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

UNCERTAINTY QUANTIFICATION IN SIMULATION MODELS: A PROPOSED FRAMEWORK AND APPLICATION THROUGH CASE STUDY

Anna Paula Galvão Scheidegger
Amarnath Banerjee

Industrial and Systems Engineering Department
Texas A&M University
3131 TAMU
College Station, Texas 77843-3131, USA

Tábata Fernandes Pereira

Instituto de Engenharia de Produção e Gestão
Universidade Federal de Itajubá – campus Itabira
R. Irmã Ivone Drumond, 200, Distrito Industrial II
Itabira, Minas Gerais, 35903-087, BRAZIL

ABSTRACT

Despite the great advances in modeling and simulation techniques, modelers and researchers acknowledge that models are simplified representations of reality and, hence, are subject to uncertainty and errors. Although models are inevitably uncertain, they can still be a valuable decision-support tool if the users are informed about the uncertainty in the results. The importance of model uncertainty identification and quantification becomes clear in this context, but there are numerous challenges that remain. In this work, an uncertainty analysis framework is proposed for simulation models. This framework comprises of the steps that must be performed to analyze the uncertainty in simulation models. Next, an application of the framework is discussed where entropy is used as a possible measure of input-uncertainty. By using this framework, stakeholders can be better advised regarding the applicability and uncertainty of the simulation model, which will lead to an appropriate adjustment of expectations on the model results.

1 INTRODUCTION

Simulation models are developed to mimic real systems. Despite the increased level of details that can be added to modern day simulation models, modelers and researchers acknowledge that a model can never precisely reconstruct the real system under investigation. As an example, in the case of infectious disease outbreaks, due to the natural stochasticity of the system and the great impacts of rare or unforeseen events, epidemiological models are never capable of accurately predicting when and whether or not an individual will become infected (Christley et al. 2013).

Hanson and Hemez (2003) declared that simulation models are naturally dependent on the modeler's understanding of the system. This idea was reinforced by Christley et al. (2013) in their paper entitled, "'Wrong, but useful': Negotiating uncertainty in Infectious Disease Modelling". According to these authors, infectious disease models, as any other type of simulation model, have a good degree of assumptions, approximations, and human influence.

As highlighted by Oberkampf et al. (2002), a simulation model is always a simplified representation of the reality and any complex system, or even simple ones, contain details that are not represented in the model. Besides, if a system is driven by random inputs, a correct input model does not exist and uncertainty will always be present in the model (Biller and Gunes 2010). Consequently, simulation models are always subject to errors and uncertainty (Marelli and Sudret 2014).

DeVolder et al. (2002) argued that the more complex the system, the harder it is to get a precise solution from the model because the uncertainties are also greater. These authors also mentioned that this is somewhat ironic, because models, especially simulation models, are most needed to represent complex systems. In the context of infectious diseases, Christley et al. (2013) mentioned that the study of the emergence of new diseases or the re-emergence of old ones is the most suitable, but also the most problematic for modeling. The issue is that the system may not be well known, data may be limited, and there may be no precedents to compare the model results within the time and place of interest.

To emphasize the counter-intuitive impacts of model complexity on data uncertainty, the Council for Regulatory Environmental Modeling (CREM) argued that on one hand, increasingly complex models reduce model uncertainty as more understanding and details are incorporated into the model (Council for Regulatory Environmental Modeling 2009). On the other hand, more details also increase data uncertainty as more input variables and data are required. Therefore, a trade-off decision must be made between model complexity and uncertainty. According to the Council for Regulatory Environmental Modeling (2009), there is an appropriate level of complexity that will lead to the minimum total uncertainty.

The model inaccuracies due to uncertainty and errors may not only lead to economic losses, but they can also hinder society's trust in using models as a decision support tool (Kitching et al. 2005). The usefulness of simulation models depends, then, on controlling their error and uncertainty (Barton et al. 2014).

Although models are inevitably uncertain, the consensus in the academic field is that based on model uncertainty identification and quantification and given the appropriate reflection about this uncertainty, models can still effectively support decision-making. According to Oberkampf et al. (2002), a model with limited, but known applicability, is more useful than a very detailed or complex model with unknown uncertainty. The appropriate reflection on model uncertainty involves informing decision makers about how uncertain the model results may be, and where, when, and under which conditions the model results are applicable.

Considering the importance of uncertainty for simulation results, DeVolder et al. (2002) claimed that an algorithm or a systematic method was needed to quantify uncertainty in simulation models. These authors believed that to be an effective decision-support tool, simulation models must provide estimates about their level of accuracy and level of applicability so that decision-makers could determine what was the appropriate level of confidence to be placed on the results.

Nevertheless, estimating or quantifying uncertainty is not an easy task. Christley et al. (2013) reasoned that only a few uncertainties can be quantified, and even this quantification is most likely uncertain. So, the questions are: "What are the uncertainties present in a simulation model?", and "How to estimate these uncertainties?".

In order to provide directions for these questions, an uncertainty analysis framework is proposed in this work. This framework presents the types and sources of uncertainties that may exist in a simulation model and the steps that must be performed to estimate the simulation model uncertainty. An application of the framework is discussed where entropy is used as a measure of uncertainty.

The remaining of this paper is divided as follows: Section 2 provides a brief overview of uncertainty analysis, with Section 2.1 discussing the concepts of uncertainty quantification and uncertainty propagation, Section 2.2 discussing the types and sources of uncertainty in simulation models, and Section 2.3 presenting the steps for uncertainty analysis and the methods that have been applied in the literature. In Section 3, a framework for uncertainty analysis is proposed. An application of the framework is discussed in Section 4, and some concluding remarks with suggestions of future work are given in Section 5.

2 OVERVIEW OF UNCERTAINTY ANALYSIS

2.1 Uncertainty Quantification and Uncertainty Propagation

Uncertainty quantification is often classified into: forward uncertainty quantification and backward uncertainty quantification. Here, it is assumed that forward uncertainty quantification refers to uncertainty propagation, as indicated by Hanson and Hemez (2003).

Since the beginning of the 21st century, the topic of uncertainty quantification and propagation has been attracting the attention of simulation modelers from a large range of fields. This trend can be viewed in Figure 1, which shows the number of publications, by year, in the Scopus® database that had cited the term “uncertainty quantification” or “uncertainty propagation” in either their title, abstract, or keywords. The search was performed in the database on March 15, 2018, but limited up to 2017.

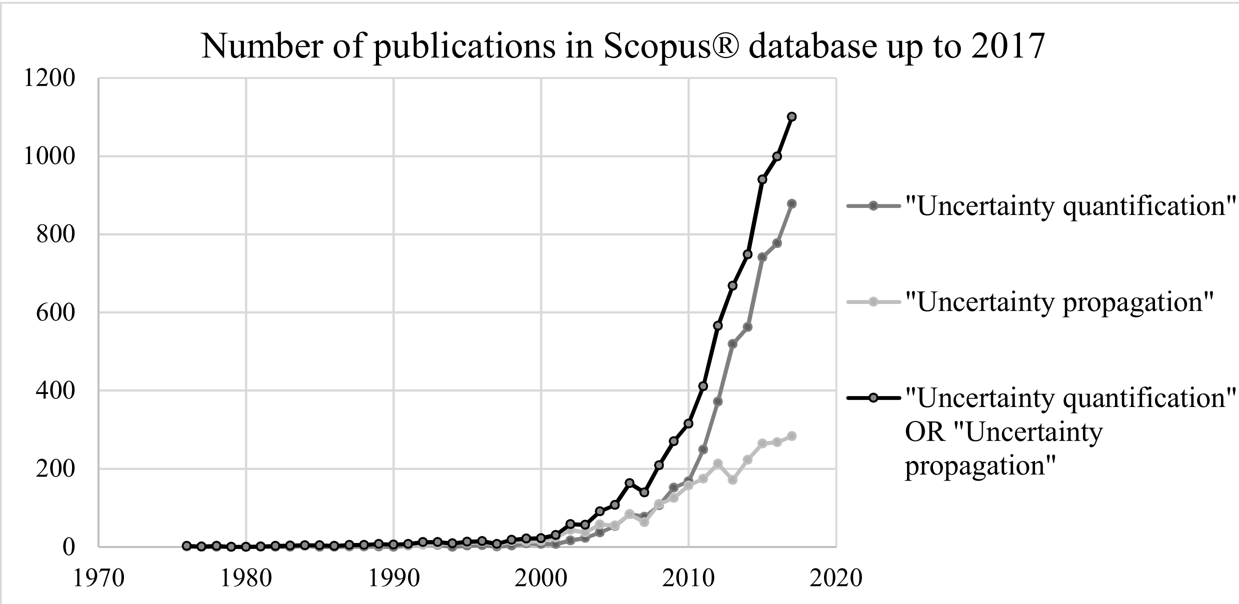


Figure 1: Number of publications, by year, in the Scopus® database citing the term “uncertainty quantification” or “uncertainty propagation” in their title, abstract, or keyword.

Due to the increased importance of the topic, researchers now talk about model verification, validation, and uncertainty quantification (VV&UQ), instead of the previous well-known model verification and validation (V&V) (Roy and Oberkampf 2011). Although the topic has gained attention in the past decade, it is still at a nascent stage and there is a lot to be researched (Chen et al. 2013).

The goal of uncertainty quantification is to identify each source of uncertainty and error and to estimate their effect on the simulation results (DeVolder et al. 2002). In other words, uncertainty quantification aims at making the best use of simulation models by rigorously estimating their variability, lack of knowledge, and possible errors (Marelli and Sudret 2014). Verification is the process of confirming the correctness of the model approximation and logic, and validation is the process of confirming the soundness of the model assumptions (Roy and Oberkampf 2011).

Some authors used to apply the terms “uncertainty quantification” and “uncertainty propagation” interchangeably. For instance, refer to Le Maıtre et al. (2004a) and Le Maıtre et al. (2004b). However, in this work, the terms have different meanings. Here, uncertainty propagation refers to the natural propagation of uncertainty, which can occur linearly or not, throughout the modeling steps, while uncertainty quantification refers to quantitative characterization or estimation of uncertainty due to the natural variability of the system, the existence of errors, and the uncertainty propagation. As it can be seen in Figure 1, the rate of use of the term “uncertainty propagation” is not increasing proportionally when compared to the rate of use of the term “uncertainty quantification”.

2.2 Types and Sources of Uncertainty

Most of the researchers classify uncertainty based on its nature. According to this classification, uncertainty is distinguished into two types: epistemic (also known as ignorance, subjective, or reducible uncertainty), and aleatory (also known as ontological, random, or irreducible uncertainty) (Oberkampf et al. 2002; Roy and Oberkampf 2011). Epistemic uncertainty occurs due to lack of knowledge and it refers to the uncertainty that could be possibly resolved or reduced by improved experiments, more data collection, and further research. Aleatory uncertainty refers to the uncertainty that is very difficult, if not impossible, to be captured due to the natural variability of the systems and the uncertainty of the occurrence of the event (Chen et al. 2013; Christley et al. 2013; Marelli and Sudret 2014).

Usually, epistemic uncertainty is characterized by a probability distribution that represents the modeler's degree of belief on that event or by an interval of possible real-world values with no associated distribution (Roy and Oberkampf 2011). For the aleatory case, uncertainty is characterized by a probability distribution that represents the frequency of occurrence of an event of interest.

The division between epistemic and aleatory uncertainty may not be always possible. Sometimes, depending on the context or the question being investigated, one may have to consider a mix of both uncertainties. Other researchers consider different uncertainty classifications. Sharda and Banerjee (2013), for instance, divided uncertainties into: (1) model uncertainties, the uncertainty related to the selection of the distribution model, and (2) parameter uncertainty, the uncertainty related to the selection of the distribution parameters. Christley et al. (2013) made the distinction between weak uncertainties (or probabilistic uncertainties or risks) that could be stated in probabilistic terms, and strong uncertainties (or scenario uncertainties or just uncertainty) that involved unanswered questions and for which probabilistic values could not be assigned.

According to Walker et al. (2003), model uncertainty is a three-dimensional concept encompassing the nature, the level, and the location of uncertainty. Based on the level of uncertainty, the authors mentioned that there is a transition from determinism to indeterminacy or total ignorance. The classification based on the uncertainty nature has been previously discussed in this section. Based on the location, Walker et al. (2003) classify uncertainty in: (1) context, (2) model, (3) input, (4) parameter, and (5) model outcome. The context refers to the identification of the system boundaries, the economic, environmental, political, social, and technological conditions, as well as the stakeholders' values and interests.

The model uncertainty refers to both conceptual and computer model and can be further classified as: (a) model structure uncertainty and (b) model technical uncertainty. Model structure uncertainty refers to the insufficient understanding of the system and involves the relationships among inputs and between inputs and outputs, the assumptions, and the model formulations. Technical uncertainty arises from software errors, hardware errors, and coding errors. The input uncertainty is associated with the external driving forces and to the system data that drives the model. Similarly, parameter uncertainty is associated with the uncertainty of the constants used in the model and this depends on the chosen context.

Finally, model outcome uncertainty, also known as the prediction error, is the total uncertainty reflected in the estimator of the response of interest due to the propagation of all other four types of uncertainties. Some authors also distinguished between uncertainties and errors. Both terms involve probabilistic concepts and affect the accuracy of the model results (DeVolder et al. 2002). Error is defined as an inaccuracy that can be identified upon examination and it is not due to lack of knowledge or inherent variability (Oberkampf et al. 2002). When an error is identified, a cost-effective decision must be made in terms of fixing it or not.

Xie et al. (2014a) stated that for any stochastic system, there are at least two sources of uncertainty in the simulation results: the input-uncertainty, due to fitting input distributions based on finite samples of real-world data; and the simulation-estimation error (or the simulation-sampling error), due to a finite amount of simulation effort. In fact, more than often, model errors and uncertainty arise from a number of sources and not only from the two aforementioned sources (Roy and Oberkampf 2011). According to Roy and Oberkampf (2011), the general rule is to consider an aspect uncertain unless there is strong evidence that this aspect results in a minimal impact on the model results. The sources of errors usually include: (1) mathematical approximations, model assumptions that do not represent the reality but make the solution feasible or cost-effective, and round-off errors; and (2) model logic or programming mistakes, typos, and measurement errors. The first group of errors is known as acknowledged errors and the second group is known as unacknowledged errors (Roy and Oberkampf 2011).

The sources of uncertainty include: the natural variability of the system, the initial conditions, the unknown variables, the incomplete knowledge of the environment and the surroundings (external forces), the system boundaries, the existence of unforeseen events (disasters, accidents, abnormal conditions, and hostile environment), the input-uncertainty (input model and parameter uncertainty), the model assumptions, the modeling tool, the modeler's understanding, the estimates used to summarize the results, the limited experiments, and the function of how the uncertainty propagates over the system (DeVolder et al. 2002; Hanson and Hemez 2003; Barton et al. 2010; Roy and Oberkampf 2011; Chen et al. 2013).

2.3 Uncertainty Analysis Steps and Methods

Uncertainty analysis methodology, or uncertainty management methodology as referred by Baudin et al. (2015), involves the following steps: (1) specification of the random inputs, the deterministic inputs, the model, the variable of interest, and the output estimator; (2) quantification of the sources of uncertainty using statistical fitting or expert judgment; (2b) quantification of the sources of uncertainty by indirect methods using real observations of the model outputs; (3) propagation of uncertainties to estimate the quantity of interest; and, (3b) execution of sensitivity analysis on the variable of interest to rank the uncertainties.

Based on Roy and Oberkampf (2011), uncertainty analysis methodology consists of six different steps: (1) identification of the sources of uncertainty (input modeling); (2) characterization of the model input uncertainties; (3) elimination of coding errors and estimation of uncertainty due to numerical errors; (4) propagation of uncertainties through the model to obtain uncertainties in the responses of interest; (5) quantification of model form uncertainty by validation process and estimation of model form uncertainty due to extrapolation to other conditions with no data available; and (6) estimation of total uncertainty. At each of the steps, a different method or a combination of methods can be used.

According to Marelli and Sudret (2014), every uncertainty analysis problem can be decomposed into input, model, and output analysis. The input and output uncertainties have been increasingly investigated in the simulation field and are known as extrinsic input-uncertainty and intrinsic output-uncertainty, respectively. For model uncertainty analysis, not many methods were found in the literature apart from model verification and validation.

Model form uncertainty is usually difficult to quantify but can be minimized through the validation process. The validation must be performed in the conditions for which the model was developed, but it must also be extrapolated to other possible conditions. Coding errors and numerical approximation errors are also difficult to eliminate, but they can be detected and minimized through the application of good modeling techniques and through the verification process (Roy and Oberkampf 2011).

In simulation models, the intrinsic output-uncertainty comes from the finite run length and the finite number of replications (Barton et al. 2010). According to Song and Nelson (2013) and Barton et al. (2014), the intrinsic uncertainty is already measured by all simulation software and it is characterized by confidence intervals on the performance measures. As in any sampling experiment, increasing the number of replications in a simulation project reduces the variance (Nelson 1987a). However, increasing the number of replications may be too costly or not feasible due to time constraints.

Some techniques such as antithetic variates (AV), control variates (CV), and common random numbers (CRN), have been developed to reduce the variance of simulation estimators without increasing the computational effort (Nelson 1987b). According to Nelson (1987a), VRTs had their origins in Monte Carlo estimation and survey sampling around 1965 and 1975, respectively. Many simulation software offer built-in features that facilitate the execution of AV and CRN, but CV usually requires some additional software support (Nelson 1987b).

According to Song et al. (2014), input uncertainty depends mainly on two factors: the amount of real-data available from which the input distribution parameters are estimated and the sensitivity of the response to those parameters. In other words, the input uncertainty depends on (1) how accurately the input was modeled, and (2) how sensitive is the system response to the input model. The most common methods used in input-uncertainty propagation and input-uncertainty quantification are: calibration, graphical techniques, sampling-based methods, sensitivity analysis, Bayesian methods, approximation methods, and meta-models or surrogate models (Barton 2012; Baudin et al. 2015).

Oberkampf et al. (2002) listed methods used for propagation of epistemic uncertainty and others used for propagation of aleatory uncertainty. Besides probability theory, the methods for propagation of epistemic uncertainty include: possibility theory, interval analysis, Dempster-Shafer theory or evidence theory, fuzzy set theory, imprecise probability theory, ensemble copula coupling, Bayesian estimation, and Bayesian model average (DeVolder et al. 2002; Chen et al. 2013; Schefzik et al. 2013). The methods used for propagation of aleatory uncertainty include: sampling methods such as Monte Carlo, Latin Hypercube, Hamiltonian Monte Carlo (HMC) algorithm, Particle Swarm Optimization (PSO) algorithm, and the Neighborhood Algorithm (NA), and statistical design approaches (Hanson and Hemez 2003; Mohamed et al. 2010).

Graphical techniques such as scatter plots and cobweb plots are used for qualitative uncertainty propagation through the identification of patterns. Correlation coefficients such as Pearson correlation coefficient and Spearman's rank correlation coefficient can also be used for uncertainty propagation (Baudin et al. 2015). Sampling methods, such as minimum energy design, stratified sampling, direct resampling, bootstrap resampling, and meta-model assisted bootstrap are other common methods for uncertainty propagation (Barton 2012). For a review of some of these methods, see Barton (2012).

Meta-model or surrogate models, such as polynomial chaos expansion (also known as Wiener chaos expansion) and kriging (also known as Gaussian process regression), are commonly used for uncertainty quantification. Approximation methods such as first-order reliability method (FORM), second-order reliability method (SORM) and δ-methods can also be applied (Marelli and Sudret 2014). Other methods of uncertainty quantification include variance-based sensitivity indices (also known as Sobol' indices).

Figure 2 provides an overview of uncertainty analysis. As shown in Figure 2, it is possible to verify some of the main sources of uncertainty, as well as the main methods being used in uncertainty propagation and quantification.

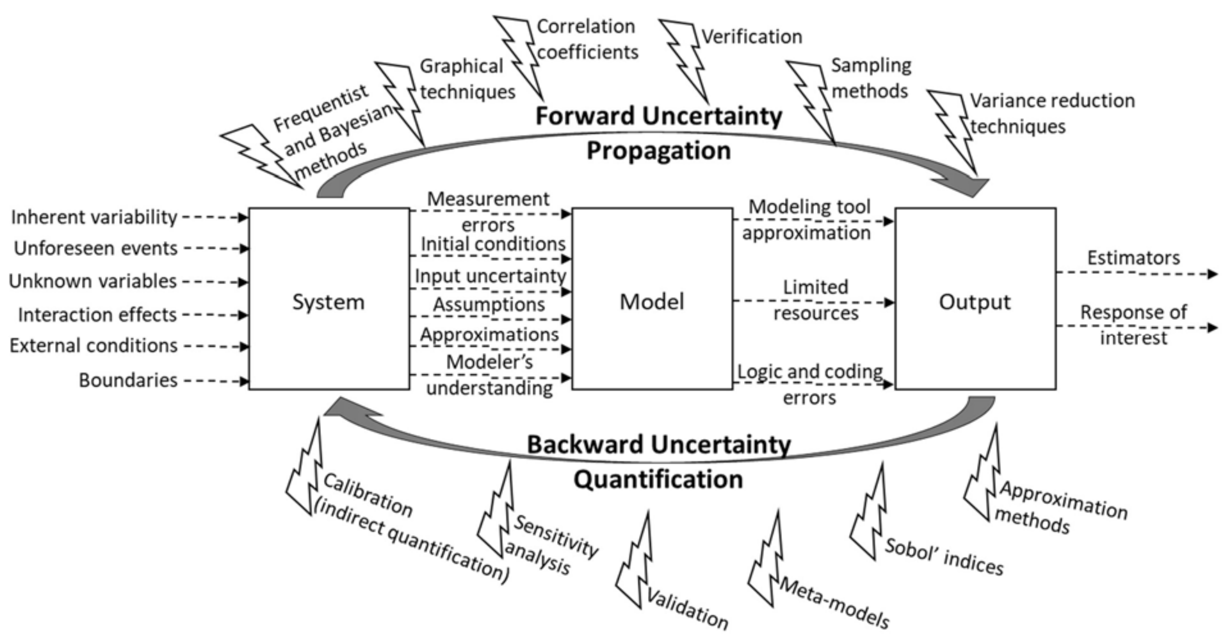


Figure 2: Uncertainty analysis overview.

Table 1 lists some of the important works in the field of uncertainty analysis and the methods used by them.

Table 1: Some of the important works in the field of simulation uncertainty analysis.

Paper	Method	Assumptions
Barton and Schruben (1993)	Uniform and bootstrap resampling	Independent univariate empirical distributions.
Barton et al. (2010)	Meta-model-assisted bootstrapping (stochastic kriging meta-model), direct bootstrap, Bayesian bootstrap, conditional confidence interval, and Bayesian credible interval.	Independent univariate parametric distributions with known families (exponential) and unknown parameters.
Song and Nelson (2013)	Bootstrap and variance decomposition	Independent univariate parametric distributions with known parametric families and unknown parameters.
Barton et al. (2014)	Meta-model-assisted bootstrapping (stochastic kriging meta-model)	Independent univariate parametric distributions with known parametric families and unknown parameters, and meta-model uncertainty can be ignored.
Xie et al. (2014a)	Bayesian credible interval assisted by meta-model (Gaussian process)	Independent univariate parametric distributions with known families and unknown parameters.
Xie et al. (2014b)	Meta-model-assisted bootstrapping (stochastic kriging meta-model) and Spearman's rank correlation	NORmal To Anything (NORTA) distribution [multivariate parametric distribution], which means unknown dependent inputs (unknown distributions).

3 UNCERTAINTY ANALYSIS FRAMEWORK

We propose a framework for simulation uncertainty analysis that consists of seven steps and a three-dimensional table for identification of the sources of uncertainty, as depicted in Figure 3. The framework was developed based on the discussion provided in Section 2.

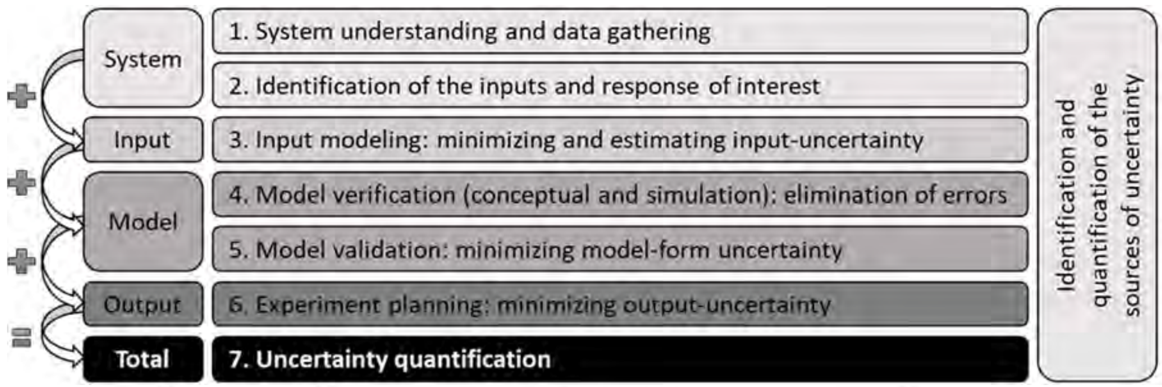
The three-dimensional table is based on the work of Walker et al. (2003) and contains the types of uncertainty that one may face in a simulation project. We used the classification by location, level, and nature found in the literature, but we adapted the classification by location for the terminology that is more frequently used in the simulation field. The classification by location is important because it is related to the phase of the simulation project, and the classification by level is also important, because, for those uncertainties where estimation is not possible, the level may be used for a thorough risk assessment. Examples include the use of failure mode effects analysis (FMEA) or event trees. The classification by nature may not be required if the modelers think it may not aid in the uncertainty quantification. All the uncertainties identified in each of the framework steps must be registered in the appropriate row of the three-dimensional table, following the appropriate classification.

The first step and second step of the framework involve understanding the system, identifying the inputs, and collecting all the data needed. These steps are related to the first phase of a simulation project. At this time, the modelers must define the goal of the project and perform a risk assessment of the worst, the most common, and the best conditions of the system. The modelers must also think about possible unforeseen events, estimate the inherent variability of the system, and elaborate a data collection protocol to minimize errors during data collection.

The third step involves modeling the inputs and using methods of input-uncertainty propagation, such as the ones discussed in Section 2.3, to minimize input-uncertainty and to estimate the uncertainty that may not be possible to eliminate. This step is related to the conceptual modeling phase. The focus of this paper is on this step, which is further described in Section 4.

The fourth and fifth steps involve model verification and validation, which are activities well-known to simulation modelers. It is important to highlight that the conceptual model must also be verified and this verification must occur before the beginning of the simulation model development.

The sixth step involves planning the experiments in a way that the output-uncertainty (also known as simulation-estimation uncertainty) is minimized. Finally, step 7 refers to the total uncertainty quantification or estimation, which must consider all the uncertainties identified and estimated in the previous steps. Step 6 and 7 correspond to the last phase of a simulation project.



Location of uncertainty	**Level of uncertainty**				**Nature of uncertainty**
	Statistical uncertainty	Scenario uncertainty	Recognized ignorance	Total ignorance	
System					Epistemic
					Aleatory
Input					Epistemic
					Aleatory
Conceptual model					Epistemic
					Aleatory
Simulation model					Epistemic
					Aleatory
Output					Epistemic
					Aleatory
Total estimation =					

Figure 3: A framework for simulation uncertainty analysis.

4 AN APPLICATION DISCUSSION

Since the focus of this paper is on Step 3 of the framework, the use of entropy is discussed as a measure of input uncertainty in simulation models. For the demonstration, we use a SEIR-SEI (susceptible-exposed-infectious-recovered / susceptible-exposed-infectious) compartmental model that studies the spread of Chikungunya. The SEIR model is used to represent humans, while the SEI model is used to represent the mosquito vector. The model was developed using System Dynamics in AnyLogic®. For a detailed description of the model and the parameters, please refer to Scheidegger and Banerjee (2017).

There are 10 input parameters used in the model. The parameter values were identified based on the literature. After analysis, it was identified that there was no consensus or there was no information available on the values of 5 parameters and, therefore, these are the parameters that will be used here to estimate uncertainty. These parameters are shown in Table 2.

Table 2: Input parameters and levels for experiments.

#	Input parameter	Level 1	Level 2	Level 3
1	Mosquito population [Tm, number of mosquitoes]	160,000	190,000	220,000
2	Daily mosquito to human infect rate [β_h]	0.126	0.140	0.154
3	Daily human to mosquito infect rate [β_m]	0.500	0.540	0.580
4	Daily human recovery rate [γ, inverse of the recovery period in days]	Uniform(0.140,1.000)	Uniform(0.250,0.500)	
5	Initial number of infectious mosquitoes [m_{i0}]	1	100	1,000

Two system responses were considered in this study: (1) total number of infected people, and (2) duration of the epidemic in days. The model was run for a duration of 5 years and 20 replications were performed for each combination of factors and levels, leading to a total number of runs (N) = 3,240. Next, we used Shannon's entropy, a measurement of information gain and uncertainty, to estimate the uncertainty impact of each input described in Table 2 on the response of interest. Before estimating the uncertainty, we had to group the responses in different clusters. The clustering was performed based on data split. For the "total number of infections" response, the data was clustered with 90% of the total population as the boundary to split the data into two subsets. For the "epidemic duration" response, the data was clustered using 2 years (730 days) as the boundary to split the data into two subsets.

According to the information entropy theory, we can use the information gain to quantify the reduction in uncertainty in the model due to one specific parameter. So, the higher the information gain of a response for input parameter x, the higher the reduction in uncertainty, and consequently, higher is the impact of parameter x uncertainty in the response. Let $I(Y,x) = H_t(Y) - \sum_{j=1}^{l_x} q_{x\,j}\, H(Y_x)_j$ be the information gain of response vector Y on factor x and let $H_t(Y) = -\sum_i \; p_i\, log_2 p_i$ be the total Shannon's entropy of Y and let $H(Y_x)_j$ be the Shannon's entropy of the cluster response subset of factor x on level j, which is also referred to as conditional Shannon's entropy. Here: p_i is the proportion of response elements in cluster i, l_x is the number of levels of factor x, and $q_{x\,j}$ is the proportion of response elements in level j of factor x. Table 3 shows the information gain of each parameter corresponding to the total number of infections and the duration of the outbreak.

Table 3: Information gain for cluster 1 of the total number of infections and duration of the outbreak.

Response		Total number of infections			Duration of the outbreak		
Cluster method		90% of the population			2 years (730 days)		
Parameter	Levels	# of <= 706,500	# of >= 706,500	Entropy	# of <= 730	# of >= 730	Entropy
Total		120	42	0.8256	120	42	0.8256
1 - Mosquito population	1	48	6	0.5033	31	23	0.9841
	2	39	15	0.8524	42	12	0.7642
	3	33	21	0.9641	47	7	0.5564
Information Gain			**0.0524**			**0.0574**	
2 - Daily mosquito to human infect rate	1	50	22	0.8880	58	14	0.7107
	2	1	0	0.0000	1	0	0.0000
	3	69	20	0.7687	61	28	0.8984

Information Gain			**0.0087**			**0.0162**	
3 - Daily human	1	56	22	0.8582	59	19	0.8010
to mosquito	2	1	0	0.0000	1	0	0.0000
infect rate	3	63	20	0.7966	60	23	0.8515
Information Gain			**0.0042**			**0.0037**	
4 - Daily human	1	116	36	0.7897	110	42	0.8504
recovery rate	2	4	6	0.9710	10	0	0.0000
Information Gain			**0.0247**			**0.0277**	
5 - Initial	1	75	28	0.8441	79	24	0.7832
number of	2	22	8	0.8366	15	15	1.0000
infectious mosquitoes	3	23	6	0.7355	26	3	0.4798
Information Gain			**0.0023**			**0.0566**	

From Table 3, we can conclude that the mosquito population is the parameter whose variability has the greatest impact on the response *total number of infections*, followed by the daily human recovery rate. For the response *duration of an outbreak*, the mosquito population, followed by the initial number of infectious mosquitoes are the parameters whose variability has the greatest impact.

As it can be seen, the response of interest is important, because the method does not agree with the ranking of impact for all parameters. Although for both responses, mosquito population was considered the parameter whose variability had the greatest impact, all the other parameters had different ranks in each response. To exemplify, the initial number of infectious mosquitoes was the second parameter based on the duration of the outbreak but it was considered the fifth (i.e. the last) based on the total number of infections.

5 CONCLUSIONS AND FUTURE WORK

Information entropy measures have been applied to a large number of fields. Although information entropy is a measure of uncertainty, to the best of our knowledge there has been no work applying the measures in the context of simulation uncertainty quantification.

We recognize that this is a simple application, but we believe this is a good starting point to show the potential of entropy measures as a method for simulation uncertainty quantification. We believe entropy measures must be further investigated as a method to inform simulation stakeholders about the importance of each input to the simulation response uncertainty. If applicable, this method can lead to better-planned data collection and to better use of simulation results, once stakeholders are informed about the potential uncertainties in the model responses. We realize that the topic of uncertainty quantification has been increasingly attracting the attention of researchers and practitioners in the field of simulation, but there are numerous challenges that still remain. In this paper, we proposed a framework for uncertainty analysis focused on the different phases of a simulation project, and which uses terminology that is common in the simulation field.

As topics for future investigation, we suggest: (1) investigating the consistency of a model for different contexts, different values of the parameters, and different clustering methods, (2) running the model for a larger number of replications to check for consistency in the results, (3) investigating if other entropy measures, such as Kolmogorov-Sinai entropy, may be more applicable to the problem, and (4) investigating how the results are affected by the input model used.

ACKNOWLEDGMENTS

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support to this work through the Science without Borders program and also FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais).

REFERENCES

Barton, R. R. 2012. "Tutorial: Input Uncertainty in Outout Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque et al., 67-78. Piscataway, New Jersey: IEEE.

Barton, R. R., and L. W. Schruben. 1993. "Uniform and Bootstrap Resampling of Empirical Distributions". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans et al., 503-508. Piscataway, New Jersey: IEEE.

Barton, R. R., B. L. Nelson, and W. Xie. 2010. "A Framework for Input Uncertainty Analysis". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson et al., 1189-1198. Piscataway, New Jersey: IEEE.

Barton, R. R., B. L. Nelson, and W. Xie. 2014. "Quantifying Input Uncertainty Via Simulation Confidence Intervals". *INFORMS Journal on Computing* 26(1):74-87.

Baudin, M., A. Dutfoy, B. Iooss, and A.-L. Popelin. 2015. "Openturns: An Industrial Software for Uncertainty Quantification in Simulation". In *Handbook of Uncertainty Quantification*, edited by R. Ghanem et al., 1-38. Cham, Switzerland: Springer.

Biller, B., and C. Gunes. 2010. "Introduction to Simulation Input Modeling". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson et al., 49-58. Piscataway, New Jersey: IEEE.

Chen, P., A. Quarteroni, and G. Rozza. 2013. "Simulation ‐ Based Uncertainty Quantification of Human Arterial Network Hemodynamics". *International Journal for Numerical Methods in Biomedical Engineering* 29(6):698-721.

Christley, R. M., M. Mort, B. Wynne, J. M. Wastling, A. L. Heathwaite, R. Pickup, Z. Austin, and S. M. Latham. 2013. ""Wrong, but Useful": Negotiating Uncertainty in Infectious Disease Modelling". *PLoS ONE* 8(10):e76277.

Council for Regulatory Environmental Modeling. 2009. "Web-Based Training on Best Modeling Practices and Technical Modeling Issues: Sensitivity and Uncertainty Analyses." Retrieved March 10th, 2018, from https://www.epa.gov/sites/production/files/2015-09/documents/mod8-saua-mod-final.pdf.

Devolder, B., J. Glimm, J. Grove, Y. Kang, Y. Lee, K. Pao, D. Sharp, and K. Ye. 2002. "Uncertainty Quantification for Multiscale Simulations". *Journal of Fluids Engineering* 124(1):29-41.

Hanson, K., and F. Hemez. 2003. "Uncertainty Quantification of Simulation Codes Based on Experimental Data". In *Proceedings of the 41st Aerospace Sciences*, edited, 1-10. Reston, VA: American Institute of Aeronautics and Astronautics.

Kitching, R. P., A. M. Hutber, and M. V. Thrusfield. 2005. "A Review of Foot-and-Mouth Disease with Special Consideration for the Clinical and Epidemiological Factors Relevant to Predictive Modelling of the Disease". *The Veterinary Journal* 169(2):197-209.

Le Maıtre, O. P., O. M. Knio, H. N. Najm, and R. G. Ghanem. 2004a. "Uncertainty Propagation Using Wiener–Haar Expansions". *Journal of Computational Physics* 197(1):28-57.

Le Maıtre, O. P., H. N. Najm, R. G. Ghanem, and O. M. Knio. 2004b. "Multi-Resolution Analysis of Wiener-Type Uncertainty Propagation Schemes". *Journal of Computational Physics* 197(2):502-531.

Marelli, S., and B. Sudret. 2014. "Uqlab: A Framework for Uncertainty Quantification in Matlab". *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*: 2554-2563.

Mohamed, L., M. A. Christie, and V. Demyanov. 2010. "Comparison of Stochastic Sampling Algorithms for Uncertainty Quantification". *SPE Journal* 15(01):31-38.

Nelson, B. L. 1987a. "A Perspective on Variance Reduction in Dynamic Simulation Experiments". *Communications in Statistics - Simulation and Computation* 16(2):385-426.

Nelson, B. L. 1987b. "Variance Reduction for Simulation Practitioners". In *Proceedings of the 1987 Winter Simulation conference*, edited by A. Thesen et al., 43-51. Piscataway, New Jersey: IEEE.

Oberkampf, W. L., S. M. Deland, B. M. Rutherford, K. V. Diegert, and K. F. Alvin. 2002. "Error and Uncertainty in Modeling and Simulation". *Reliability Engineering & System Safety* 75(3):333-357.

Roy, C. J., and W. L. Oberkampf. 2011. "A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing". *Computer Methods in Applied Mechanics and Engineering* 200(25-28):2131-2144.

Schefzik, R., T. L. Thorarinsdottir, and T. Gneiting. 2013. "Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling". *Statistical Science* 28(4):616-640.

Scheidegger, A. P. G., and A. Banerjee. 2017. "An Agent-Based Model to Investigate Behavior Impacts on Vector-Borne Disease Spread". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan et al., 2833-2844. Piscataway, New Jersey: IEEE.

Sharda, B., and A. Banerjee. 2013. "Robust Manufacturing System Design Using Multi Objective Genetic Algorithms, Petri Nets and Bayesian Uncertainty Representation". *Journal of Manufacturing Systems* 32(2):315-324.

Song, E., and B. L. Nelson. 2013. "A Quicker Assessment of Input Uncertainty". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy et al., 474-485. Piscataway, New Jersey: IEEE.

Song, E., B. L. Nelson, and C. D. Pegden. 2014. "Advanced Tutorial: Input Uncertainty Quantification". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk et al., 162-176. Piscataway, New Jersey: IEEE.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. Van Der Sluijs, M. B. Van Asselt, P. Janssen, and M. P. Krayer Von Krauss. 2003. "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support". *Integrated Assessment* 4(1):5-17.

Xie, W., B. L. Nelson, and R. R. Barton. 2014a. "A Bayesian Framework for Quantifying Uncertainty in Stochastic Simulation". *Operations Research* 62(6):1439-1452.

Xie, W., B. L. Nelson, and R. R. Barton. 2014b. "Statistical Uncertainty Analysis for Stochastic Simulation with Dependent Input Models". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk et al., 674-685. Piscataway, New Jersey: IEEE.

AUTHOR BIOGRAPHIES

ANNA PAULA GALVÃO SCHEIDEGGER is a Ph.D. student in the Industrial and Systems Engineering department at Texas A&M University. She holds a master's degree in Industrial Engineering from Universidade Federal de Itajubá, Brazil. Her research interests include modeling and simulation, uncertainty quantification and experiment design. Her email address is apscheidegger@tamu.edu.

AMARNATH BANERJEE is a Professor and Corrie and Jim Furber '64 Faculty Fellow of Industrial and Systems Engineering at Texas A&M University. He received his Ph.D. in Industrial Engineering and Operations Research from the University of Illinois at Chicago, and BS in Computer Science from Birla Institute of Technology and Science, Pilani, India. His research interests are in modeling, simulation, and visualization, with applications in manufacturing, healthcare, and information systems. His email address is banerjee@tamu.edu.

TÁBATA FERNANDES PEREIRA is a Professor of Business Management at Universidade Federal de Itajubá – campus Itabira. She received her Ph.D. and her master's degree in Industrial Engineering from the Universidade Federal de Itajubá, and BS in Information Systems from the Research and Education Foundation of Itajubá. During her doctorate program, she spent one year at Texas A&M University as a Visitor Scholar. Her email address is tabatafp@unifei.edu.br.