Laboratório Nacional de Computação Científica

Programa de Pós-Graduação em Modelagem Computacional

Generalized Lambda Distribution for Uncertainty Quantification of Large-scale Spatio-temporal Models

Noel Moreno Lemus

Petrópolis, RJ - Brasil

Abril de 2018

Noel Moreno Lemus

Generalized Lambda Distribution for Uncertainty Quantification of Large-scale Spatio-temporal Models

Thesis submitted to the examining committee in partial fulfillment of the requirements for the degree of Doctor of Sciences in Computational Modeling.

Laboratório Nacional de Computação Científica

Programa de Pós-Graduação em Modelagem Computacional

Supervisor: Fábio André Machado Porto

Petrópolis, RJ - Brasil

Abril de 2018

XXXX

Moreno Lemus, Noel

Generalized Lambda Distribution for Uncertainty Quantification of Large-scale Spatio-temporal Models / Noel Moreno Lemus. – Petrópolis, RJ - Brasil, Abril de 2018-

41 p. : il. ; 30 cm.

Orientador(es): Fábio André Machado Porto e

Thesis (D.Sc.) – Laboratório Nacional de Computação Científica
Programa de Pós-Graduação em Modelagem Computacional, Abril de 2018.

1. Uncertainty Quantification. 2. Big Data. 3. Information Entropy. I. Machado Porto, Fábio André. II. LNCC/MCTI. III. Title

CDD: XXX.XXX

Noel Moreno Lemus

Generalized Lambda Distribution for Uncertainty Quantification of Large-scale Spatio-temporal Models

Thesis submitted to the examining committee in partial fulfillment of the requirements for the degree of Doctor of Sciences in Computational Modeling.

Approved by:

Prof. Fábio André Machado Porto, D.Sc.
(Presidente)

Prof. Fernando Alves Rochinha, D.Sc.

Prof. Hugo de La Cruz, Ph.D.

Prof. Antonio Tadeu, Ph.D.

Petrópolis, RJ - Brasil
Abril de 2018

Dedication

To my little and special family.

Acknowledgements

O autor manifesta reconhecimentos às pessoas e instituições que colaboraram para a execução de seu trabalho.

"Essentially, all models are wrong, but some are useful."
(George Edward Pelham)

Abstract

Segundo a **??**, 3.1-3.2), o resumo deve ressaltar o objetivo, o método, os resultados e as conclusões do documento. A ordem e a extensão destes itens dependem do tipo de resumo (informativo ou indicativo) e do tratamento que cada item recebe no documento original. O resumo deve ser precedido da referência do documento, com exceção do resumo inserido no próprio documento. (. . .) As palavras-chave devem figurar logo abaixo do resumo, antecedidas da expressão Palavras-chave:, separadas entre si por ponto e finalizadas também por ponto.

Keywords: latex. abntex. editoração de texto.

Abstract

Large-scale spatio-temporal simulations with quantified uncertainty enable scientists/decision-makers to make precise statements about the degree of confidence they have in their simulation-based predictions. This uncertainty could be quantified or characterized in different ways, from the use of low order statistical moments (the most used), to trying to evaluate the complete *PDF* (the best way). Only the characterization of the uncertainty by using the *PDF* allows aware decisions. The uncertainty also needs to be characterized in a way that allows researchers to answer queries that arise in the future.

In this thesis, we propose a new method to quantify the uncertainty in large-scale spatio-temporal models using the Generalized Lambda Distribution (*GLD*). We show how the use of the *GLD* allow us to characterize the uncertainty on each spatio-temporal location in a way that can be used latter to answer questions that arise in the *UQ* context. To ilustrate this, we answer the following questions: (*i*) how to group the output of the *UQ* process based in the simillarity of the uncertainty?, (*ii*) what is the uncertainty in some spatio-temporal locations not analysed previously?, (*iii*) what is the uncertainty of an specific spatio-temporal region?, (*iv*) how to compare two regions as a function of its uncertainty?, and (*v*) what is the less uncertain model from a set of model, to predict some quantity of interest in an spatio-temporal region?

The method was tested in realistic use cases from various scientific areas.

Keywords: Uncertainty Quantification, Large-scale spatio-temporal models, Big Data, Generalized Lambda Distribution

List of Figures

List of Tables

List of abbreviations and acronyms

UQ	Uncertainty Quantification
FP	Forward Problem
QoI	Quantity of Interest
GLDEX	r package to compute the GLD

List of symbols

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta
$\in$	Pertence

Contents

1 Introduction

The rapid growth of high-performance computing and the advances in numerical techniques in the last two decades have provided an unprecedented opportunity to explore complex physical phenomena using large-scale spatio-temporal modeling and simulation. At the same time, scientific community is leaving behind the traditional deterministic approach, which offers point predictions with no associated uncertainty (JOHNSTONE et al., 2016); to include Uncertainty Quantification (UQ) as a common practice in their researches.

Large-scale spatio-temporal simulations with quantified uncertainty enable scientists to make precise statements about the degree of confidence they have in their simulation-based predictions. These approaches find practical applicability in models for predicting the behavior of weather, hurricane forecasts (TOBERGTE; CURTIS, 2013), subsurface hydrology (BARONI; TARANTOLA, 2014), geology (GUERRA et al., 2016), nuclear reactor design, financial portfolios (CHEN; FLOOD; SOWERS, 2008), and biological phenomena, just to name a few. They also allow to study physical phenomena that are impossible to assess experimentally, for example: simulate nuclear accidents, or the conditions that some spatial vehicle will find at landing in Mars, and so on. The success of these techniques has made them increasingly important tools for high impact predictions and decision making.

UQ includes different aspects that warranty the predictive fidelity of a numerical simulation, such as the uncertainty in the experimental data, which is used for defining the parameter values of a model; the propagation of uncertain parameters through the model; and the choice of the model itself. UQ is a complex process that covers the following main tasks: (i) uncertainty characterization (CRESPO; KENNY; GIESY, 2014), also called model calibration (FARRELL, 2015) or statistical inverse problem (ESTACIO-HIROMS; PRUDENCIO, 2012); (ii) sensitivity analysis; (iii) forward problem or uncertainty propagation; and (iv) model selection.

This paper is focused on *forward propagation*, whose objective is to quantify the uncertainties in model output(s) propagated from uncertain inputs. The targets of *forward propagation* analysis can be: (i) evaluate low-order moments (i.e. mean and variance) of the outputs, (ii) evaluate the reliability of the outputs, and/or (iii) assess the complete probability distribution (PDF) of the outputs.

When dealing with large-scale spatio-temporal models, a huge among of data is generated as a result of the simulation process. Indeed, on each spatio-temporal location $(s_i, t_j) \in \mathcal{S} \times \mathcal{T} \subseteq \mathbb{R}^3 \times \mathbb{R}$, usually more than 10^4 simulations are performed. Then, the size of the output dataset is in the order of $N_s \times N_t \times N_{sim}$, where: N_s is the number of spatial locations, N_t is the number of time steps, and N_{sim} is the number of simulations. An example of the volume of data generated by these simulations is given in the experimental section **??** of this paper, where the output dataset is about 2.4 TB. This turn *forward propagation* in a data intensive problem.

Another important aspect, which is often not taken into account, is that the uncertainty need to be quantified in some way that can be used after, to answer questions that arise in the *UQ* context. In that sense, assess the complete *PDF* could be the best way to quantify uncertainty, because if you can find the *PDF* that best fit the dataset with reasonably accurately, you can get all the statistical properties under one roof. At the same time, we can substitute the original data by the *PDFs*, which represents a huge reduction in the volume of data to manipulate.

Contradictorily, statistical moments (e.g. mean and standard deviation) are possibly the most used ways to quantify the uncertainty, despite the fact that they doesn't have information about the manner in which the data are distributed (LAMPASI; Di Nicola; PODESTA, 2006). This is because of the difficulty to find the *PDF* that best fit a dataset (KARIAN; DUDEWICZ, 2011), even more, when dealing with large-scale spatio-temporal models where the *PDF* needs to be derived on each spatio-temporal location, and therefore the *forward propagation* problem becomes time consuming and computationally intensive too.

However, the use of low order moments alone prevents us from making accurate analysis with respect to the uncertainty. They are not enough neither for the characterization nor for the quantification of the uncertainty, and questions such as:

- What is the uncertainty in the spatio-temporal region $\mathcal{S}_i \times \mathcal{T}_j$ associated to the *QoI* q_k and a computational model $\mathcal{M}_m$?
- How to compare different spatio-temporal regions $\mathcal{S}_i \times \mathcal{T}_j$ with respect to the uncertainty?
- What is the less uncertain model from the set of models $\mathcal{M} = \mathcal{M}_1, \mathcal{M}_2, ...\mathcal{M}_m$, to predict the value of a QoI q_k, over a spatio-temporal region $\mathcal{S}_i \times \mathcal{T}_j$?

can be poorly answered. So, we emphasize that only the characterization of the uncertainty by using the *PDF* allows aware decisions.

A first effort to try to estimate the *PDFs* on large-scale spatio-temporal simulations was done by Ji et. al. in ***Parallel Computation of PDFs on Big Spatial Data Using Spark***. They propose a new solution to efficiently compute the *PDFs* in parallel using Spark, through three methods: data grouping, machine learning prediction and sampling. The main drawback of the proposed approach is that you should try many different distributions, to find the PDF that best fits the dataset on each specific spatio-temporal location. Another drawback is that, as we mentioned above, the uncertainty needs to be quantified in the way that facilitates its further use; and the heterogeneity of the functions used in the approach doesn't facilitate it.

To face these challenges, in this paper we propose a general framework to quantify the uncertainty in large-scale spatio-temporal models. It uses a data-driven approach and combines the generalized lambda distribution (*GLD*), clusters algorithms and information entropy, for helping researchers to answer the above questions and many others that arise in *UQ* context. Our proposal provides a generally applicable and easy-to-use tool that supports the representation and analysis of uncertainty, as was suggested in the "*Workshop on Quantification, Communication, and Interpretation of Uncertainty in Simulation and Data Science*" (TOBERGTE; CURTIS, 2013).

In order to illustrate the use of the proposed framework, a case study is discussed. The main results obtained are: (i) the *GLD* good fits for more than the 80 % of the dataset, (ii) the use of the *GLD* allows to include clustering algorithms to group the spatio-temporal locations with similar uncertainty, (iii) the centroids of the clusters can be used as a faithful representation of the rest of the spatio-temporal locations, which significantly reduces the data corresponding to the simulation outputs, (iv) with the use of these centroids we can characterize the uncertainty in any spatio-temporal region as a mixture of *GLDs*.

The rest of the paper is organized as follows: Section **??** gives the theoretical foundations of *UQ* and highlights some interesting aspects included in our proposal. Section **??** describes the principal characteristics of the *GLD* that make it suitable for this proposal. Section **??** presents the proposed approach, the workflow we implement and some considerations of the implementation. Section **??** presents a use case and discusses the results. This use case allows us to explain our approach in the context of a real problem, which facilitates its understanding. Section **??** covers the related works and finally, section 8 concludes the paper and proposes some future works.

1.1 Research Objectives

1.2 Highlights of the Dissertation

1.3 Organization of the Dissertation

2 Related Works

2.1 Overview

HPC and computational modeling play a dominant role in shaping the methodological developments and research in uncertainty qualification. Depending on the complexity of the uncertainty qualification investigation, anywhere from 10^2 to 10^8 runs of the computational model may be required. Thus, uncertainty qualification investigations may require extreme-computing environments (e.g., exascale) to obtain results in a useful time frame, even if a single run of the computational model does not require such resources.

Advances in computing over the past few decades—both in availability and power—have led to an explosion in computational models available for simulating a wide variety of complex physical (and social) systems. These complex models—which may involve millions of lines of code, and require extreme-computing resources—have led to numerous scientific discoveries and advances. This is because these models allow simulation of physical processes in environments and conditions that are difficult or even impossible to access experimentally. However, scientists' abilities to quantify uncertainties in these model-based predictions lag well behind their abilities to produce these computational models. This is largely because such simulation-based scientific investigations present a set of challenges that is not present in traditional investigations.

Until recently, the original approach of describing model parameters using single values has been retained, and consequently the majority of mathematical models in use today provide point predictions, with no associated uncertainty. (JOHNSTONE et al., 2016)

An immediate challenge in the development of an appropriate treatment of uncertainty in an analysis of a complex system is the selection of a mathematical structure to be used in the representation of uncertainty. (HELTON et al., 2010a) Traditionally, probability theory has provided this structure [48-55]. However, in the last several decades, additional mathematical structures for the representation of uncertainty such as evidence theory [56-63], possibility theory [64- 70], fuzzy set theory [71-75], and interval analysis [76-81] have been introduced. This introduction has been accompanied by a lively discussion of the strengths and weaknesses of the various mathematical structures for the representation of uncertainty [82-90]. For perspective, several comparative discussions of these different approaches to the representation of uncertainty are available [72; 91-98]

a 'typical' UQ problem involves one or more mathematical models for a process of interest, subject to some uncertainty about the correct form of, or parameter values for, those models.

Often, though not always, these uncertainties are treated probabilistically.

but how will you actually go about evaluating that expected value when it is an integral over a million-dimensional parameter space? Practical problems from engineering and the sciences can easily have models with millions or billions of inputs (degrees of freedom).

the language of probability theory is a powerful tool in describing uncertainty

UQ cannot tell you that your model is 'right' or 'true', but only that, if you accept the validity of the model (to some quanti-fied degree), then you must logically accept the validity of certain conclusions (to some quantified degree). (SULLIVAN, 2015)

"UQ studies all sources of error and uncertainty, including the following: systematic and stochastic measurement error; ignorance; limitations of theoretical models; limitations of numerical representations of those models; limitations of the accuracy and reliability of computations, approximations, and algorithms; and human error. A more precise definition is UQ is the end-to-end study of the reliability of scientific ∞erences."

UQ is not a mature field like linear algebra or single-variable complex analysis, with stately textbooks containing well-polished presentations of classical theorems bearing August names like Cauchy, Gauss and Hamilton. Both because of its youth as a field and its very close engagement with applications, UQ is much more about problems, methods and 'good enough for the job'. There are some very elegant approaches within UQ, but as yet no single, general, over-arching theory of UQ.

In

Probability theorists usually denote the sample space of a probability space by Ω; PDE theorists often use the same letter to denote a domain in $\Re^n$ on which a partial differential equation is to be solved. In UQ, where the worlds of probability and PDE theory often collide, the possibility of confusion is clear. Therefore, this book will tend to use Θ for a probability space and $\mathbf{X}$ for a more general measurable space, which may happen to be the spatial domain for some PDE.

2.2 Types of Uncertainty

It is sometimes assumed that uncertainty can be classified into two categories, (KIUREGHIAN; DITLEVSEN, 2009) although the validity of this categorization is open to debate.

Aleatory uncertainty arises from an inherent randomness in the properties or behavior of the system under study. For example, the weather conditions at the time of a reactor accident are inherently random with respect to our ability to predict the future. Other examples include the variability in the properties of a population of weapon components and the variability in the possible future environmental conditions that a weapon component could be exposed to. Alternative designa- tions for aleatory uncertainty include variability, stochastic, irreducible and type A. (HELTON, 2009)

Epistemic uncertainty derives from a lack of knowledge about the appropriate value to use for a quantity that is assumed to have a fixed value in the context of a particular analysis. For example, the pressure at which a given reactor containment would fail for a specified set of pressurization conditions is fixed but not amena- ble to being unambiguously defined. Other examples include minimum voltage required for the operation of a system and the maximum temperature that a system can withstand before failing. Alternative designations for epistemic uncertainty include state of knowledge, subjective, reducible and type B. (HELTON, 2009)

2.2.1 Aleatory uncertainty

Aleatory uncertainty arises from an inherent randomness in the properties or behavior of the system under study. For example, the weather conditions at the time of a reactor accident are inherently random with respect to our ability to predict the future. Other examples include the variability in the properties of a population of weapon components and the variability in the possible future environmental conditions that a weapon component could be exposed to. Alternative designa- tions for aleatory uncertainty include variability, stochastic, irreducible and type A. (HELTON, 2009)

2.2.2 Epistemic uncertainty

2.3 Uncertainty Representation

The question of how to represent and communicate uncertainties is a topic of research both from a practical and theoretical point of view. A fair bit of theoretical research is aimed at the mathematical calculus of uncertainty. This includes extensions and alternatives to standard probabilistic reasoning, such as Dempster-Schafer theory and imprecise probabilities. When uncertainties are needed for investigations requiring computational models, additional considerations arise. For example, if the simulation output is a daily surface-temperature field over the globe for the next 200 years, representing uncertainty and dependencies is complex. Should ensembles be used to represent plausible outcomes? How should these ensembles of simulation output be stored? How can high-consequence/low-probability outcomes be discovered in this massive output? Here some research investigations attempt to leverage theory that exploits high dimensionality to bound probabilities and system behavior. Finally, even when uncertainties are well captured,

Table 1 – The range of the GLD parameters and the minimum and maximum values corresponding to the labeling of the regions given in Figure

Region	λ_1	λ_2	λ_3	λ_4	Minimum	Maximum
1 and 5	all	< 0	< -1	> 1	$-\infty$	$\lambda_1 + \frac{1}{\lambda_2}$
2 and 6	all	< 0	> 1	< -1	$\lambda_1 - \frac{1}{\lambda_2}$	∞
3	all	> 0	> 0	> 0	$\lambda_1 - \frac{1}{\lambda_2}$	$\lambda_1 + \frac{1}{\lambda_2}$
	all	> 0	$= 0$	> 0	λ_1	$\lambda_1 + \frac{1}{\lambda_2}$
	all	> 0	> 0	$= 0$	$\lambda_1 - \frac{1}{\lambda_2}$	λ_1
4	all	< 0	< 0	< 0	$-\infty$	∞
	all	< 0	$= 0$	< 0	λ_1	∞
	all	< 0	< 0	$= 0$	$-\infty$	λ_1

how best to communicate such uncertainties to the public or to decision-makers is also a topic of ongoing research.

(HELTON et al., 2010b)

2.3.1 Representation of Uncertainty with Probability

2.3.2 Dempster-Shafer theory

2.3.3 The Bayesian Methodology

2.4 Methods for Uncertainty Propagation

2.5 Probabilistic Background

2.5.1 The Generalized Lambda Distribution

The Generalized Lambda Distribution (GLD) was defined by Ramberg and Schmeiser in 1974 by the quantil function:

$$F^{-1}(p|\lambda) = F^{-1}(p|\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2} \tag{2.1}$$

where p are the probabilities, $p \in [0, 1]$, λ_1 and λ_2 are the location and scale parameteres, and λ_3 and λ_4 determine the skewness and kurtosis of the $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$.

Some restrictions in the values of $\lambda_1, \lambda_2, \lambda_3$ and λ_4 define if the GLD is valid. Those restrictions define 6 regions as is shown in table

The probability density function of the GLD at the point $x = F^{-1}(p)$ is given by:

$$f(x) = f(F^{-1}(p)) = \frac{\lambda_2}{\lambda_3 p^{\lambda_3 - 1} + \lambda_4 (1-p)^{\lambda_4 - 1}} \tag{2.2}$$

Note that the valid parameteres of λ guaranty that:

$$f(x) \geqslant 0 \tag{2.3}$$

$$\int f(x)dx = 1 \tag{2.4}$$

2.5.2 Fitting Mixture Distributions Using a Mixture of Generalized Lambda Distributions

Esto esta en (TOBERGTE; CURTIS, 2013)

2.5.3 Sampling Methods

2.5.3.1 Monte Carlo

2.6 Summary

2.7 Concepts

high-dimensional parameter spaces computationally demanding forward models nonlinearity and/or complexity in the forward model

2.8 Software and Tools for UQ

These include both free software, like OpenTURNS (Andrianov et al., 2007), DACOTA (Adams et al., 2009) and DUE (Brown and Heuvelink, 2007), commercial, like COSSAN (Schuëller and Pradlwarter, 2006), or free, but written for a licenced software, e.g. SAFE (Pianosi et al., 2015) or UQLab (Marelli and Sudret, 2014) toolboxes for MATLAB. A broad review of existing software packages is available in Bastin et al. (2013). To the best of our knowledge, however, none of the existent software is specifically designed to be extended by the environmental science community. The use of powerful but complex languages like C++ (e.g. Dakota), Python (e.g. OpenTURNS) or Java (e.g. DUE) often discourages relevant portions of the non-highly-IT trained scientific community from the adoption of otherwise powerful tools. spup-R package (K. Sawicka; SOIL, 2016). De aqui saque lo de arriba tambien, aunque lo de arriba lo puedo buscar en sus respectivos papers y hablar un poco de cada uno de ellos.

Currently, advances in uncertainty propagation and assessment have been paralleled by a growing number of software tools for uncertainty analysis, but none has gained recognition for a universal applicability, including case studies with spatial models and spatial model inputs. (K. Sawicka; SOIL, 2016)

3 Uncertainty Quantification Process

3.1 Measures of Information and Uncertainty

3.1.1 Variance, Information and Entropy

Variance.

Information and Entropy.

3.1.2 Information Gain, Distances and Divergences

3.2 Sensitivity Analysis

Sensitivity analysis is the systematic study of how model inputs—parameters, initial and boundary conditions—affect key model outputs. Depending on the application, one might use local derivatives or global descriptors such as Sobol's functional decomposition or variance decomposition. Also, the needs of the application may range from simple ranking of the importance of inputs to a response surface model that predicts the output given the input settings. Such sensitivity studies are complicated by a number of factors, including the dimensionality of the input space, the complexity of the computational model, limited forward model runs due to the computational demands of the model, the availability of adjoint solvers or derivative information, stochastic simulation output, and high-dimensional output. Challenges in sensitivity analysis include dealing with these factors while addressing the needs of the application. (**??**)

$$E = mc^2 \tag{3.1}$$

4 Parallel Computation of PDFs

4.1 Introduction

4.2 Architecture for Computing PDFs in Spark

4.3 Experimental Evaluation

5 The Generalized Lambda Distribution

5.1 From Emperimental Data to GLD Paremeters

(LAMPASI; Di Nicola; PODESTA, 2006)

5.2 GLD Shapes

5.3 The GLDEX R package

5.4 GLD mixture

6 Our Approach

6.1 Fit the spatio-temporal dataset to the GLD

Aqui tengo que poner:

- Fit each spatio-temporal point to a corresponding GLD.
- Evaluate if the resulting GLD is valid on each spatio-temporal location.
- Perform a ks-test to evaluate if the quality of the fit on each spatio-temporal location.

6.2 Clusterizing the GLD based in its lambda values

6.3 Use of GLD mixture to characterize the uncertainty in an spatio-temporal region

6.4 Information entropy as a measure of the uncertainty in an spatio-temporal region

6.5 Information entropy and model selection

6.6 Conclusions

7 Applicability

In the present chapter we are going to test the UQMS in three different scenarios, spatial only domain, section 7.1 , spatio-temporal domain, section 7.2, and finally a multidisciplinary system, section 7.3.

7.1 Case Study: Wave Propagation Problem

The first one is a geophysical tests for wave propagation problems

As a first case study we use the "HPC4E Seismic Test Suite", a collection of four 3D models and sixteen associated tests that can be downloaded freely at the project's website (https://hpc4e.eu/downloads/datasets-and-software). The models include simple cases that can be used in the development stage of any geophysical imaging practitioner (developer, tester ...) as well as extremely large cases that can only be solved in a reasonable time using ExaFLOPS supercomputers. The models are generated to the required size by means of a Matlab/Octave script and hence can be used by users of any OS or computing platform. The tests can be used to benchmark and compare the capabilities of different and innovative seismic modelling approaches, hence simplifying the task of assessing the algorithmic and computational advantages that they pose.

In our case, we are going to use the "HPC4E Seismic Test Suite" as a case study of the porposed UQMS. As we mention in the introduction of this chapter this model is a spatial only domain problem, because we are going to consider a multidimentional array as an Input and a multidimentional array as an output, but of them time independet.

7.1.1 Mathematical Formulation

7.1.2 Model and Dataset Description

The models have been designed as a set of 16 layers with constant physical properties. The top layer delineates the topography and the other 15 different layer interface surfaces or horizons. In the following, an interface horizon is associated with properties that apply to the layer that exists between itself and the immediately next layer horizon. The model covers an area of 10 x 10 x 5 km, with maximum topography at about 500 m and maximum depth at about 4500 m. The layer horizons have been sampled very finely with 1.6667 m spacing so that a highly accurate representation can be honored at high frequencies. For simulation schemes based on unstructured grids, the layer horizons can be used easily to constrain model blocks. For simulation schemes based upon Cartesian grids, a simple

Table 2 – Layer constant properties and their depth range. "Star" layers are only used in the flat case, in substitution of their non-star equivalents

Layer Id	**Vp (m/s)**	**Vs (m/s)**	**Density (Kg/m3)**	**Max. depth (m)**	**Min. depth (m)**
1	1618.92	500.00	1966.38	-135.55	-476.35
2	1684.08	765.49	1985.88	41.50	-394.90
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
2*					
3*					

script is provided that can generate 3D grids for any desired spatial sampling. Table 2 shows the properties of each of the layers included in the models.

7.1.3 Adding uncertainty into the model

The "HPC4E Seismic Test Suite" does not provide uncertainty sources, because all the input parameters of the model have fixed values. Then, to the purpose of our work we need to add some uncertainties into the inputs. Let's suppose the variable V_p is uncertain. As this variable have 16 different values, one for each layer, we can consider it as a random vector, equation 7.1. We associate to each of the V_{p_i} a Normal distribution with μ_i equal to the value reported in Table 2 and $\sigma = 2$.

$$V_p = < V_{p_i}, \mathcal{N}(\mu_i, \sigma_i) > \tag{7.1}$$

7.2 Case Study: Austin, queso library

7.3 Case Study: Multidisciplinary System (NASA)

7.4 Case Study: Spatio-temporal Nicholson-Bailey model

Este esta en el software uqlab, en la carpeta Doc Manuals

8 Conclusions and Future Works

8.1 Revisiting the Research Questions

8.2 Significance and Limitations

8.3 Open Problems and Future Work

8.4 Final Considerations

Bibliography

BARONI, G.; TARANTOLA, S. A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. *Environmental Modelling and Software*, Elsevier Ltd, v. 51, p. 26–34, 2014. ISSN 13648152. Disponível em: <http://dx.doi.org/10.1016/j.envsoft.2013.09.022>. Citado na página 16.

CHEN, J.; FLOOD, M. D.; SOWERS, R. B. Measuring the Unmeasurable: An Application of Uncertainty Quantification to Financial Portfolios Measuring the Unmeasurable An application of uncertainty quantification to financial portfolios. *Quantitative Finance*, v. 7688, n. January, p. 1–18, 2008. ISSN 14697696. Disponível em: <http://dx.doi.org/10.1080/14697688.2017.1296176>. Citado na página 16.

CRESPO, L. G.; KENNY, S. P.; GIESY, D. P. The NASA Langley Multidisciplinary Uncertainty Quantification Challenge. *16th AIAA Non-Deterministic Approaches Conference*, n. January, p. 1–9, 2014. Disponível em: <http://arc.aiaa.org/doi/abs/10.2514/6.2014-1347>. Citado na página 16.

ESTACIO-HIROMS, K. C.; PRUDENCIO, E. E. User's Manual: Quantification of Uncertainty for Estimation, Simulation, and Optimization (QUESO). 2012. Citado na página 16.

FARRELL, K. A. Selection , Calibration , and Validation of Coarse-Grained Models of Atomistic Systems. 2015. Citado na página 16.

GUERRA, G. M. et al. Uncertainty quantification in numerical simulation of particle-laden flows. *Computational Geosciences*, v. 20, n. 1, p. 265–281, 2016. ISSN 1420-0597. Disponível em: <http://link.springer.com/10.1007/s10596-016-9563-6>. Citado na página 16.

HELTON, J. Conceptual and computational basis for the quantification of margins and uncertainty. n. June, 2009. Disponível em: <http://www.osti.gov/energycitations/product.biblio.jsp?osti{_}id=958>. Citado na página 21.

HELTON, J. C. et al. Representation of analysis results involving aleatory and epistemic uncertainty. *International Journal of General Systems*, Taylor & Francis Group, v. 39, n. 6, p. 605–646, 2010. ISSN 0308-1079. Citado na página 19.

HELTON, J. C. et al. Representation of analysis results involving aleatory and epistemic uncertainty. *International Journal of General Systems*, Taylor & Francis Group, v. 39, n. 6, p. 605–646, 2010. ISSN 0308-1079. Citado na página 22.

JOHNSTONE, R. H. et al. Uncertainty and variability in models of the cardiac action potential: Can we build trustworthy models? *Journal of Molecular and Cellular Cardiology*, The Authors, v. 96, p. 49–62, 2016. ISSN 10958584. Disponível em: <http://dx.doi.org/10.1016/j.yjmcc.2015.11.018>. Citado 2 vezes nas páginas 16 and 19.

K. Sawicka, G. H.; SOIL. spup- an R package for uncertainty propagation in spatial environmental modelling. *International symposium on "Spatial Accuracy Assessment in Natural Resources and Environmental Sciences"*, v. 53, n. 9, p. 1689–1699, 2016. ISSN

1098-6596. Disponível em: <http://spatial-accuracy.org/Accuracy2016>. Citado na página 23.

KARIAN, Z. A.; DUDEWICZ, E. J. *Handbook of fitting statistical distributions with R*. [S.l.: s.n.], 2011. ISSN 1098-6596. ISBN 9788578110796. Citado na página 17.

KIUREGHIAN, A. D.; DITLEVSEN, O. Aleatory or epistemic? Does it matter? *Structural Safety*, v. 31, n. 2, p. 105–112, mar 2009. ISSN 01674730. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0167473008000556>. Citado na página 20.

LAMPASI, D. A.; Di Nicola, F.; PODESTA, L. Generalized lambda distribution for the expression of measurement uncertainty. *IEEE Transactions on Instrumentation and Measurement*, v. 55, n. 4, p. 1281–1287, 2006. ISSN 00189456. Citado 2 vezes nas páginas 17 and 26.

SULLIVAN, T. J. *Introduction to Uncertainty Quantification*. Springer, 2015. ISBN 9783319233949. Disponível em: <http://www.springer.com/series/1214>. Citado na página 20.

TOBERGTE, D. R.; CURTIS, S. Workshop on Quantification, Communication, and Interpretation of Uncertainty in Simulation and Data Science. *Journal of Chemical Information and Modeling*, v. 53, n. 9, p. 1689–1699, 2013. ISSN 1098-6596. Citado 3 vezes nas páginas 16, 18, and 23.

Appendix

APPENDIX A – uqms R package

Quisque facilisis auctor sapien. Pellentesque gravida hendrerit lectus. Mauris rutrum sodales sapien. Fusce hendrerit sem vel lorem. Integer pellentesque massa vel augue. Integer elit tortor, feugiat quis, sagittis et, ornare non, lacus. Vestibulum posuere pellentesque eros. Quisque venenatis ipsum dictum nulla. Aliquam quis quam non metus eleifend interdum. Nam eget sapien ac mauris malesuada adipiscing. Etiam eleifend neque sed quam. Nulla facilisi. Proin a ligula. Sed id dui eu nibh egestas tincidunt. Suspendisse arcu.

A.1 Título da seção

Aqui temos uma seção dentro do Apêndice.

Figure 1 – Legenda para a figura.

APPENDIX B – Título do apêndice B

Nunc velit. Nullam elit sapien, eleifend eu, commodo nec, semper sit amet, elit. Nulla lectus risus, condimentum ut, laoreet eget, viverra nec, odio. Proin lobortis. Curabitur dictum arcu vel wisi. Cras id nulla venenatis tortor congue ultrices. Pellentesque eget pede. Sed eleifend sagittis elit. Nam sed tellus sit amet lectus ullamcorper tristique. Mauris enim sem, tristique eu, accumsan at, scelerisque vulputate, neque. Quisque lacus. Donec et ipsum sit amet elit nonummy aliquet. Sed viverra nisl at sem. Nam diam. Mauris ut dolor. Curabitur ornare tortor cursus velit.

Morbi tincidunt posuere arcu. Cras venenatis est vitae dolor. Vivamus scelerisque semper mi. Donec ipsum arcu, consequat scelerisque, viverra id, dictum at, metus. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut pede sem, tempus ut, porttitor bibendum, molestie eu, elit. Suspendisse potenti. Sed id lectus sit amet purus faucibus vehicula. Praesent sed sem non dui pharetra interdum. Nam viverra ultrices magna.

Aenean laoreet aliquam orci. Nunc interdum elementum urna. Quisque erat. Nullam tempor neque. Maecenas velit nibh, scelerisque a, consequat ut, viverra in, enim. Duis magna. Donec odio neque, tristique et, tincidunt eu, rhoncus ac, nunc. Mauris malesuada malesuada elit. Etiam lacus mauris, pretium vel, blandit in, ultricies id, libero. Phasellus bibendum erat ut diam. In congue imperdiet lectus.

APPENDIX C – Título do apêndice C

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Annex

ANNEX A – Título do anexo A

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

Maecenas non massa. Vestibulum pharetra nulla at lorem. Duis quis quam id lacus dapibus interdum. Nulla lorem. Donec ut ante quis dolor bibendum condimentum. Etiam egestas tortor vitae lacus. Praesent cursus. Mauris bibendum pede at elit. Morbi et felis a lectus interdum facilisis. Sed suscipit gravida turpis. Nulla at lectus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Praesent nonummy luctus nibh. Proin turpis nunc, congue eu, egestas ut, fringilla at, tellus. In hac habitasse platea dictumst.

A.1 Título da seção

Aqui temos uma seção dentro do Anexo.

Vivamus eu tellus sed tellus consequat suscipit. Nam orci orci, malesuada id, gravida nec, ultricies vitae, erat. Donec risus turpis, luctus sit amet, interdum quis, porta sed, ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempor orci, et tincidunt urna nunc a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed vitae sem. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Maecenas ante. Duis ullamcorper enim. Donec tristique enim eu leo. Nullam molestie elit eu dolor. Nullam bibendum, turpis vitae tristique gravida, quam sapien tempor lectus, quis pretium tellus purus ac quam. Nulla facilisi.

ANNEX B – Título do anexo B

Curabitur nunc magna, posuere eget, venenatis eu, vehicula ac, velit. Aenean ornare, massa a accumsan pulvinar, quam lorem laoreet purus, eu sodales magna risus molestie lorem. Nunc erat velit, hendrerit quis, malesuada ut, aliquam vitae, wisi. Sed posuere. Suspendisse ipsum arcu, scelerisque nec, aliquam eu, molestie tincidunt, justo. Phasellus iaculis. Sed posuere lorem non ipsum. Pellentesque dapibus. Suspendisse quam libero, laoreet a, tincidunt eget, consequat at, est. Nullam ut lectus non enim consequat facilisis. Mauris leo. Quisque pede ligula, auctor vel, pellentesque vel, posuere id, turpis. Cras ipsum sem, cursus et, facilisis ut, tempus euismod, quam. Suspendisse tristique dolor eu orci. Mauris mattis. Aenean semper. Vivamus tortor magna, facilisis id, varius mattis, hendrerit in, justo. Integer purus.

Vivamus adipiscing. Curabitur imperdiet tempus turpis. Vivamus sapien dolor, congue venenatis, euismod eget, porta rhoncus, magna. Proin condimentum pretium enim. Fusce fringilla, libero et venenatis facilisis, eros enim cursus arcu, vitae facilisis odio augue vitae orci. Aliquam varius nibh ut odio. Sed condimentum condimentum nunc. Pellentesque eget massa. Pellentesque quis mauris. Donec ut ligula ac pede pulvinar lobortis. Pellentesque euismod. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent elit. Ut laoreet ornare est. Phasellus gravida vulputate nulla. Donec sit amet arcu ut sem tempor malesuada. Praesent hendrerit augue in urna. Proin enim ante, ornare vel, consequat ut, blandit in, justo. Donec felis elit, dignissim sed, sagittis ut, ullamcorper a, nulla. Aenean pharetra vulputate odio.

ANNEX C – Título do anexo C

Nunc velit. Nullam elit sapien, eleifend eu, commodo nec, semper sit amet, elit. Nulla lectus risus, condimentum ut, laoreet eget, viverra nec, odio. Proin lobortis. Curabitur dictum arcu vel wisi. Cras id nulla venenatis tortor congue ultrices. Pellentesque eget pede. Sed eleifend sagittis elit. Nam sed tellus sit amet lectus ullamcorper tristique. Mauris enim sem, tristique eu, accumsan at, scelerisque vulputate, neque. Quisque lacus. Donec et ipsum sit amet elit nonummy aliquet. Sed viverra nisl at sem. Nam diam. Mauris ut dolor. Curabitur ornare tortor cursus velit.

Morbi tincidunt posuere arcu. Cras venenatis est vitae dolor. Vivamus scelerisque semper mi. Donec ipsum arcu, consequat scelerisque, viverra id, dictum at, metus. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut pede sem, tempus ut, porttitor bibendum, molestie eu, elit. Suspendisse potenti. Sed id lectus sit amet purus faucibus vehicula. Praesent sed sem non dui pharetra interdum. Nam viverra ultrices magna.

Aenean laoreet aliquam orci. Nunc interdum elementum urna. Quisque erat. Nullam tempor neque. Maecenas velit nibh, scelerisque a, consequat ut, viverra in, enim. Duis magna. Donec odio neque, tristique et, tincidunt eu, rhoncus ac, nunc. Mauris malesuada malesuada elit. Etiam lacus mauris, pretium vel, blandit in, ultricies id, libero. Phasellus bibendum erat ut diam. In congue imperdiet lectus.