes of interest, different levels of uncertainty, a variety of sampling strategies, and different size of experiments can create a diversity of choices to make in experimental design. The implication of these choices can impact the inclusion or exclusion of some critical future possibilities, the computational burden of simulation runs, and the insights to be gained from the results of exploratory modeling. Some previous studies have shown the implications of different designs of exploratory modeling in practice. For example, Kwakkel et al. (2016b) showed how the size of experiments can influence the computational cost in the implementation of two exploratory modeling approaches. They concluded that the availability of computational resources can impact the selection of exploratory modeling methods. In another example, Haasnoot et al. (2014) discussed how to design a fit-for-purpose model for exploratory modeling, agile enough to perform many calculations quickly and realistically enough to represent sufficient details of the whole system.

Given the significance of experimental design in exploratory modeling and considering limited research on this area, this article investigates (1) which dimensions need to be considered in the design of experiments in exploratory modeling and (2) how techniques, mostly from sensitivity analysis and uncertainty estimation literature, can help to deal with these different dimensions. In particular, we discuss what approaches and techniques have been suggested for deciding about outcomes of interest, critical uncertainty factors, an appropriate space of uncertainty, an efficient sampling technique, and the sufficient size of experiments. This investigation and technical review lay the ground for the development of an analytical framework for experimental design in the future.

The article is structured in four sections. After introduction in Section 1, Section 2 reviews previous frameworks which were introduced within the exploratory modeling for experimental designs. Section 3 explains various dimensions of experimental designs, elaborating why they are important and how different techniques can deal with them. Finally, Section 4 concludes the article and identifies future research directions.

2 FRAMEWORKS FOR EXPERIMENTAL DESIGN IN PREVIOUS EXPLORATORY MODELING STUDIES

While a large group of previous exploratory modeling studies have included different dimensions of experimental design in their work, only a few of them have explained the approach they used to design experiments and systematically discussed the choices they made. Among them are the framework of experimental design (and sometimes for the broader domain of problem formulation) presented in Robust Decision Making (Lempert et al. 2003), Adaptive Policymaking (Kwakkel et al. 2010), Dynamics Adaptive Policy Pathways (Haasnoot et al. 2013), Multi-Objective Robust Decision Making (Kasprzyk et al. 2013), and Epoch-Era Analysis (Rader et al. 2010; Fitzgerald and Ross 2012).

In Robust Decision Making, Walker et al. (2013a) discussed some dimensions of experimental design in the first and second steps called *participatory scoping* and *case generation* respectively. The participatory scoping step specifies key uncertainties in interactions with stakeholders while the case generation step generates computational experiments using a simulation model and for the identified ranges of uncertainties. Lempert et al. (2003) also presented a framework for the generation of experiments called *XLRM framework*. The XLRM framework specifies four aspects in experimental design: future uncertainties, near-term policy levers, performance measures for decision making, and quantitative relationship(s) which link uncertainties to measures. In a similar approach to the XLRM framework, Walker et al. (2013b) introduced a framework consisting of four components for policy analysis under uncertainty. The framework has been adopted by previous exploratory modeling studies (Eker and van Daalen 2015) for problem formulation. The components include objectives and preferences of stakeholders (W), policy variables to play with the model (P), outcome indicators to assess the model performance (O), system model to generate outcomes from a set of inputs (R), and external uncertainties (X) which affect the model performance in long-term.

In Adaptive Policymaking, Kwakkel et al. (2010) discussed several aspects of experimental design in the first step, called *setting the stage,* where existing conditions, including constraint, policies, objectives, and definition of success, are specified. In a similar vein, Haasnoot et al. (2013) in Dynamics Adaptive Policy Pathways talked about the design of experiments in the first step (*describe the study area*) where system characteristics, objective, uncertainties, current situation and the definition of success are described, and in the second step (*problem analysis*) where possible future situations, consisting of transient scenarios for the identified ranges of uncertainties, are generated. In Multi-Objective Robust Decision Making, Kasprzyk et al. (2013) and Watson and Kasprzyk (2017) addressed some elements of experimental design in the first step (*problem formulation*) where epistemic uncertainties from various sources are taken into account, and in the third step (*uncertainty analysis*) where a set of outcomes of interest is generated based on the uncertainty ensemble. They also discussed three design considerations in the third step: how to sample from (exogenous) deeply uncertain factors, how many measures (outcomes) reflect robustness, and what statistical threshold define the robustness on the selected measures. Finally, in Epoch-Era Analysis, Rader et al. (2010) and Curry and Ross (2015) addressed some of the choices of experimental design in the first step (*multi-stakeholder value definition*) where the normative value of a system performance and the contextual uncertainties influencing the future performance is identified. They also discussed experimental design in the second step (*epoch enumeration*) where discrete ranges for uncertainties are specified in interaction with stakeholders, and the full factorial combination of values from uncertainty ranges create an ensemble of epochs, each characterizing a plausible future state of the system context.

The review of the exploratory modeling literature shows few previous studies discussed the design of experiments. Those few studies also mostly presented only an overview of the components that should be included in experimental design and did not introduced tools and techniques for choosing among various design alternatives. However, the techniques for designing computational experiments have been extensively discussed in the broader area of sensitivity analysis and uncertainty estimation (Shin et al. 2013; Pianosi and Wagener 2015; Pianosi et al. 2016) and in agent-monitored simulation works (Yilmaz et al. 2017; Yilmaz et al. 2016). This becomes the motivation for the rest of this paper to investigate what techniques exist and how experimental design in exploratory modeling can be informed by these existing techniques.

3 DIMENSIONS TO BE CONSIDERED IN THE DESIGN OF EXPERIMENTS

In this section we elaborate the dimensions of experimental design which elicit, analyze, and organize the inputs required for performing exploratory modeling. We should emphasize that the design of experiments is iterative; it is not set fully *a priori* and can be modified by feedback from other steps of the exploratory modeling process. We also highlight the role of stakeholders since experiments, such as selection scenarios, should be designed in a participatory process in interaction with stakeholders. In the following sub-sections (also in Table 1), we first explain why each dimension is important in the design of experiments. Second, we discuss how previous exploratory modeling studies (if exist) have dealt with that specific dimension of experimental design. Third, we introduce techniques and approaches from the exploratory modeling literature and also from sensitivity analysis and uncertainty estimation (Beven and Binley 2014; Pianosi et al. 2016) for choosing among various design alternatives in each dimension.

Table 1: An overview of potential techniques to be used in experimental design.

Aspect	Technique	
Outcomes of interest	Multiple outcomes	Multi-criteria analysis
		Pareto ranking
	Non-scalar outcomes	Objective function
		Prediction function
Critical uncertainty factors	Prioritizing uncertainties (Factor Periodization):	Correlation analysis methods
		Regression analysis methods
		Variance-based methods

	Identifying insignificant uncertainties (Factor Fixing)	Density-based methods
Appropriate space of uncertainty	Setting lower and upper bounds	Sensitivity analysis
	A priori filtering of the uncertainty space	Scenario discovery
	Estimating likelihood weights for the uncertainty space	Generalized Likelihood Uncertainty Estimation (GLUE)
Efficient sampling technique	Factorial sampling	
	Latin Hypercube sampling	
	Monte Carlo sampling	
Sufficient size of experiments	Rules of thumb	
	A posteriori evaluation of the size of experiments (Convergence analysis and Robustness analysis)	

3.1 Outcomes of Interest

Choosing outcomes of interest is a prerequisite, and therefore the first step, of the experiment design. The outcomes are used as inputs for other dimensions of the experiment design, e.g. for identifying critical uncertainty factors (see Section 3.2) and for limiting the uncertainty space (see Section 3.3). Outcomes of interest are also used in the analysis of experiments in exploratory modeling for various purposes, including: showing the state of the system performance (e.g. in boxplots and Kernel Density Estimates (Moallemi et al. 2017)), specifying a minimum performance threshold expected from a system (e.g. in scenario discovery (Bryant and Lempert 2010)), and characterizing decision objectives (e.g. in Multi-Objective Robust Decision Making (Kasprzyk et al. 2013) and in Moallemi et al. (2018a; 2018b)). Outcomes of interest are used in the exploratory modeling process as indicators for measuring the variation in the model response under uncertainty. There can be a single outcome or multiple outcomes of interest for representing different aspects of the model response. Outcomes can be in form of scalar values, such as the financial burden of a renewable-based electricity system for government after 20-year operation. They can also be in the form of non-scalar values and time series, such as the 20-year growth of renewable generated electricity. In dynamic models where differential equations are integrated over a temporal domain, the outcomes of interest can be the time-series values of a variable in the entire time period.

Some previous exploratory modeling studies with an interest in multi-objective optimization have used a collective measure of *robustness*, e.g. to what extent a function of multiple outcomes of interest remains stable over time, instead of individual outcome variables (Kasprzyk et al. 2013; Hamarat et al. 2014; Halim et al. 2015). An appropriate threshold is defined for this collective measure to assess robustness. The insensitivity of the model response, in terms of this measure and its threshold, to potential changes in the input parameters is used to quantify robustness (Maier et al. 2016). Herman et al. (2015) have extensively discussed the concept of robustness and how it should be defined. Some previous studies exist in exploratory modeling which defined robustness based on multiple outcome variables (Dixon et al. 2008; Popper et al. 2009). As an example, Kasprzyk et al. (2013) used multiple outcomes—market use, reliability, and cost—for a multi-objective robust optimization of a case study of urban water supply management. They assumed the robustness thresholds to be the extreme 90th or 10th percentile of the model response. They defined robustness based on the deviation of the baseline simulation from the extreme 90th or 10th percentile in multiple model outcomes over the uncertainty ensemble.

We did not identify a specific technique for choosing outcomes of interest in the literature as it very much depends on the purpose of analysis. Outcomes are usually selected among those model variables which can represent different aspects of the model response and which decision makers are interested to track after model execution. However, we identified techniques used for handling the multiplicity of outcomes and non-scalar outcomes in exploratory modeling:

- Regarding the multiplicity of outcomes, having multiple outcomes of interest can create ambiguity in identification and ranking of critical uncertainties in experimental design. Critical uncertainty factors are chosen based on the sensitivity of outcomes to them. Having multiple outcomes can result in variation in the sensitivity of the model, and therefore, in a different list of critical uncertainties (Pappenberger et al. 2008). In this case, some studies (Minunno et al. 2013) have addressed multiple outcomes jointly with multi-criteria analysis and by selecting critical uncertainties based on a Pareto set.
- Regarding non-scalar outcome values, they can be used as time series outcomes for assessing the model response over time. They can be also converted to a target function with scalar values (Pianosi et al. 2016). This target function can be in form of an objective function or a prediction function (Shin et al. 2013). The objective function is a measure of the model response resulted from the comparison of model generated values with their respective observed/documented values while the prediction function is a scalar value model response, chosen from a specific time step in the temporal domain or a function (such as average) of the entire temporal domain, meaningful for the analysis.

3.2 Critical Uncertainty Factors

Exploratory modeling tends to consider a wide of range uncertainties—such as techno-economic, social, and political uncertainties—in the surrounding environment of a system. However, computational time and cost limit the inclusion of every identifiable uncertainty in the simulation process and necessitate the use of a method for filtering only critical uncertainty factors. Critical uncertainty factors are considered those *exogenous* uncertain input parameters, imposed by the environment and out of the control of decision makers. Critical uncertainty factors should also be those exogenous uncertainties whose variation could significantly influence the model behavior. The selection of critical uncertainty factors impacts the space of uncertainty (possible variations of values) that we delineate later in the experimental design (see Section 3.3). The selection also impacts the reliability of conclusions obtained from the analysis of experiments as exclusion of one uncertainty can lead to disregarding an important area of future possibilities and can come at the cost of making vulnerable decisions.

Previous exploratory modeling studies have taken different approaches for specifying the list of critical uncertainties. Some studies took a qualitative approach and made assumptions about the criticality of uncertainties based on: a participatory approach involving stakeholders (Halim et al. 2015; Malekpour et al. 2017), a narrative-based approach being underpinned by theories from social sciences (Moallemi et al. 2017), and the modeler understanding of the model structure and the specific features of the context of the study (Hamarat et al. 2013; Kwakkel and Pruyt 2013; Hamarat et al. 2014; Eker and van Daalen 2015). Few studies took a quantitative approach and used a form of sensitivity analysis, such as correlation indices and Standardized Regression Coefficients, to identify and to rank critical uncertainties (Pye et al. 2015). Moallemi and Malekpour (2018) suggested a mixed participatory and computational approach (i.e. standard sensitivity analysis) for the identification of critical uncertainties.

The identification of critical uncertainty factors has been discussed more systematically in the sensitivity analysis literature. Sensitivity analysis identifies critical uncertainty factors to separate them from those factors that have no or less influence on outcome(s) of interest and should be ignore from the rest of analysis (van Werkhoven et al. 2009). Two groups of methods in sensitivity analysis can be adopted in exploratory modeling for selecting critical uncertainty factors: Factor Priorization and Factor Fixing (Saltelli et al. 2008). The latter identifies uncertainty factors with negligible influence on outcomes and the former prioritizes uncertainty factors based on their impact on outcomes. Several methods exist in each group:

- Correlation analysis (Iman and Helton 1988) and regression analysis (Saltelli and Marivoet 1990) are among those methods introduced for Factor Priorization. These methods prioritize uncertainties based on the statistical analysis of an input-output dataset generated by Monte Carlo simulations

(Pianosi et al. 2016). Several methods, such as Pearson Correlation Coefficient and Partial Correlation Coefficient for linear correlation among uncertainty factors and Canonical Correlation Analysis for non-linear correlation among uncertainty factors have been used in correlation analysis. Methods, such as Linear Regression, Standardized Regression Coefficients, and Classification And Regression Trees (CART), have been also used for regression analysis. See (Kleijnen and Helton 1999; Helton and Davis 2002) for detailed reviews of these methods. The choice of these methods depends on the degree of linearity and/or monotonicity being assumed between inputs and output (Pianosi et al. 2016).

- Variance-based (Borgonovo 2007) and density-based (Pappenberger et al. 2008; Anderson et al. 2014) methods are among those methods introduced for Factor Fixing. Variance-based methods treat uncertainty factors as stochastic variables which are used by the model and result in an outcome distribution (Pianosi et al. 2016). The variance of the observed distribution is used to measure the sensitivity to an uncertain factor. Variance-based methods can be *first-order* when the direct contribution of each individual uncertainty to the outcome is analyzed separately, or *total-order* (Homma and Saltelli 1996) when the separate contribution of each uncertainty factor and the contributions from their interactions are analyzed together. Density-based methods—such as entropy-based indices (Pappenberger et al. 2008), δ-sensitivity measure (Anderson et al. 2014), and PAWN (Pianosi and Wagener 2015)—also treat uncertainty factors as stochastic variables. However, they measure the sensitivity to a single uncertainty by comparing the variance (divergence) of the Probability Density Function of outcomes when all uncertainties can vary except one uncertainty of interest which is fixed to a single value (Pianosi et al. 2016).

Uncertainty can also exist in the model structure and model conceptualization. Haasnoot et al. (2014) argued that a fit-for-purpose model in exploratory modelling needs to be integrated in the sense that it integrates knowledge and methods from different disciplines to represent various aspects of the system. They argued that the model also needs to be agile in the sense that it can run many simulations under various assumptions fast and with low computational burden. One way of coping with uncertainty in the model development process in exploratory modelling is by informing it using narratives (storylines) which bring a qualitative understanding of the system based on the participation of stakeholders. Narratives can specify the scope and boundary of the system with its main components and interactions informed by stakeholders (Moallemi et al. 2017).

3.3 An Appropriate Space of Uncertainty

The combination of critical uncertainty factors with various ranges of variation forms a space of uncertainty from which inputs for running simulations in exploratory modeling are sampled. On the one hand, an uncertainty space that is too broad increases the computational burden and makes the conclusions from the analysis of experiments too plural. On the other hand, an uncertainty space that is too narrow increases the risk of missing some future possibilities and the risk of making conclusions vulnerable to some unforeseen (future) circumstances. Consequently, delineating an appropriate space of uncertainty is one of the most delicate parts of the design of experiments in exploratory modeling. The space of uncertainty can be well-characterized with known probability distributions, or can be, as it is called, a *deep uncertainty* (Lempert et al. 2003) or a *severe uncertainty* (Ben-Haim 2006). Deep (or severe) uncertainty is a term used for describing the Knightian form of uncertainty (Knight 1921) rather than probabilistic uncertainty. It is a condition with no agreement or knowledge on the value or probability distribution of uncertainty variables. The focus of exploratory modeling is on deep uncertainties (Bankes 1993).

An appropriate space of uncertainty is usually delineated in a way feasible with respect to the feature of the case study and the natural or physical constraints of uncertainties. It should be also in accordance with the common sense of experts and existing literature. Moreover, the delineation of the space of uncertainty is appropriate if exploratory modeling results are sensitive to it. Previous studies within the exploratory modeling literature and the broader area of sensitivity analysis and uncertainty estimation have used different approaches and methods for delineating the uncertainty space:

- A common approach is to consider lower and upper bounds and to assume independent uniform distributions for uncertainties (Pianosi et al. 2016). Different methods are used to set the bounds and to narrow the uncertainty space. For example, some studies (Spear et al. 1994) used sensitivity analysis to investigate how the model responds to the interactions of multiple uncertainties in different areas of the uncertainty space and to limit the uncertainty space accordingly. When uncertainties are in the form of a time-series, the process of sensitivity analysis would be different. For example, Singh et al. (2014) and Moallemi et al. (2017) specified the space of uncertainty related to the time series parameters through a scalar multiplier which can regenerate a similar trend in combination with a function of time and a fixed growth value. As an example from energy sectors, the electricity demand over time is a time series uncertainty, and its ranges of uncertainty is dynamic over time and cannot be set as fixed lower and upper bounds at the beginning of simulations. The electricity demand can be assumed as a non-linear function of scalar parameters: a deterministic initial demand, the uncertain effects of economic growth and electricity price on demand. They then defined the ranges of uncertainty for each scalar parameter.
- Other studies (Pappenberger et al. 2008; Kasprzyk et al. 2013; Kelleher et al. 2013) initially considered a feasible space of uncertainty, but limited this uncertainty space by applying *a priori* for filtering those areas which result in a certain behavior. This behavior is defined based on a threshold on outcomes of interest. An example is Kasprzyk et al. (2013) who generated an ensemble of the states of the world by Latin Hypercube sampling from upper and lower bounds of uncertainty factors. They also used a scaling factors method (Dixon et al. 2008) to renormalize the distribution of uncertainties within the specified bounds. However, not all generated states of the world were considered for analysis. They considered worst-case, i.e. the most extreme 10th percentile of states of the world to make sure that decisions made will remain robust even under extreme circumstances. Kasprzyk et al. (2013) used scenario discovery (Bryant and Lempert 2010) to identify states of the world (areas of the uncertainty space) which result in outcome values outside the 10th and 90th percentiles.
- Another approach from the uncertainty estimation literature which can be adopted in exploratory modeling for delineating the space of uncertainty is Generalized Likelihood Uncertainty Estimation (GLUE). GLUE was introduced by Beven and Binley (1992, 2014) for the calibration and estimation of uncertainty based on generalized likelihood measures. GLUE was introduced primarily for a situation where a set of observed datasets exist for the estimation of uncertainties. The core argument is that every combination of values from the space of uncertainty is capable of being considered equally likely as inputs for simulations. This approach argues that it is short-sighted to fit an optimal estimate for the baseline value of uncertainty factors and then to consider upper and lower bounds around the baseline to delineate the space of uncertainty. GLUE rejects this as it may exclude value sets in a distant area of the selected bounds which could still result in similar behavior. GLUE instead advocates assigning a likelihood weight to each value set based on comparing predicted (model-generated) values with some qualitative and quantitative evidences. Any value sets with a likelihood over zero will be taken into account as a potential value for uncertainties. The approach runs simulations based on value sets randomly chosen from the specified distribution of uncertainties.

3.4 An Efficient Sampling Technique and A Sufficient Size of Experiments

Sampling is a strategy for choosing random samples from the space of uncertainty for model execution to generate experiments, and the size of experiments is the number of simulation runs that are be performed to produce an ensemble of possibilities in exploratory modeling. Choosing the efficient sampling technique is important as an efficient technique can cover the space of uncertainty without leaving any areas (possibilities) with no samples. Choosing the sufficient size of experiments is important as a high number of experiments can increase the computational burden, and a low number of experiments does lead to an inadequate density of results, and therefore to less reliable conclusions.

Although the exploratory modeling literature has not sufficiently discussed the relationship between sampling techniques and experiment size, the issue is widely analyzed in the sensitivity analysis and uncertainty estimation literature and several techniques have been introduced:

- The efficiency of the sampling technique depends on how fully and evenly the searching algorithm of the technique covers the space of uncertainty. It also depends on the complexity of the likelihood surface of the uncertainty space (Beven and Binley 2014). This complexity comes from the interaction of multiple dimensions in the uncertainty space or from the model structure (Beven and Binley 2014). Adaptive sampling techniques have been introduced in the literature for an efficient search of the space of uncertainty based on their likelihood of uncertainty estimation (Spear et al. 1994; Khu and Werner 2003; Blasone et al. 2008a; Blasone et al. 2008b). These adaptive sampling techniques partition the space of uncertainty and consider the area of higher likelihood to improve the density of sampling. Latin Hypercube sampling has been suggested as another way of sampling when prior information about the distributions of the uncertainty space exist (Looms et al. 2008; Beven and Binley 2014). There are also other sampling techniques such as Full Factorial sampling, and Monte Carlo sampling, each with specific features, which can be used in different contexts. See (Press et al. 1996; Forrester and Keane 2008) for an introduction to these sampling techniques.
- To select the sufficient size of experiments, the ideal is to include the widest possible response of outcomes while keeping the size of experiments as small as possible. Choosing the size of experiments is dependent highly on the application, and a general rule may not remain valid from one example to another. However, there are some rules of thumb in the sensitivity analysis literature to choose *a priori* an appropriate the size of experiments. For example, Pianosi et al. (2016) suggested that the number of experiments for M input uncertainties increases by a factor between 100 and 1000 for regional sensitivity analysis and by a factor around 1000 and more for variance-based and density-based methods of global sensitivity analysis. There are also some techniques from the sensitivity analysis literature (Pianosi et al. 2016), which can be used to verify *a posteriori* the appropriateness of the size of experiments in exploratory modeling. Two such techniques are convergence analysis and robustness analysis. In convergence analysis (Nossent et al. 2011; Wang et al. 2013), the degree of independence between the results and the size of experiments is assessed by taking sub-samples from the original experiments and by analyzing whether the same results are achieved or not. In robustness analysis (Romano and Shaikh 2012), the degree of independence between the results and the set of experiments employed is assessed by taking different sets of experiments of the same size and analyzing the similarity of results.

4 CONCLUSIONS

This article assumed that the way we design experiments in exploratory modeling can have a significant impact on the computational burden of simulation runs and the reliability of the results and conclusions. Although the computing power constraint for large simulation runs has been relatively relaxed, the limit on computational power can still cause the issue of simulation speed when a specific model structure or simulation platform is slow to run and/or when a high dimension of uncertainty space exists. This signifies the importance of further investigating experimental design in exploratory modeling; a discussion of various techniques which can be used for answering different questions in the design of experiments.

From a review of the literature we observe that a variety of techniques from sensitivity analysis and uncertainty estimation can be good complements to the limited research on experimental design in exploratory modeling. Each of these techniques has different features and is more suitable for some specific conditions. The selection among these techniques for an experimental design depends greatly on the harmony between the feature of techniques and the characteristics of the application (context) as well as on the purpose of analysis in exploratory modeling. This offers opportunities for two future research directions.

The first is to move towards developing an analytical framework which can assist modelers, as a toolkit, in the design of efficient and appropriate experiments suiting their purpose of analysis in exploratory modeling the best. The second is to investigate further to what extent the choice of modelers in each dimension of experimental design can change the results of exploratory modeling in practice and how sensitive the final conclusions are to different choices of design.

REFERENCES

Anderson, B., E. Borgonovo, M. Galeotti, and R. Roson. 2014. "Uncertainty in Climate Change Modeling: Can Global Sensitivity Analysis Be of Help?". *Risk Analysis* 34(2):271–293.

Axelrod, R. 1997. "Advancing the Art of Simulation in the Social Sciences." in Simulating Social Phenomena". In *Simulating Social Phenomena*, edited by R. Conte, R. Hegselmann, and P. Terna, 21–40. Berlin: Springer-Verlag.

Bankes, S. 1993. "Exploratory Modeling for Policy Analysis". *Operations Research* 41(3):435–449.

Bankes, S. C., R. J. Lempert, and S. W. Popper. 2001. "Computer-Assisted Reasoning". *Computing in Science & Engineering* 3(2):71–77.

Ben-Haim, Y. 2006. *Info-Gap Decision Theory: Decisions under Severe Uncertainty*. London, UK: Academic Press.

Beven, K., and A. Binley. 1992. "The Future of Distributed Models: Model Calibration and Uncertainty Prediction". *Hydrological processes* 6(3):279–298.

Beven, K., and A. Binley. 2013. "Glue: 20 Years On". *Hydrological Processes* 28(24):5897–5918.

Blasone, R.-S., H. Madsen, and D. Rosbjerg. 2008a. "Uncertainty Assessment of Integrated Distributed Hydrological Models Using Glue with Markov Chain Monte Carlo Sampling". *Journal of Hydrology* 353(1):18–32.

Blasone, R.-S., J. A. Vrugt, H. Madsen, D. Rosbjerg, B. A. Robinson, and G. A. Zyvoloski. 2008b. "Generalized Likelihood Uncertainty Estimation (Glue) Using Adaptive Markov Chain Monte Carlo Sampling". *Advances in Water Resources* 31(4):630–648.

Borgonovo, E. 2007. "A New Uncertainty Importance Measure". *Reliability Engineering & System Safety* 92(6):771–784.

Bryant, B. P., and R. J. Lempert. 2010. "Thinking inside the Box: A Participatory, Computer-Assisted Approach to Scenario Discovery". *Technological Forecasting and Social Change* 77(1):34–49.

Curry, M. D., and A. M. Ross. 2015. "Considerations for an Extended Framework for Interactive Epoch-Era Analysis". *Procedia Computer Science* 44:454–465.

Davis, P.K., 2000. "Exploratory Analysis Enabled By Multiresolution, Multiperspective Modeling". In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines et al., 293–302. Piscataway, New Jersey: IEEE.

Dixon, L., R. J. Lempert, T. LaTourrette, and R. T. Reville. 2008. "The Federal Role in Terrorism Insurance: Evaluating Alternatives in an Uncertain World". RAND, Santa Monica, CA.

Eker, S., and E. van Daalen. 2015. "A Model-Based Analysis of Biomethane Production in the Netherlands and the Effectiveness of the Subsidization Policy under Uncertainty". *Energy Policy* 82:178–196.

Fitzgerald, M. E., and A. M. Ross. 2012. Mitigating Contextual Uncertainties with Valuable Changeability Analysis in the Multi-Epoch Domain. Systems Conference (SysCon), 2012 IEEE International.

Forrester, A., and A. Keane. 2008. *Engineering Design Via Surrogate Modelling: A Practical Guide*: John Wiley & Sons.

Guivarch, C., J. Rozenberg, and V. Schweizer. 2016. "The Diversity of Socio-Economic Pathways and Co2 Emissions Scenarios: Insights from the Investigation of a Scenarios Database". *Environmental Modelling & Software* 80:336–353.

Guivarch, C., R. Lempert, and E. Trutnevyte. 2017. "Scenario Techniques for Energy and Environmental Research: An Overview of Recent Developments to Broaden the Capacity to Deal with Complexity and Uncertainty". *Environmental Modelling & Software* 97:201–210.

Haasnoot, M., J. H. Kwakkel, W. E. Walker, and J. ter Maat. 2013. "Dynamic Adaptive Policy Pathways: A Method for Crafting Robust Decisions for a Deeply Uncertain World". *Global Environmental Change* 23(2):485–498.

Haasnoot, M., W. P. A. van Deursen, J. H. A. Guillaume, J. H. Kwakkel, E. van Beek, and H. Middelkoop. 2014. "Fit for Purpose? Building and Evaluating a Fast, Integrated Model for Exploring Water Policy Pathways". *Environmental Modelling & Software* 60(Supplement C):99–120.

Halim, R. A., J. H. Kwakkel, and L. A. Tavasszy. 2015. "A Scenario Discovery Study of the Impact of Uncertainties in the Global Container Transport System on European Ports". *Futures* 81:148–160.

Hamarat, C., J. H. Kwakkel, and E. Pruyt. 2013. "Adaptive Robust Design under Deep Uncertainty". *Technological Forecasting and Social Change* 80(3):408–418.

Hamarat, C., J. H. Kwakkel, E. Pruyt, and E. T. Loonen. 2014. "An Exploratory Approach for Adaptive Policymaking by Using Multi-Objective Robust Optimization". *Simulation Modelling Practice and Theory* 46:25–39.

Helton, J. C., and F. Davis. 2002. "Illustration of Sampling ‐ Based Methods for Uncertainty and Sensitivity Analysis". *Risk Analysis* 22(3):591–622.

Herman, J. D., P. M. Reed, H. B. Zeff, and G. W. Characklis. 2015. "How Should Robustness Be Defined for Water Systems Planning under Change?". *Journal of Water Resources Planning and Management* 141(10):04015012.

Homma, T., and A. Saltelli. 1996. "Importance Measures in Global Sensitivity Analysis of Nonlinear Models". *Reliability Engineering & System Safety* 52(1):1–17.

Iman, R. L., and J. C. Helton. 1988. "An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models". *Risk Analysis* 8(1):71–90.

Kasprzyk, J. R., S. Nataraj, P. M. Reed, and R. J. Lempert. 2013. "Many Objective Robust Decision Making for Complex Environmental Systems Undergoing Change". *Environmental Modelling & Software* 42:55–71.

Kelleher, C., T. Wagener, B. McGlynn, A. Ward, M. Gooseff, and R. Payn. 2013. "Identifiability of Transient Storage Model Parameters Along a Mountain Stream". *Water Resources Research* 49(9):5290–5306.

Khu, S.–T., and M. G. Werner. 2003. "Reduction of Monte-Carlo Simulation Runs for Uncertainty Estimation in Hydrological Modelling". *Hydrology and Earth System Sciences Discussions* 7(5):680–692.

Kleijnen, J. P. C., and J. C. Helton. 1999. "Statistical Analyses of Scatterplots to Identify Important Factors in Large-Scale Simulations, 1: Review and Comparison of Techniques". *Reliability Engineering & System Safety* 65(2):147–185.

Knight, F. H. 1921. Risk Uncertainty and Profit. Boston, MA: Houghton Mifflin Company.

Kwakkel, J. H., and E. Pruyt. 2013. "Exploratory Modeling and Analysis, an Approach for Model-Based Foresight under Deep Uncertainty". *Technological Forecasting and Social Change* 80(3):419–431.

Kwakkel, J. H., and E. Pruyt. 2015. "Using System Dynamics for Grand Challenges: The Esdma Approach". *Systems Research and Behavioral Science* 32(3):358–375.

Kwakkel, J. H., W. E. Walker, and V. A. W. J. Marchau. 2010. "Adaptive Airport Strategic Planning". *European Journal of Transport and Infrastructure Research* 10:249–273.

Kwakkel, J., W. Walker, and M. Haasnoot. 2016a. "Coping with the Wickedness of Public Policy Problems: Approaches for Decision Making under Deep Uncertainty". *J. Water Resource Planning Management* 10.1061/(ASCE)WR.1943–5452.

Kwakkel, J. H., M. Haasnoot, and W. E. Walker. 2016b. "Comparing Robust Decision-Making and Dynamic Adaptive Policy Pathways for Model-Based Decision Support under Deep Uncertainty". *Environmental Modelling & Software* 86:168–183.

Lempert, R. J., S. W. Popper, and S. C. Bankes. 2003. *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis*: Rand Corporation.

Looms, M. C., A. Binley, K. H. Jensen, L. Nielsen, and T. M. Hansen. 2008. "Identifying Unsaturated Hydraulic Parameters Using an Integrated Data Fusion Approach on Cross-Borehole Geophysical Data". *Vadose Zone Journal* 7(1):238–248.

Maier, H. R., J. H. A. Guillaume, H. van Delden, G. A. Riddell, M. Haasnoot, and J. H. Kwakkel. 2016. "An Uncertain Future, Deep Uncertainty, Scenarios, Robustness and Adaptation: How Do They Fit Together?". *Environmental Modelling & Software* 81:154–164.

Malekpour, S., R. R. Brown, F. J. de Haan, and T. H. F. Wong. 2017. "Preparing for Disruptions: A Diagnostic Strategic Planning Intervention for Sustainable Development". *Cities* 63:58–69.

Minunno, F., M. v. Oijen, D. R. Cameron, and J. S. Pereira. 2013. "Selecting Parameters for Bayesian Calibration of a Process-Based Model: A Methodology Based on Canonical Correlation Analysis". *SIAM/ASA Journal on Uncertainty Quantification* 1(1):370–385.

Moallemi, E. A., and S. Malekpour. 2018. "A Participatory Exploratory Modelling Approach for Long-Term Planning in Energy Transitions". *Energy Research & Social Science* 35:205–216.

Moallemi, E. A., F. de Haan, J. Kwakkel, and L. Aye. 2017. "Narrative-Informed Exploratory Analysis of Energy Transition Pathways: A Case Study of India's Electricity Sector". *Energy Policy* 110:271–287.

Moallemi, E. A., S. El Sawah, and M. Ryan. 2018a. "Model-Based Multi-Objective Decision Making Under Deep Uncertainty From A Multi-Method Design Lens". *Simulation Modelling Practice & Theory* 84: 232–250.

Moallemi, E.A., S. El Sawah, and M. Ryan. 2018b. "Towards An Integrated Approach for Multi-Objective Decision Making Under Deep Uncertainty", Systems Evaluation Test and Evaluation Conference 2018: Unlocking the Future Through Systems Engineering: SETE 2018. Sydney: Engineers Australia 119–133.

Nossent, J., P. Elsen, and W. Bauwens. 2011. "Sobol'sensitivity Analysis of a Complex Environmental Model". *Environmental Modelling & Software* 26(12):1515–1525.

Pappenberger, F., K. J. Beven, M. Ratto, and P. Matgen. 2008. "Multi-Method Global Sensitivity Analysis of Flood Inundation Models". *Advances in Water Resources* 31(1):1–14.

Pianosi, F., and T. Wagener. 2015. "A Simple and Efficient Method for Global Sensitivity Analysis Based on Cumulative Distribution Functions". *Environmental Modelling & Software* 67:1–11.

Pianosi, F., K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson, and T. Wagener. 2016. "Sensitivity Analysis of Environmental Models: A Systematic Review with Practical Workflow". *Environmental Modelling & Software* 79:214–232.

Popper, S., J. Griffin, C. Berrebi, T. Light, and E. Y. Min. 2009. "Natural Gas and Israel's Energy Future: A Strategic Analysis under Conditions of Deep Uncertainty". RAND, Santa Monica, CA, USA.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1996. *Numerical Recipes in C*. Vol. 2: Cambridge University Press, Cambridge.

Pye, S., N. Sabio, and N. Strachan. 2015. "An Integrated Systematic Analysis of Uncertainties in UK Energy Transition Pathways". *Energy Policy* 87:673–684.

Rader, A. A., A. M. Ross, and D. H. Rhodes. 2010. A Methodological Comparison of Monte Carlo Simulation and Epoch-Era Analysis for Tradespace Exploration in an Uncertain Environment. Systems Conference, 2010 4th Annual IEEE.

Romano, J. P., and A. M. Shaikh. 2012. "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap". *The Annals of Statistics* 40(6):2798–2822.

Saltelli, A., and J. Marivoet. 1990. "Non-Parametric Statistics in Sensitivity Analysis for Model Output: A Comparison of Selected Techniques". *Reliability Engineering & System Safety* 28(2):229–253.

Saltelli, A., M. Ratto, T. Andres, F. Campologno, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. 2008. *Global Sensitivity Analysis*. Chichester, England: John Wiley and Sons.

Shin, M.-J., J. H. A. Guillaume, B. F. W. Croke, and A. J. Jakeman. 2013. "Addressing Ten Questions About Conceptual Rainfall–Runoff Models with Global Sensitivity Analyses in R". *Journal of Hydrology* 503:135–152.

Singh, R., T. Wagener, R. Crane, M. Mann, and L. Ning. 2014. "A Vulnerability Driven Approach to Identify Adverse Climate and Land Use Change Combinations for Critical Hydrologic Indicator

Thresholds: Application to a Watershed in Pennsylvania, USA". *Water Resources Research* 50(4):3409–3427.

Spear, R. C., T. M. Grieb, and N. Shang. 1994. "Parameter Uncertainty and Interaction in Complex Environmental Models". *Water Resources Research* 30(11):3159–3169.

Trutnevyte, E., C. Guivarch, R. Lempert, and N. Strachan. 2016. "Reinvigorating the Scenario Technique to Expand Uncertainty Consideration". *Climatic Change* 135(3):373–379.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang. 2009. "Sensitivity-Guided Reduction of Parametric Dimensionality for Multi-Objective Calibration of Watershed Models". *Advances in Water Resources* 32(8):1154–1169.

Walker, W. E., M. Haasnoot, and J. H. Kwakkel. 2013a. "Adapt or Perish: A Review of Planning Approaches for Adaptation under Deep Uncertainty". *Sustainability* 5(3):955–979.

Walker, W. E., V. A. Marchau, and J. H. Kwakkel. 2013b. "Uncertainty in the Framework of Policy Analysis". In *Public Policy Analysis.*, edited by W. A. H. Thissen, and W. E. Walker, 215–261. Berlin: Springer.

Wang, J., X. Li, L. Lu, and F. Fang. 2013. "Parameter Sensitivity Analysis of Crop Growth Models Based on the Extended Fourier Amplitude Sensitivity Test Method". *Environmental modelling & software* 48:171–182.

Watson, A. A., and J. R. Kasprzyk. 2017. "Incorporating Deeply Uncertain Factors into the Many Objective Search Process". *Environmental Modelling & Software* 89:159–171.

Yilmaz, L., S. Chakladar, and K. Doud. 2016. "The Goal-Hypothesis-Experiment Framework: A Generative Cognitive Domain Architecture For Simulation Experiment Management". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder et al., 1001–1012. Piscataway, New Jersey: IEEE.

Yilmaz, L., S. Chakladar, K. Doud, A. E. Smith, A. Teran-Somohano, H. Oğuztüzün, S. Çam, O. Dayıbaş, and B. K. Görür. 2017. "Models As Self-Aware Cognitive Agents And Adaptive Mediators For Model-Driven Science". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan et al., 1300–1311. Piscataway, New Jersey: IEEE.

Yücel, G. 2010. "Analyzing Transition Dynamics: The Actor-Option Framework for Modelling Socio-Technical Systems". Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands.

AUTHOR BIOGRAPHIES

ENAYAT A. MOALLEMI is Research Associate at the Capability Systems Centre, the University of New South Wales. His research lies at the intersection of robust decision making, exploratory modelling, and sustainability transitions. E–mail: e.moallemi@unsw.edu.au.

SONDOSS ELSAWAH is Senior Lecturer at the University of New South Wales. Her research program focuses on understanding the behavior of large complex problems and systemic risks that arise from the interactions between social, ecological, and technological systems. Email: s.elsawah@adfa.edu.au.

MICHAEL J. RYAN is Associate Professor and the Director of the Capability Systems Centre, University of New South Wales, Canberra. He is the author or co-author of twelve books, three book chapters, and over 200 technical papers and reports. E-mail: M.Ryan@adfa.edu.au.