Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

An Empirical Workflow of Uncertainty Quantification to Evaluate Agent-based Simulation Outputs Aiming Analytical Confidence

Carolina Gonçalves Abreu

Tese apresentada como requisito parcial
para conclusão do Doutorado em Informática

Orientadora
Prof.ª Dr.ª Célia Ghedini Ralha

Brasília
2018

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Doutorado em Informática

Coordenador: Prof. Dr. Bruno Luiggi Macchiavello Espinoza

Banca examinadora composta por:

Prof.ª Dr.ª Célia Ghedini Ralha (Orientadora) — CIC/UnB
Prof. Dr. Anthony Jakeman — ANU
Prof. Dr. Christophe Yvon Le Page — CIRAD, PTARH/UnB
Prof. Dr. Maria Emília Machado Telles Walter — CIC/UnB

CIP — Catalogação Internacional na Publicação

Abreu, Carolina Gonçalves.

An Empirical Workflow of Uncertainty Quantification to Evaluate Agent-based Simulation Outputs Aiming Analytical Confidence / Carolina Gonçalves Abreu. Brasília : UnB, 2018.
154 p. : il. ; 29,5 cm.

Tese (Doutorado) — Universidade de Brasília, Brasília, 2018.

1. análise de incerteza, 2. análise de sensibilidade, 3. avaliação integrada, 4. validação de modelo, 5. modelo baseado em agentes, 6. uso da terra

CDU 004.4

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília–DF — Brasil

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

An Empirical Workflow of Uncertainty Quantification to Evaluate Agent-based Simulation Outputs Aiming Analytical Confidence

Carolina Gonçalves Abreu

Tese apresentada como requisito parcial
para conclusão do Doutorado em Informática

Prof.ª Dr.ª Célia Ghedini Ralha (Orientadora)
CIC/UnB

Prof. Dr. Anthony Jakeman
ANU

Prof. Dr. Christophe Yvon Le Page
CIRAD, PTARH/UnB

Prof. Dr. Maria Emília Machado Telles Walter
CIC/UnB

Prof. Dr. Bruno Luiggi Macchiavello Espinoza
Coordenador do Doutorado em Informática

Brasília, 23 de agosto de 2018

Aos meus pais,
Carla Vieira Gonçalves Abreu e
Wilson José Rodrigues Abreu.

Agradecimentos

Registro o agradecimento à minha orientadora, Prof.Dr. Célia Ghedini Ralha, pela incansável disposição em discutir cada etapa dessa pesquisa. Pelo incentivo para que eu não deixasse nenhuma questão sem resposta. Pela convivência e amizade que, ao longo dos anos, muito me ensinou.

Ao Cassio, por toda a paciência e apoio durante todos os processos de concepção, construção e testes da ferramenta. Ao Rafael e ao Jeremias, pelo suporte durante a extensiva fase de experimentos. Ao Murilo pelo auxílio com a integração das ferramentas estatísticas. Ao Jailton, por toda a paciência e dedicação ao me ensinar sobre mapas e sistemas de informação geográfica.

Aos professores Conceição, Genaína e George, por suas valiosas sugestões e pela indicação de excelentes fontes bibliográficas.

Aos professores e colegas da Escola Multiplataforma Internacional de Verão sobre Modelagem e Simulação Baseada em Agentes para o Gerenciamento de Recursos Renováveis, por proporem questões e reflexões que me acompanharam durante o período de finalização da tese.

Também gostaria de expressar minha gratidão a todos os co-autores dos artigos publicados nesse período, uma vez que esse trabalho avançou a cada nova publicação escrita.

À Universidade de Brasília, por acolher e apoiar a confecção desse trabalho.

"Research is the act of going up alleys to see if they're blind."

Plutarch

Resumo Expandido

A investigação sobre a mudança de uso e cobertura da terra é importante para promover o gerenciamento criterioso do território, como meio de contenção de danos ambientais. Além disso, é um processo complexo que relaciona a interação entre sistemas ambientais, econômicos e sociais em diferentes escalas temporais e espaciais. O entendimento da dinâmica desses sistemas foca não somente nas partes, mas do comportamento que emerge da interação entre elas. Modelo Baseado em Agente (MBA) é uma boa técnica para o estudo desses fenômenos, uma vez que modela as interações entre agentes autônomos e o seu ambiente. As simulações computacionais são a técnica mais utilizada para avaliar esses modelos, para testar explicitamente os efeitos das decisões humanas em situações complexas.

Conquanto os MBAs forneçam uma ferramenta poderosa para analisar fenômenos emergentes, sua utilidade é limitada por dificuldades na análise dos seus resultados, o que fomenta críticas e questionamentos sobre a contribuição real dos *frameworks* para o suporte à decisão. A ferramenta mais difundida para avaliação é a análise de sensibilidade, pois quantifica os efeitos das alterações nos fatores de entrada do modelo nas previsões do modelo. Entretanto, grande parte dos métodos mais difundidos de análise de sensibilidade não são adequados ou são insuficientes para lidar com as especificidades advindas da complexidade dos MBAs. Dentre elas, destacam-se a estocasticidade, não-linearidade e a parametrização *ad hoc*, que implicam uma considerável incerteza epistêmica. Sem uma investigação apropriada, há chances significativas de que os resultados derivados da simulação sejam a consequência de vieses.

Embora reconhecendo as diferenças particulares dos inúmeros MBAs, a presente tese examina se esses desafios podem ser superados, no contexto de um estudo de caso de uso e cobertura da terra no Cerrado do Distrito Federal, usando a ferramenta multiagente MASE-BDI (Coelho et al., 2016). O objetivo dessa tese é avaliar a aplicação de várias metodologias de quantificação de incerteza e análise de sensibilidade na análise de resultados de MBAs. O foco da pesquisa é efetuar uma aplicação integrada de técnicas de análise de incerteza e sensibilidade e avaliar os impactos que as diferenças nos tamanhos de amostra, técnicas de amostragem e métodos de análise de sensibilidade podem ter na

saída do modelo. Além disso, propõe-se um *workflow* para que essas técnicas possam ser aplicadas de forma organizada e sistemática. De modo mais abrangente, aplica-se uma metodologia geral de avaliação de MBAs, que inclui diferentes abordagens para produzir versões simplificadas do modelo que podem ser usadas para explorar os resultados ou realizar uma análise exploratória. Estas abordagens para análise, calibração e verificação do modelo requerem um grande número de execuções de simulação de cenários repetidos e com muitas combinações de parâmetros e de configurações do modelo. Para facilitar esse processo, foi implementada a integração da ferramenta de simulação MASE-BDI com o conjunto de bibliotecas estatísticas para quantificação de incerteza PSUADE. Houve a criação de um *driver* e de uma interface para automatizar o pré e pós-processamento de entradas e saídas para muitas execuções do modelo.

Todos os experimentos foram testados em um modelo espacialmente explícito de uso e cobertura da terra. A ferramenta de simulação é o MASE-BDI, desenvolvido pela Universidade de Brasília. MASE é o acrônimo para *MultiAgent System for Environmental simulation* que implementa o modelo de racionalidade *Belief-Desire-Intention* (BDI). No BDI, os agentes possuem crenças, um conjunto de informações que se tem sobre o ambiente que habitam e que alteram tanto a percepção quanto o seu pensamento sobre o mundo. Desejos, que representam as atitudes motivacionais dos agentes que os conduzem a um curso de ação, e intenções, que são o conjunto de planos montados pelo agente para que ele atinja os seus objetivos. A função objetivo da análise dos resultados da simulação é uma métrica estatística de aptidão denominada figura de mérito (FoM), determinada pela razão entre as mudanças na terra que foram preditas corretamente sobre a soma das mudanças observadas (Pontius et al., 2008). Essa métrica quantifica se os acertos de um mapa de uso e cobertura da terra são maiores que os erros na predição da quantidade de conversão entre os diferentes usos e coberturas da terra e da alocação dessas mudanças no espaço.

A metodologia utilizada na tese foi incremental e evolutiva. Inicialmente, foi realizada uma avaliação do modelo com a utilização dos métodos mais difundidos na literatura: análise de sensibilidade um-fator-de-cada-vez (OAT - *One-factor-At-a-Time*) para quatro fatores de entrada e um número variável de replicações. Para avaliar a qualidade da saída do modelo, a métrica de aptidão foi avaliada por meio de intervalos de confiança. Os resultados mostraram que apesar de ser possível diferenciar os fatores de entrada sensíveis e não sensíveis, a variabilidade da saída era tão grande que a incerteza impedia qualquer análise mais robusta. Percebeu-se que diferentes replicações da amostra afetavam consideravelmente os resultados.

A revisão de literatura apresentou um cenário apelidado por Angus and Hassani-Mahmooei (2015) de "anarquia metodológica". Partindo da premissa que há grandes discrepâncias nas orientações provenientes da revisão de literatura, optou-se por uma investigação profunda e abrangente dos itens que eram passíveis de influenciar os resultados do modelo. Esse segundo passo da investigação propôs uma adaptação e detalhamento do *workflow* para análise de saída do modelo, disponíveis na literatura. A partir da proposta de Pianosi et al. (2016) propôs-se uma metodologia com três passos: 1) projeto do experimento; 2) análise de incerteza; e 3) análise de sensibilidade. A contribuição baseia-se na inserção explícita de métodos para a definição da estabilidade da variância, ou seja, o tamanho mínimo da amostra para o estudo de caso específico. Os pesquisadores divergem consideravelmente sobre qual deve ser o tamanho mínimo de uma amostra, dado um determinado número de fatores de entrada. Postula-se que a variabilidade do parâmetro de saída sob investigação deve nortear essa escolha. Apenas quando a variância atingir um ponto de estabilidade, é possível obter o número mínimo de simulações necessárias para que as conclusões sejam válidas.

Além disso, os experimentos foram projetados para investigar a eficácia e eficiência da estratégia de amostragem e do método de análise de sensibilidade. Foram avaliadas todas as possíveis combinações entre as estratégias de amostragem comuns na literatura (Monte Carlo, Hipercubo Latino, *Array* Ortogonal, Fourier, entre outros) e os métodos de sensibilidade (regressão, correlação, OAT, Sobol, Teste Delta, processos gaussianos, entre outros). Todas as possíveis combinações resultaram em uma miríade de simulações. Para executar esse grande número de testes, foi necessário implementar uma integração entre a modelo de simulação MASE-BDI e a ferramenta estatística de quantificação de incerteza PSUADE (Tong, 2005). Dessa forma, por meio de uma interface de usuário é possível determinar os fatores de entrada e saída, o tamanho da amostra e a técnica de amostragem. O sistema automaticamente informa esses parâmetros para a ferramenta e simula cada um desses cenários. Após esse cálculo, é possível selecionar os métodos de análise de incerteza e sensibilidade e calcular os respectivos índices. De forma surpreendente, métodos amplamente difundidos apresentaram resultados controversos quando aplicados no estudo de caso. Ademais, diferentes métodos de amostragem produziram diferentes saídas para o mesmo método de análise de sensibilidade. Em alguns casos, diferentes tamanhos de amostra indicaram resultados conflitantes para uma mesma métrica de sensibilidade.

A partir dessas observações é possível afirmar que nenhum MBA pode aplicar um método sem antes questioná-lo. Uma série de investigações preliminares são obrigatórias para garantir que os métodos de incerteza e sensibilidade são adequados para o estudo de caso em questão. Para tornar os experimentos mais eficientes, uma utilização integrada de análise de incerteza e sensibilidade foi a opção metodológica escolhida. Os resultados da análise de incerteza alimentavam a análise de sensibilidade, promovendo uma análise mais completa das saídas do modelo. O *workflow* proposto é a ferramenta para guiar outros

pesquisadores da área de MBA e evitar que erros comuns sejam cometidos. Um exemplo são os métodos de regressão-linear e correlação, amplamente difundidos em modelos ecológicos, mas que se mostraram inadequados para a avaliação do MBA em questão.

Na última etapa da tese, optou-se por enquadrar os experimentos em um *framework* geral para avaliação de modelos inicialmente proposto por Augusiak, Van den Brink, and Grimm (2014). "Avalidação"é a composição entre avaliação e validação que se ancora no ciclo de modelagem e propõe atividades específicas para verificar cada passo da concepção e simulação de um modelo. O foco desse trabalho concentrou-se nos métodos de verificação das saídas, análise e corroboração das saídas do modelo. Para cada item, fornece-se o passo a passo de atividades, aplicadas ao modelo MASE-BDI. Para ilustrar o potencial dessa metodologia, foram propostos dois experimentos, um exploratório e um explanatório, para gerar versões simplificadas, computacionalmente eficientes, e que exploram comportamentos específicos do sistema em questão. A simplificação baseia-se na redução da variabilidade dos fatores de entrada, de modo a aumentar a confiança nos resultados das predições. O experimento exploratório possibilitou a investigação de comportamentos extremos do sistema, mantendo a variabilidade dos fatores. O experimento explanatório reduz a variabilidade de saída. Ao refinar o fator de entrada que mais influencia o resultado, foi possível reduzir as incertezas. Ambos os experimentos mantêm a média da variável de saída constante.

O resultado é uma avaliação integral do modelo, no que concerne a variável de saída de interesse. A sequência de experimentos identificou quais os métodos mais adequados e eficientes para o estudo de caso. Entretanto, a aplicação desses métodos ilustra como deveria ser uma análise integrada de incerteza e sensibilidade em um MBA. Essa iniciativa favorece a transparência e permite o escrutínio e a replicabilidade por parte da comunidade de pesquisa. O resultado é um modelo ajustado e avaliado, cuja média registrada para função objetivo é maior que 51%, melhorando significativamente os resultados iniciais obtidos com as orientações provenientes da literatura.

Apesar de os testes terem sido realizados em um modelo específico, as considerações podem ser generalizadas para todo o campo de pesquisa. A integração de análise de incerteza e sensibilidade deve ser feita rotineiramente nos processos de avaliação de um modelo. Seguindo as etapas estabelecidas pelo *workflow*, pesquisadores podem aumentar o nível de confiança nos resultados de suas simulações e promover um uso mais racional e eficiente dos MBAs.

Palavras-chave: análise de incerteza, análise de sensibilidade, avaliação integrada, validação de modelo, modelo baseado em agentes, uso da terra

Extended Abstract

Research on land use change and land cover are essential to promote insightful management of land use to refrain environmental damage. Also, it is a complex process that relates to the interaction between environmental, economic and social systems at different temporal and spatial scales. Understanding the dynamics of these systems focuses not only on the parts but on the behavior that emerges from the interaction between them. Agent-based model (ABM) is a useful technique for studying these phenomena since ABMs model the interactions between autonomous agents and their environment. Computational simulations are the most used technique to evaluate these models, to explicitly test the effects of human decisions in complex situations.

While ABMs provide a powerful tool for analyzing emerging behavior, their usefulness is limited by difficulties in analyzing their results, which encourages criticism and questioning about the actual contribution of frameworks to decision support. The most popular tool for model evaluation is sensitivity analysis, as it quantifies the effects of the changes in the input factors of the model in the predictions of the model. However, most of the sensitivity analysis methods are not adequate or are insufficient to deal with the specificities arising from the complexity of ABMs. Among these, we highlight the stochasticity, non-linearity and the *ad hoc* parametrization of ABMs, which imply a considerable epistemic uncertainty. Without proper investigation, there are significant chances of finding results that can be a consequence of biases.

Although recognizing the particular differences of the numerous ABMs, this thesis examines whether these challenges can be overcome in the context of a case study of land use and land cover in the Cerrado of the Federal District, using the MASE-BDI multiagent tool. The objective of this thesis is to evaluate the application of several methodologies of uncertainty quantification and sensitivity analysis to analyze ABM output. We aim to perform an integrated application of uncertainty and sensitivity techniques and evaluate the impacts that differences in sample sizes, sampling techniques, and SA methods may have on model output. In addition, a workflow is proposed so that these techniques can be applied in an organized and systematic way. More broadly, a general ABM assessment methodology is applied, which includes different approaches to produce simplified versions

of the model that can be used to explore the results of the model or perform exploratory analysis.

These approaches for model analysis, calibration, and verification require a large number of repeated scenario simulation runs, with many combinations of model parameters and configurations. To facilitate this process, we implemented the integration of the MASE-BDI simulation tool with PSUADE, a set of statistical libraries for uncertainty quantification. A driver and an interface have been created to automate pre and post-processing of inputs and outputs for many models' runs.

All experiments were performed in a spatially explicit model of land use and land cover change. The simulation tool is MASE-BDI, developed at the University of Brasilia. MASE is the acronym for *MultiAgent System for Environmental simulation* that implements the *Belief-Desire-Intention* (BDI) rationality model. In BDI, agents have beliefs, a set of information about the environment they inhabit that change both perception and thinking about the world. Desires, which represent the motivational attitudes of the agents and leading them to a course of action, and, moreover, intentions, a set of plans mounted by the agent to achieve his goals. The objective function of the output analysis is a statistical metric called figure of merit (FoM), determined by the ratio between the changes in the land use that were predicted correctly over the sum of the observed changes. This metric quantifies whether the correctness of land use and land cover map is higher than the errors in predicting the amount of conversion between the different uses and land cover and the allocation of those changes in space.

The methodology used in the thesis was incremental. Initially, an evaluation of the model was performed using the most used method in the literature: one-factor-at-a-time (OAT) sensitivity analysis. We investigated four factors and sampled it within its range with a variable number of replications. To assess the quality of the output of the model, the fitness metrics were evaluated through confidence intervals. The results showed that although it is possible to differentiate between the sensitive and non-sensitive input factors, the variability of the output was so significant that the uncertainty prevented any more robust analysis. It was found that different replications of the sample affected the results considerably.

The literature review performed by Angus and Hassani-Mahmooei (2015) presented a scenario of "methodological anarchy". Based on the premise that there are major discrepancies in the guidelines found in the literature, we prosecuted an in-depth and comprehensive investigation of the items that were likely to influence the results of the model. The second step of the research proposed an adaptation and detailing of the workflow for model output analysis. Based on the framework proposed by Pianosi et al. (2016), we tailored a methodology with three necessary steps: 1) design of experiment; 2) uncertainty analysis; and 3) sensitivity analysis. The contribution is the explicit insertion of methods to define the variance stability, i.e., the minimum sample size for the specific case study. Researchers diverge considerably on what is the minimum sample size, given a number of input factors. We postulate that the variability of the output parameter under investigation should guide this choice. Only when variance reaches a stability point, we can define the minimum number of simulations necessary for the conclusions to be valid.

Besides, the experiments were designed to investigate the effectiveness and efficiency of the sampling strategy and the method of sensitivity analysis. We assessed all possible combinations of the sampling strategies shared in the literature (Monte Carlo, Latin Hypercube, Orthogonal Array, Fourier, among others) and methods of sensitivity (regression, correlation, OAT, Sobol, Delta test, Gaussian processes, among others). To test all these combinations resulted in a myriad of simulations. To perform this large number of tests, it was necessary to implement integration between the MASE-BDI simulation model and the statistical uncertainty quantification tool PSUADE (Tong, 2005). It is possible to determine input and output factors, sample size and sampling techniques through a user interface. The system automatically informs these parameters to the MASE-BDI tool and simulates each of these scenarios. After the simulation, it is possible to select the uncertainty and sensitivity analysis methods and calculate the respective indices. Surprisingly, some of the methods that are used continuously in ABM presented controversial results when applied in our case study. Also, different sampling methods produced different outputs for the same sensitivity analysis method. In some cases, different sample sizes indicated conflicting results for the same sensitivity metric.

From these observations, it was possible to affirm that no ABM can apply a method without first questioning it. Many preliminary investigations are required to ensure that the methods chosen for uncertainty and sensitivity analyzes are reliable to the particular case. To raise the computational efficiency of these tests, we applied an integrated use of uncertainty analysis and sensitivity analysis as the baseline assessment. The results of the uncertainty analysis were the input of the sensitivity analysis, promoting a complete exploration of the model outputs. The proposed workflow is a tool to guide other ABM researchers and prevent common mistakes from being made. An example is the methods of linear regression and correlation, widely diffused in ecological models but which proved inadequate for the evaluation of the ABM under study.

Finally, we chose to apply a general framework for model evaluation, initially proposed by Augusiak et al. (2014). "Evaludation" is the composition between model evaluation and validation. It is anchored in the modeling cycle and proposes specific activities to check and verify each step of the design and simulation of a model. We focused on the last three stages of the evaludation process: model output verification, model analysis,

and model output corroboration. For each item, we provide the step-by-step of activities, applied to the MASE-BDI model. To illustrate the potential of this methodology, two experiments were proposed to generate simplified, computationally efficient versions that exploit specific behaviors of the system in question: an exploratory and an explanatory experiment. The simplification is based on the reduction of the variability of the input factors to increase confidence in the prediction results. The exploratory experiment allowed the investigation of boundary behaviors of the system while maintaining the variability of the factors. The explanatory experiment reduces output variability. By refining the input factor that most influences the result it was possible to reduce the uncertainties. Both experiments maintain the mean of the output variable of interest.

The overall result is an integral evaluation of the model, regarding the output variable of interest. The sequence of experiments identified the most appropriate and efficient methods for the case study. However, the application of these methods illustrates how integrated analysis of uncertainty and sensitivity in an ABM should be. This initiative promotes transparency and allows scrutiny and replicability by the research community. The result is an adjusted and evaluated model whose average for the objective function is higher than 51%, significantly improving the initial results obtained with the literature guidelines.

Although the tests have been performed in a specific model, the considerations can be generalized for the entire field of research. The integration of uncertainty and sensitivity analysis should be done routinely in the evaluation processes of a model. Following the steps established by the workflow, researchers can increase the confidence level in the results of their simulations and promote more rational and efficient use of ABMs.

Keywords: uncertainty analysis, sensitivity analysis, integrated assessment, model validation, agent-based model, land use

Contents

List of Figures

List of Tables

List of Acronyms

ABM	Agent-based Model
ABS	Agent-based Simulation
AI	Artificial Intelligence
ANOVA	Analysis of Variance
ANU	Australian National University
BCE	Biblioteca Central
BDI	Belief-Desire-Intention
CI	Confidence Interval
CIC	Departamento de Ciência da Computação
CIRAD	Centre de Coopération Internationale en Recherche Agronomique pour le Développement
CV	Coefficient of Variation
DOE	Design of Experiments
DT	Delta Test
EE	Elementary Effects
E-FAST	Extended Fourier Amplitude Sensitivity Testing
EPA	Environmental Protected Area
FAST	Fourier Amplitude Sensitivity Testing
FoM	Figure of Merit
GE	Group Exploration

GIS	Geographic Information System
GP	Gaussian Process
GUI	Grapahical User Interface
ID	Identification
IFFD	Iterated Fractional Factorial Design
IPCC	Intergovernmental Panel on Climate Change
IE	Individual Exploration
IPA	Image Producer's Accuracy
IUA	Image User's Accuracy
JADEX	Java Agent DEvelopment framework BDI Extension
KIDS	Keep It Descriptive, Stupid
KISS	Keep It Simple, Stupid
LH	Latin Hypercube
LUCC	Land Use and Cover Change
LP	Local Perturbation
LR	Linear Regression
MARS	Multivariate Adaptive Regression Splines
MASE	Multi-agent System for Environmental Simulation
MC	Monte-Carlo
MOAT	Morris screening One-At-a-Time
OA	Orthogonal Array
OALH	Orthogonal Array-based Latin Hypercube
OAT	One-factor-At-a-Time
ODD	Overview, Design concepts, Details

PEST	Parameter Estimation Software
PSUADE	Problem Solving environment for Uncertainty Analysis and Design Exploration
PTARH	Programa de Pós-Graduação em Tecnologia Ambiental e Recursos Hídricos
QOI	Quantity of Interest
RC	Right Change
rLH	replicated Latin Hypercube
RSA	Regionalized Sensitivity Analysis
SA	Sensitivity Analysis
SAC-SMA	Sacramento Soil Moisture Accounting
SOBOL-QR	Sobol Quasi-Random
SOT	Sum-of-Trees
SPEA	Spearman Correlation Coefficient
SRC	Standardizes Regression Coefficient
TA	Transformation Agents
TG	Transformation Group Agents
TM	Total time of the simulation
TOP	Transparency and Openness Promotion
UA	Uncertainty Analysis
UnB	Universidade de Brasília
UI	User Interface
UQ	Uncertainty Quantification
WC	Wrong Change
WP	Wrong Persistence

Chapter 1

Introduction

The Earth's environment is changing at an unprecedented pace. An important area of research is the modeling of land use and land cover change (LUCC). These models try to determine what are the factors of land use change, envision when changes will happen and where, and assess how choices in public policy can influence this change. Agent-based model (ABM) is the most applied approach in LUCC research (Matthews, Gilbert, Roach, Polhill, & Gotts, 2007; Parker, Manson, Janssen, Hoffmann, & Deadman, 2003; Pontius, 2000; Pontius & Neeti, 2010; Rindfuss, Entwisle, Walsh, Li, et al., 2008; Verburg, 2006).

Agent-based modeling of social-ecological systems has been a valuable tool for understanding and supporting sustainable management of resources. ABM - in ecology also referred to as individual-based model, has become a preferred modeling tool across a wide range of fields. The main reason is that ABMs represent individual agents explicitly, and are ideally suited for including agent diversity and interactions between individual agents (Railsback & Volker, 2011). Also, it can capture the continuous changes due to the feedback of internal or external factors and can take place across different temporal and spatial scales (Schulze, Müller, Groeneveld, & Grimm, 2017).

1.1 Motivation

LUCC models require proper computational frameworks. Model simulation is the act of reproducing the behavior of a phenomenon in a computer environment (Parker, Berger, & Manson, 2001). In the last two decades, computer simulation, specifically agent-based simulation (ABS) has become indispensable in many scientific fields such as social sciences, environmental sciences, economics, and computer sciences. This intensive use of simulation is a shift in the scientific paradigm itself. Research methods usually are based on induction, the discovery of patterns in empirical data, or deduction, the specification of some axioms to prove logical consequences derived from them. G. Gilbert (1996) argues that simulation is the third alternative to science. Axelrod (2003) states that:

> Like deduction, [simulation] starts with a set of explicit assumptions. But unlike deduction, it does not prove theorems. Instead, a simulation generates data that can be analyzed inductively. Unlike typical induction, however, the simulated data comes from a rigorously specified set of rules rather than a direct measurement of the real world. While induction can be used to find patterns in data, and deduction can be used to find consequences of assumptions, simulation modeling can be used as an aid [to] intuition.

Computer modeling can be defined as the computer-aided construction of an abstraction of an observed system for a specific reason (Sterman, 1991). Thus, a computer simulation has the purpose of driving a model of a system with proper inputs and observing the corresponding outputs (Bratley, Fox, Schrage, & Schouten, 1984). There are different methodologies to build a computer model, and therefore an ABM. Each methodology attempts to provide a systematic guideline to researchers. Multi-agent systems, a paradigm from the computer science based on distributed artificial intelligence (AI), is one of the approaches that have tried to provide robust methodologies, such as Tropos (Bresciani, Perini, Giorgini, Giunchiglia, & Mylopoulos, 2004), Prometheus (Padgham & Winikoff, 2003), and Gaia (Wooldridge, Jennings, & Kinny, 2000), to guide researchers in the modeling process.

Regardless of 'how' the computer model was built, under which framework, the resulting ABS presents several features which attract multidisciplinary research teams to simulate complex and adaptive system. The idea beneath ABS is that the researcher may be able to understand the complexity of the different components not by trying to model it at the global level but analyzing emergent properties resulting from local interactions between autonomous agents and the environment. This bottom-up emergence was the new way of thinking proposed by Epstein and Axtell (1996), which allows the researcher to explain complex social phenomena from simple but dynamic representations. Today, a new approach is the pattern-oriented modelling proposed by Grimm and Railsback (2012).

The modeling process may also have different designs. Models may be conceived from a theoretical approach or a data-driven, descriptive approach. Theoretical models are abstractions that try to extract the basic mechanisms and decision points of some phenomena, usually simple enough to be used as an illustration of a specific theory or hypothesis. This modeling approach is also known as KISS, *Keep It Simple, Stupid*, which requires the modeler to make preliminary choices and to eliminate elements that seem unimportant at first (Bommel, 2017). Another alternative proposed by Edmonds and Moss (2005), is the KIDS, *Keep It Descriptive, Stupid*. The authors state that the simulated model must relate to the target phenomena in the most straight-forward way possible, taking into account the widest possible range of evidence. This methodology considers the key role of empirical data throughout the modeling stages.

Both modeling approaches have its limitations. In KISS, the modeler takes the risk to eliminate information that could be fundamental to describe the structure and dynamics of the studied system correctly. Even in simple models, there is a risk of the modeler unintentionally introduce simulation bias. In KIDS approach there is a higher probability because of the larger more substantial of data and assumptions. Also, a KIDS model may become so complicated that it is not possible to explain the results.

There are many reasons for an ABM other than prediction (Epstein, 2008). The purpose of the ABM, whether it is a simple or a descriptive approach, is a different issue to consider. There is still much debate whether ABS should be viewed as a heuristic tool to explore ideas, gain system understanding, and test hypothesis or whether they can also serve as a management and decision support tool for specific case studies (Matthews et al., 2007). Both model purposes are important and sometimes there is no clear boundary between them. However, Groeneveld et al. (2017) extensive review of land use ABMs showed that the overwhelming majority of ABMs are used for system understanding. In fact, there is a gap in their use for solving real-world problems by guiding for the design of management and policy strategies in specific case studies (Schulze et al., 2017). This lack of predictive power of ABMs is still an open challenge to be overcome.

1.2 Problem

The use of ABM is associated with some challenges that arise, such as data requirements, process uncertainty, and model validity. There is a need for available datasets to reflect the actual heterogeneity of the agents, environment, and processes that are required to make use of ABM's full power. Observation data can be scarce, and modelers will often have to resort to *ad hoc* implementation and parametrization. The parameters of these models exert a great influence on the performance of the models, and each represents assumptions regarding the modeled system. How to specify the model parameters is not a trivial problem (Duan et al., 2006). The combined effect of several factors, including errors in observed data, method options, calibration criteria, and errors in model formulation make parameter estimation difficult. This problem of over-parametrization aggravates this difficulty, as the models are progressively more complex. There is a tendency to include more and more physical layers and information, while the calibration of the models is still done with a limited amount of data (Gan et al., 2014).

The common criticism on ABM/ABS begins with the stochasticity problem, because some factors will change randomly or following some probabilities and therefore sometimes the same initial parameters result in a different output. This require exploring the model under different parameter settings. A second criticism is the subjectivity due to unclear assumptions and because of the great amount of degrees of freedom. ABMs have solid methodological foundations but researchers have a lot of freedom regarding the design of agent structure, interactions, adaptations and strategies. Another aspect of the criticism is the equifinality or identifiability problem. Multiple combinations of parameters and discrepancy function can yield the same experimental prediction (Walter, 1987). Besides, there is also the dimensionality problem. Modelers tend to include more and more layers of data and submodels just because the data is available. The computational cost increases dramatically with the number of input parameters.

Thus, data gaps, process uncertainty, and *ad hoc* parameterization entail considerable epistemic uncertainty. This raises doubts about the validity of agent-based modeling approaches, primarily, since a sharing understanding of suitable validation and calibration procedures for ABMs has not yet been established. Other aspects of ABM validation can include metamorphic validation (Olsen & Raunak, 2016), agent-based services for the validation and calibration of multi-agent models (Y. Li, Brimicombe, & Chao, 2008), or different validation methodologies Klügl (2008). Behavioral validation of ABMs, if conducted at all, has so far been restricted to the comparison of overall trends in simulation datasets. While most ABM modelers perform scenario analysis, formal uncertainty and sensitivity analysis on parameters still have rarely been used. Exceptions are the use of Monte-Carlo techniques in connection with stochastic submodels (Valbuena, Verburg, Veldkamp, Bregt, & Ligtenberg, 2010) or sampling of agent characteristics (Schreinemachers & Berger, 2011).

As a consequence, it is not surprising that a perceived lack of established formal measures for validation and calibration is one of the frequently cited problems of ABMs (Zimmermann, Heckelei, & Domínguez, 2009). Therefore, it is important to know how various parameters influence the model behavior, especially in stochastic ABM. This requires exploring the model behavior under different parameter settings. However, running a model for all possible parameter combinations is usually not practically feasible. If relationships between model parameters and output are not too complex, statistical tools may be used to gain an understanding of model behavior for various parameter settings, based on a limited number of model runs.

All these characteristics imply that quantitative analysis should be performed to test the veracity of the modeler's claims, to provide transparency and to grant some scientific rigor to the simulation results. However, a review of the application of quantitative analysis in ABM performed by Angus and Hassani-Mahmooei (2015) shows that this is still not a practice in ABM science. Richiardi, Leombruni, Saam, and Sonnessa (2006) states:

> Agent-based models have solid methodological foundations. However, the greater freedom they have granted to researchers (regarding model design) has often degenerated in a sort of anarchy (concerning design, analysis, and presentation).

Concerning the analysis, there is a general guideline (Saltelli & Annoni, 2010) that recommends that at least two activities should accompany modeling. The first is to characterize the empirical probability density function and the confidence bounds for model output, i.e., answer how uncertain is the inference. This task is also referred to as uncertainty analysis (UA). The second task is to identify factors or groups of factors most responsible for the uncertainty in the prediction, i.e., to identify where this uncertainty is coming from. This is the sensitivity analysis (SA). SA is generally recognized as a worthwhile step of analysis. However, the work of Shin, Guillaume, Croke, and Jakeman (2013) points to a standard omission on the application of this technique. Also, according to Saltelli and Annoni (2010), most of the times, researches perform a perfunctory quantitative analysis. Their review showed that rather often, modelers apply popular but proven inefficient methods of UA and SA. Although not yet widespread, UA-SA have been applied to ABMs in a few previous studies (Fonoberova, Fonoberov, & Mezić, 2013; Ligmann-Zielinska, 2013; Ligmann-Zielinska & Sun, 2010a; Parry, Topping, Kennedy, Boatman, & Murray, 2013).

1.3 Research Question

The research question we want to answer is how uncertainty quantification may be applied to improve analytical confidence in LUCC ABM outputs? To answer this question we need to investigate which UA and SA methods should a modeler use for his ABM. The existing reviews (Pianosi et al., 2016; ten Broeke, van Voorn, & Ligtenberg, 2016) are a good start, but they did not consider issues such as the empirical initialization of the agents, the limitations of data collections, the throughout empirical validation or the role of data in the calibration and validation processes.

A second question to be answered is how the quantitative analysis of a model output can help the overall validation of the model? There is much debate on the correct approaches to validate ABMs but at the end, in the words of Jain (2011), one should "not trust the results of a simulation model until they have been validated by analytical modeling or measurements."

1.4 Objective

Although recognizing the particular differences between the numerous ABMs, this work attempts to fill the gap between model output analysis and a general validation, to build an empirical guide to improve confidence in data-driven ABM. Therefore, it can be advocated that additional stages must be incorporated to the state-of-the-art reviews, to ensure reproducibility, to incorporate observation data when available, and to avoid perfunctory model analysis.

The present thesis examines whether these challenges can be overcome in the context of a case study of LUCC in the Cerrado of the Brazilian Federal District, using the MASE-BDI multi-agent tool. It discusses different approaches to model validation, calibration, and uncertainty analysis to deal with the uncertainty involved using *ad hoc* parametrization, especially in the initialization of the ABS. As these approaches require large numbers of simulation run, it presents the integration of the MASE-BDI simulation framework to a set of statistical libraries for uncertainty quantification, to automate the pre-and post-processing of MASE-BDI model' inputs and outputs. Another contribution is the final outputs itself. A verified and statistical sound prediction for the land use of the Brazilian Cerrado.

The objective of this thesis is to evaluate the application of several methodologies of uncertainty quantification in the analysis of results of ABMs. Specifically, to perform an integrated application of UA and SA techniques and evaluate the impacts that differences in sample sizes, sampling techniques, and SA methods may have on model output. The accomplishment of this will result on three main contributions:

- The proposal of an empirical workflow of uncertainty quantification to perform model output analysis, adherent to a evaluation/validation model framework;
- To evaluate the impacts that differences in sample sizes, sampling techniques, and sensitivity analysis methods may have on model output;
- Apply those recommendations in the MASE-BDI case study, in a general experimental ABM assessment, based on observation, hypothesis testing an reproducibility to produce more transparent, reproducible, and statistical sound ABM results.

The exploration of the model gives us a better understanding of the model significance. Another contribution of this work is a series of scientific publications produced during this Ph.D. The references are detailed in Appendix A.

1.5 Thesis outline

The document is structured in a way that there is not a specific chapter of state of the art. Each chapter tackles a review of the main research works concerning the scope of the subject under assessment. Thus, each chapter is somehow self-contained.

The development of the thesis takes place in three stages: the literature review and an initial investigation, followed by an extensive comparison of methods and approaches, included in a flow to facilitate the application, and finally a general evaluation of the model, applying best practices to generate simplified, robust, and statistically reliable versions of the model.

Chapter 2 presents an overview of the literature on uncertainty assessment in ABS. The MASE-BDI ABM is introduced, and the parameters used in the initialization of the simulation were eligible as a case study. We perform initial experimentation: an exploratory study was performed based on the One-Factor-At-a-Time (OAT) method, which is widely used in analyses of ABMs results. The results show that even the most popular practice in the literature may be inadequate for all ABMs. It is evident that the method and sample size affects the model analysis. The exploratory experiment demonstrated that the results have great uncertainty and that the predictions of the simulation were not reliable.

In Chapter 3, more information about the MASE-BDI ABM is provided. We propose an experimental design to search for the best methods to be applied in ABMs when comparing different UA and SA techniques for efficiency and effectiveness. A baseline scenario was established and derived from several lines of research around three main issues: i) impact of different sampling methods; ii) impact of different sample sizes; iii) impact of different SA methods, besides the verification of the convergence between different experiments. To execute this large number of simulations, we implemented the integration of the MASE-BDI framework with PSUADE statistical calculations tool. We discuss the discrepancy found in the literature and compare it with our results. We postulate that the minimum sample size should be at least equal to the stability point of the variance. Finally, we propose a workflow to perform model analysis, organizing and detailing the activities systematically.

Chapter 4 applies the concepts of the previous chapters in the form of UA-SA integrated output assessment and develops a simplified and more computationally efficient version of an ABM. Two simplifications are proposed: exploration and explanation. Exploratory experiments make it possible to investigate the extreme behavior of the system, maintaining the variability of the factors. Explanatory experiments reduce output variability. In the next step, we chose to integrate these analytical experiments with a validation structure of the model as a whole, in an "evaludation" process (evaluation +

validation). The steps for verifying the outputs of the model, analyzing the model and corroborating the results are detailed and exemplified step by step to provide a guide for similar work. This initiative promotes transparency and allows scrutiny and replicability by the research community.

Chapter 5 discusses how the application of a UA and SA integrated assessment, organized within a workflow and viewed under a macro prism of evaluation of the modeling process, can increase the reliability and usefulness of ABMs. The lack of specific methodologies for ABMs is one of the reasons that affect reliability in the results predicted by these models. In the end, we evaluate the strengths and limitations of existing SA methods. It should be remembered that SA can have several goals. In the context of this thesis, SA methods are designed to evaluate which parameters produce greater uncertainty in the model result. Thus, it is necessary to limit the number of factors studied.

Chapter 2

Uncertainty Assessment in Agent-Based Simulation: An Exploratory Study

Book Chapter published in:
Sukthankar G., Rodriguez-Aguilar J. (eds) Autonomous Agents and Multiagent Systems. AAMAS 2017. Lecture Notes in Computer Science, vol 10642. Springer, Cham.

2.1 Introduction

LUCC investigation is of importance to promote insightful management of Earth's land use to refrain environmental damage. Moreover, LUCC is a complex process that relates to the interaction between environmental, economic and social systems at different temporal and spatial scales. Computational frameworks are the most used techniques to simulate LUCC models for its ability to cope with its complexity.

ABM has been incorporated into LUCC models, and many other real-world problems, to explicitly simulate the effects of human decisions in complex situations. They are based on the multiagent system paradigm that features autonomous entities that interact and communicate in a shared environment. These entities perceive the environment, reason about it and act on it to achieve an internal objective. Therefore, ABM can capture emergent phenomena and provide an original description of the modeled system.

The Multi-Agent System for Environmental simulation (MASE) is a freeware software developed at the University of Brasilia. MASE-BDI is an extension of MASE for exploring potential impacts of land use policies that implement a land use ABM (Ralha & Abreu, 2017). Considering the purpose and reliance upon external data, MASE-BDI may be characterized as a predictor-type ABS (Heath, Hill, & Ciarallo, 2009): a data-driven model with the overall goal of performing medium to long-term predictions. MASE-BDI simulations were calibrated to match available GIS data (Coelho et al., 2016). Simulation results were validated according to a standard methodology for spatially explicit simulations (Pontius et al., 2008) and then compared to similar frameworks (Ralha et al., 2013). MASE-BDI performance was found to be higher than other 13 LUCC modeling applications with nine different traditional peer-reviewed LUCC models according to Pontius et al. (2008). Despite this fact, the lack of uncertainty assessment and sound experimentation is the main reason for criticism and questioning about the real contribution of frameworks to decision support for LUCC.

According to Bommel (2017), any ABS has levels of uncertainty and errors associated with it. ABS continues to harbor subjectivity and hence degrees of freedom in the structure and intensity of agent's interactions, learning, and adaptation (Lee et al., 2015). There are significant chances of finding results which may be the consequence of biases. Furthermore, almost every ABS review have expressed the need for statistical methods to validate models and evaluate the results to improve the transparency, replicability and general confidence in results derived from ABS. These problems continue to be underestimated and often neglected. Some authors such as Heath et al. (2009), likewise, argued that validation is one of the most critical aspects of a model building because it is the only means that provides some evidence that a model can be used for a particular purpose. However, at least 65% of the models in their survey were incompletely validated. Of the models validated in some way, surprisingly less than 5% used statistical validation techniques. Traditionally, ABS types of systems are difficult to analyze given their non-linear behavior and size (Casti, 1995).

Treatment of uncertainty is particularly important and usually difficult to deal with in the case of ABM's stochastic models. While acknowledging the differences in data sources and the causes of inconsistencies, there is still the need to develop methods to optimally extract information from the data, to document the uncertainties and to assess common methodological challenges. To look away could reinforce inconsistent results and damage the integrity and quality of simulation results.

This work aims to discuss how uncertainty is being portrayed in ABS and to perform an exploratory study to use statistical methods to estimate uncertainty in an LUCC agent-based prediction simulation tool. The MASE-BDI system will be the simulator under study. The Cerrado case study simulations (Ralha et al., 2013) will be the basis for the analysis. As a first investigation step, we assessed the uncertainty within the inputs and configuration parameters of the simulation. Our final goal would be to document, to quantify and to foresee its propagation impacts in the results. A particular challenge in performing measurements is coming up with appropriate metrics. The thorough experimentation and repeatability would, therefore, improve our understanding of the uncertainty and relations among the variables that characterize a simulation. The remainder of the paper is structured as follows. In Section 2.2, we present some background on uncertainty and in Section 2.3 some related work. In Section 2.4, we summarize the MASE-BDI characteristics and case study. We also present the methodology for the exploratory study. In Section 2.5 we show results together with discussions. In Section 2.6 we conclude with a summary.

2.2 Overview of uncertainty in ABS

The relevance of the treatment of uncertainty is dependent on the modeling objective. Requirements regarding model uncertainty may be less critical for social learning models, where communication and interaction among stakeholders would be of more significance. Conversely, parameters, measurements, and conditions used for model runs influence much more data-based predictions of future states. Projection, forecasting and prediction models are usually very affected by the variation of a system output from observed models.

Also, there are different sources of uncertainty that can influence the prediction of a simulation model. It can arise from simulation variability in stochastic simulation models or from structural uncertainty within assumptions of a model. We will emphasize input uncertainty, what McKay, Morrison, and Upton (1999a) defined as incomplete knowledge of "correct" values of model inputs, including model parameters. If the inputs of a model are uncertain, there is an inherent variability associated with the output of that model. Therefore it is crucial to communicate it effectively to stakeholders and technical audiences when outputting model predictions.

Uncertainty in environmental prediction simulations may limit the reliability of predicted changes. This issue is one of the recurrent conclusions of the Intergovernmental Panel on Climate Change (IPCC). Back at 1995, IPCC stated that "uncertainties in the simulation of changes in the physical properties have a significant impact on confidence in projections of future regional climate change"(Houghton et al., 1996) and that was necessary to reduce uncertainties to increase future model capabilities and improve climate change estimates. Since 2010, IPCC dedicates an integral feature of its reports to the communication of the degree of certainty within IPCC assessment findings (Mastrandrea et al., 2010). In the most recent report, IPCC assesses a substantially larger knowledge base of scientific, technical and socio-economic literature to reduce uncertainty and uses a large number of methods and formalization (IPCC, 2014). Especially for future predictions, validating a model's predictive accuracy is not straightforward due to a lack of appropriate data and methods for "validation" (Kelly (Letcher) et al., 2013). That is another reason why applications, frameworks, and methods of formalization in this research area are relevant and should be promoted.

Regarding the type of modeling, there are approaches such as Bayesian networks, able to explicitly deal with uncertainty in the interpretation of data, measurements or conditions. In contrast, other approaches such as ABMs require the development of comprehensive or compelling analysis of output data and a lot of resource-intensive attention (Lee et al., 2015). The level of testing required to develop this understanding is rarely carried out, mainly due to time and other resource constraints (Kelly (Letcher) et al., 2013).

Indeed, uncertainty assessment in ABM can be a hard task for even relatively small models. Due to their inherent complexity, ABS is often perceived as a "black box", where there is no purpose in explaining why the agents acted as they did, as long as the modeler presents some form of validation (i.e., shows a good fit). According to Marks (2007), ABMs simulations can prove existence, but not in general necessity. Despite that, there is a research effort to make ABS more transparent and to demonstrate that the simulations behave as intended through efforts in standardization in simulation model analysis and result sharing (Lorscheid, Heine, & Meyer, 2012). Besides from verification, uncertainty assessment aims to increase understanding, to improve the reliability of the predicted changes and to inform the degree of certainty of critical findings. To achieve this effort, some techniques and methods such as uncertainty and sensitivity analysis should be part of the modeling process.

Uncertainty Quantification (UQ) is defined as the identification, characterization, propagation, analysis, and reduction of uncertainties. Sensitivity analysis (SA) is defined as the study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input (Saltelli et al., 2008) and is a method to assess propagation of uncertainties. SA responds to the question of which inputs are responsible for the variability of outputs. Local SA explores the output changes by varying one parameter at a time, keeping all the others constant. Although it is a useful and straightforward approach, it may be location dependent. Global SA gives a better estimate of uncertainty by varying all parameters at the same time by using probability density functions to express the uncertainty of model parameters. Uncertainty analysis is a related broader uncertainty propagation practice to SA. It focuses instead on quantifying uncertainty in model output, addressing the variability of results. Ideally, uncertainty and SA should be run in tandem.

2.3 Related work

There are a growing number of attempts to assess uncertainty in ABS. However, there is a lack of specific guidance on effective presentation and analysis of the simulation output data. There is a variety of approaches to quantifying or reducing uncertainty. The work of Lee et al. (2015) offers an overview of the state-of-the-art methods in the social simulation area, in particular examining the issues around variance stability, SA, and spatiotemporal analysis. Because of our interest in LUCC simulations, we chose to review how those approaches are being applied and communicated on spatially-explicit simulations.

In Albrecht and Ramamoorthy (2015), the authors propose an algorithm as an alternative to goodness-of-fit traditional validation to answer if the agents in a simulation are behaving as expected. To them, the key to effective interaction in multi-agent applications is to reason explicitly about the behavior of other agents, in the form of a hypothesized behavior. This approach would allow an agent to contemplate the correctness of a hypothesis. In the form of a frequentist hypothesis test, the algorithm allows for multiple metrics in the construction of the test statistic and learns its distribution during the interaction process. It is an interesting approach to addressing the uncertainties within the model and agents behavior. We believe it would be even more useful if coupled with an uncertainty quantification technique.

The work of Paegelow, Camacho Olmedo, Mas, and Houet (2014) assesses uncertainty that is characteristic of spatially explicit models and simulations. The authors propose a benchmarking scheme of LUCC modeling tools by various validation techniques and error analysis. The authors investigate LUCC tools that are based on map comparisons to analyze the accuracy of LUCC models concerning quantity, pixel by pixel correctness and LUCC components such as persistence and change. Also, they investigated the map outputs of these simulations to test the fidelity of spatial patterns and the congruency of the simulation maps from different modeling tools. Although the variability of LUCC models does not allow strict comparisons, there is still room for improvements in methodologies, validation and uncertainty quantification.

The work of Gan et al. (2014) assesses model output analysis through a global SA, a commonly used approach for identifying critical parameters that dominate model behaviors. They use the Problem Solving environment for Uncertainty Analysis and Design Exploration (PSUADE) software, to evaluate the effectiveness and efficiency of widely used qualitative and quantitative SA methods. Each method is tested using a variety of sampling techniques to screen out the most relevant parameters from the insensitive ones. The Sacramento Soil Moisture Accounting (SAC-SMA) model, which has thirteen tunable parameters, is used for illustration. The South Branch Potomac River basin near Springfield, West Virginia in the U.S. is chosen as the study area. The authors show how different sampling methods and SA measurements can indicate different sensitive and insensitive parameters and that a comprehensive SA is paramount to avoid misleading results.

The work of J. D. Li et al. (2013) also performed a global SA to show which model parameters are critical to the performance of land surface models. The authors considered forty adjustable parameters in The Common Land Model and therefore compare different SA methods and sampling. The size of each sample would vary as well. The sampling techniques and SA measures that were considered optimal were distinct from the results found by Gan et al. (2014), meaning that not all LUCC ABS propagate uncertainty the same way.

Gao and Hailu (2012) integrated a recreational fishing ABM model with fuzzy logic to incorporate uncertainties over the preferences of outcomes or criteria. Although this work assesses the treatment of uncertainty in ABMs, the solution is based on a function that can be used to convert observed/simulated outcomes to qualitative measurements that reflect uncertainty regarding the outcomes.

Another approach was performed by Le, Seidl, and Scholz (2012), also in an LUCC model. They use the method of independent replication. In the case study, the authors replicated the simulation 12 times for each mechanism and computed the mean values of the impact indicators and the confidence interval (CI) at the reliability of 95%. They used uncertainty quantification to define a minimum certainty threshold in the simulation outputs.

Schreinemachers and Berger (2011) proposed the Monte Carlo initialization of the agents of the simulation that generates many possible and statistically consistent agent populations that are used for repetitions of simulation experiments. The authors tested the sensitivity of the LUCC simulation outcomes for crucial policy indicators. The variation of these indicators was measured by standard deviations (expressed as percentages of the normalized mean) in fifty different agent populations for the baseline scenario.

New interesting frameworks are being created to support SA in ABS. Herd, Miles, McBurney, and Luck (2015) work focus on the applicability of formal verification methods such as statistical testing of large-scale ABS. They created MC2MABS, a Monte Carlo Model Checker for MultiAgent-Based Simulations which incorporates the idea of statistical runtime verification, a combination of runtime verification and statistical model checking. The framework can provide conventional model checking for probabilistic systems by the use of a sampling approach and the employ of statistical techniques to generalize the results to the overall state space. Runtime verification focuses on the execution trace of a system, using temporal logic and checking automatically.

All these authors used several indicators to measure the variability of model results based on changing input parameters. Table 2.1 illustrates a brief comparison among those works. MASE-BDI exploratory uncertainty assessment will be described in the next sections. A large panel of statistical tools exist to help with the accuracy of the predictions such as Dakota[1], PSUADE (Tong, 2015), UQ-PyL[2] and MEME Suite[3]. There are initiatives to apply the potential of classic Design of Experiments (DOE) for ABS (Kleijnen, Sanchez, Lucas, & Cioppa, 2005; Lorscheid et al., 2012). ABS field of research would benefit from systematic empirical research with standardized procedures, but ABS idiosyncrasies in model output turn the task even harder. Researchers so far failed to reach consensus and to determine sound methodological guidelines. Therefore, the studies are still mostly investigative and exploratory.

Table 2.1: Overview of the general characteristics of each related work

Reference	**Model**	**Uncertainty Methods**
Albrecht et al. (2015)	Generic ABS	Correctness Hypothesis test and runtime statistical verification in the agent's behavior
Paegelow et al. (2014)	Land use models	Image statistical comparison of pixel/maps and error analysis to find uncertainty drivers
Gan et al. (2014)	SAC-SMA hydrological model	Global SA with 15 sampling 9 different sample sizes and 12 SA methods
J. D. Li et al. (2013)	Land surface model	Local SA and 4 Global SA methods with 3 sampling techniques, and 6 sample sizes
Le et al. (2012)	LUDAS: land use ABS	Independent Replications and Confidence Intervals to assess output variation
Ralha et al. (2013)	MASE-BDI: land use ABS	Global SA with different sample configurations, independent replications, and Confidence Intervals

[1] https://dakota.sandia.gov/
[2] http://www.uq-pyl.com/
[3] http://meme-suite.org/

2.4 MASE-BDI exploratory study

The MASE Project[4] objective is to define and implement a multi-agent tool for simulating environmental change. MASE-BDI enables modeling and simulations of LUCC dynamics using a configurable user model. The multi-agent architecture is composed of three hierarchical layers (from top to bottom) (Ralha et al., 2013): a User Interface (UI), a Pre-processing and an Agent layer. In the agent layer, there are cell agents representing land units hosting natural processes, such as crop/forest grow, and there are transformation agents, representing human agents and their behavior as farmers or cattle rancher.

The Cerrado-LUCC model of MASE-BDI is used as a test problem. The simulations depict the land use and cover changes of the most endangered biome in Brazil. The Cerrado is the second largest biome in South America and harbors significant endemism and biodiversity. The landscape has been undergoing severe transformation due to the advance of cattle ranching and soy production. The Cerrado-LUCC simulation model was documented and described employing the standard ODD-protocol (Overview, Design concepts, and Details) (Grimm et al., 2006, 2010) to promote transparency and replicability. We also applied empirically grounding ABM mechanisms for the characterization of agent behaviors and attributes in socio-ecological systems (Smajgl, Brown, Valbuena, & Huigen, 2011). In this article, we provide some core information about MASE-BDI and the Cerrado-LUCC Model, mainly about the parameters and outputs. Readers who are interested in the details of this model and the implementation of MASE-BDI multi-agent system should refer to Ralha et al. (2013) and Ralha and Abreu (2017), respectively.

The input of the simulation is a couple of grid raster maps consisting of the land cover of the region, from two different time periods (a reference map of the initial time t_0 and a reference map of a subsequent time t_1). Also, each simulation carries a set of maps to describe the physical characteristics of the environment, such as water courses, water bodies, slope, buildings, highways, environmental protected areas, and territorial zoning maps.

The simulations are calibrated from the two time-steps and project the land use and cover change for future steps. The result of a MASE-BDI simulation is a couple of predicted maps (Figure 2.1), with the allocation of change and a set of metrics calculated during runtime. The resulting image is submitted to a goodness-of-fit measurement, and the quality and errors of the quantity of change and allocation of land use change are calculated.

[4]Software Availability: `http://mase.cic.unb.br/`

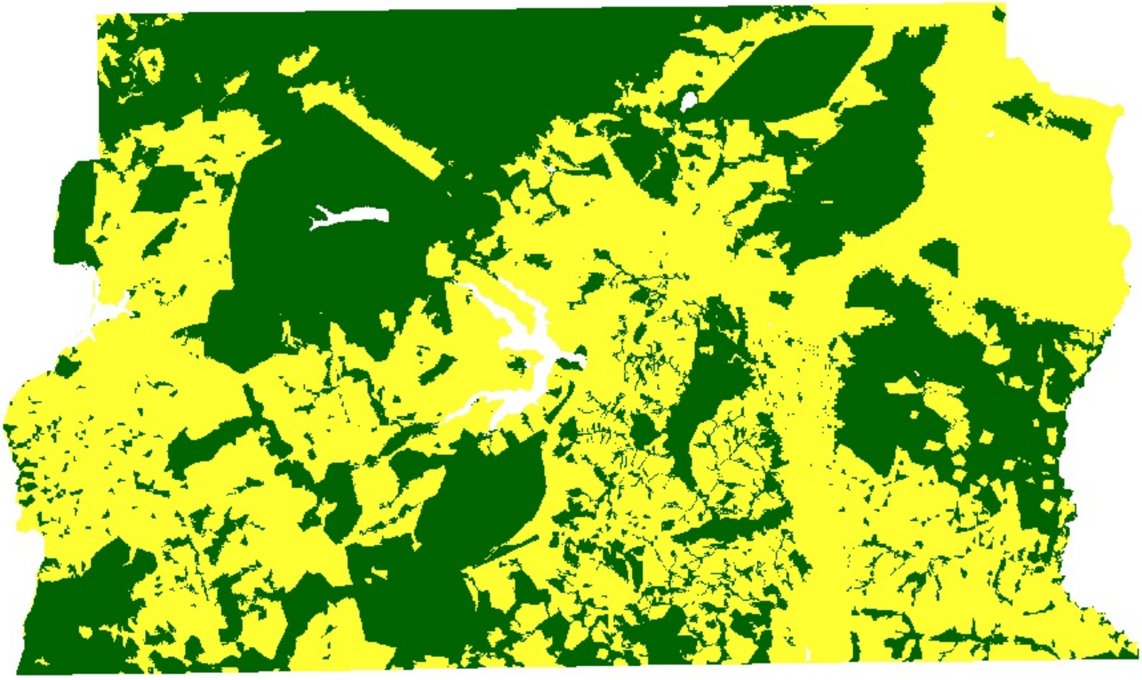

Figure 2.1: A land cover predicted map of the Cerrado in Federal District, Brazil

Methodology

The objective is to perform exploratory analysis, based on classical statistics, to reduce uncertainty and to understand how the model behaves. MASE-BDI LUCC model is under input uncertainty investigation, to calculate their influence in the simulation output. For exploratory purposes, we want insight on the parameters that affect the multi-agent system implementation, so we selected a subset of Cerrado-LUCC model inputs for this demonstration. The subset of input parameters of the multi-agent system is displayed in Table 2.2: TA-Number of Transformation Agents, TG- Number of Group Transformation Agents, IE- Potential of Individual Exploration and GE- Potential of Group Exploration. These parameters characterize the instantiation of MASE-BDI agents and therefore, should be analyzed regarding uncertainty. For the sake of clarity, a brief note on the terminology of the word input. We are aware that the ODD protocol (Grimm et al., 2010) classifies input as an amount of data that is added during a simulation. The word input has a more general use in this manuscript. We use the words input, parameter and factor to describe any entry of the model, such as a submodel, or an initialization configuration. The MASE-BDI input configuration parameters are the initial conditions to start a simulation.

Table 2.2: MASE-BDI multi-agent input configuration parameters

ID	Parameter	Description	Range
I1	TA	Number of Transformation Agents	[1, 100]
I2	TG	Number of Group Transformation Agents	[10, 100]
I3	IE	Potential of Individual Exploration	[1, 500]
I4	GE	Potential of Group Exploration	[1, 1500]

The number of transformation agents is a parameter that reflects the number of computational agents (in the multi-agent system paradigm) instantiated in a simulation run. In this study case, one agent does not represent one single individual. The Cerrado-LUCC model was formulated based on an empirical characterization of agent behaviors, proposed by Smajgl et al. (2011), with two necessary steps: the development of behavioral categories and the scaling to the whole population of agents. TA was derived from the Brazilian Agricultural Census of 2006 and comprises a set of Producer legal status. The range of 1 to 100 is an abstraction to the 3407 register producers in the region that may be active or inactive in a given period. The details of this agent characterization are thoroughly illustrated in Ralha et al. (2013). Likewise, a particular type of agent is TG, which represent not an individual but an organization, cooperative, business or so. The range is an abstraction of the 548 group producers, 10 of which have permanent exploration licenses. All the explanation of this parameters are described within the ODD protocol (Grimm et al., 2010) in the work of Ralha et al. (2013).

The potential of exploration, individual or of a group, represent the impact an agent can produce in the natural vegetation cover of a cell during a step. In the Cerrado LUCC Model, considering the deforestation process, the potential of exploration is again an abstraction for the amount of m^3 of wood that can be obtained from a particular grid cell, until a theoretical limit that represents resource depletion.

In addition to the final LUCC maps, the simulation generates a set of metrics as results, mainly spatial analysis measurements, which includes pixel by pixel comparison, a quantitative and an allocation agreement. Those measurements are specific statistical LUCC indices to determine the produced map accuracy, proposed by Pontius et al. (2008). It includes an objective function called the figure of merit (FoM), a ratio between correct predicted changes and the sum of observed and predicted changes. To evaluate the response of the model to the different parameters, the experiments considered the outputs described in Table 2.3 and tried to identify and quantify the influence of the simulation input configurations on the model outputs. The identification (ID) of each of the outputs follows the numbering of its generation in the file *.csv* produced by MASE-BDI at the end of each simulation.

Table 2.3: MASE-BDI output parameters

ID	Output	Description
O1	TM	Total time of the simulation
O4	FoM	Figure of Merit
O5	IPA	Image Producer's Accuracy
O6	IUA	Image User's Accuracy
O7	WC	*Pixel's Wrong Change*: observed change predicted as persistence
O8	RC	*Pixel's Right Change*: observed change predicted as change
O9	WP	*Pixel's Wrong Persistence*: observed persistence predicted as change

To identify and analyze these uncertainties we performed a method of elementary effects (EE) of global SA on the MASE-BDI LUCC model. For this calculation, we used the software package developed by Tong (2015) called PSUADE, containing various methods for parameter study, numerical optimization, uncertainty analysis and SA.

Screening methods are based on a discretization of the inputs in levels, allowing a fast exploration of the system behavior (Iooss & Lemaître, 2015). This type of method aims to identify the non-influential inputs with a small number of model calls. The most used screening method is based on the one-parameter-at-a-time (OAT) design, where each input is varied while fixing the others. The simplicity is one of OAT's advantages, but there are drawbacks when applying to ABM. For one, it does not consider parameter interactions and may cover a slight fraction of the input space. Nevertheless, OAT is still one of the most applied SA technique in ABMs.

The EE method we chose to apply is the Morris method (MOAT) proposed by Morris (1991) and refined by Campolongo and Braddock (1999), an expansion of the OAT approach that forsakes the strict OAT baseline. It means that a change in one input is maintained when examining a switch to the next input and the parameter set is multiply repeated while randomly selecting the initial parameters settings. EE is suited for spatially explicit simulations, usually computationally expensive models with large input sets.

MOAT allows classifying the inputs into three groups: inputs having a negligible effect, inputs having substantial linear effects without interactions and inputs having significant non-linear and interaction effects. In overall effect and interaction effect of each parameter can be approximated by the mean μ and standard deviation σ of the gradients of each parameter sampled from r, the number of replications.

The MOAT sampling technique was designed for the particular MOAT method. The work of Gan et al. (2014) details how the MOAT sampling works: the range of each parameter is partitioned into $p-1$ equal intervals. Thus the parameter space is an n-

dimension p-level orthogonal grid, where each parameter can take on values from these p determined values.

First, r points are randomly generated from the orthogonal grid; and then, for each of the r points, other sample points are generated by perturbing one dimension at a time. Therefore, sample size will be $(n+1) \cdot r$. For the sampling size, Levy and Steinberg (2010) report that one needs at least $10 \cdot n$ samples to identify key factors among the parameters.

To avoid the effect size on the sample, we determined a minimum sample size of $800 (= 20 \cdot 4)$, for four inputs. For MOAT sampling we used 160 replications, resulting in sample size of 800 $(= (4 + 1) \cdot 160)$.

Moreover, as in other stochastic models, it is not advisable to conclude from a single MASE-BDI simulation run. For an initial uncertainty assessment, we applied the method of independent replications proposed by Goldsman and Tokol (2000). We run the model approximately eighty-five thousand times (an arbitrary choice to explore all the input parameter space) and randomly clustered the results into five independent replication groups. We computed the mean values of the outputs and their CIs at the reliability of 95%. Another approach to estimating the uncertainty of the model output is to study the variance in the model outputs by using the Coefficient of Variation (CV) (the ratio of the standard deviation σ of a sample to its mean μ), to compare the variance of different frequency distributions.

2.5 Results

In the current work, we analyzed four input parameters, displayed in Table 2.2, regarding the multi-agent configuration of MASE-BDI LUCC model. First, we present the results of the SA. Figure 2.2 presents the EE of CERRADO-LUCC model parameters. Figure 2.2 (left) illustrates the modified means of MOAT gradients and also their spreads based on Monte Carlo bootstrapping. The results show that GE and TA are the most sensitive parameters in term of having the largest average median (26.466 and 25.205, respectively). The other two parameters have median sensitivities close to zero, denoting the impact of these parameters on the simulation output is minimal.

Figure 2.2 (right) is a MOAT diagram that shows a consensus view among mean μ and standard deviation σ of the gradients of each parameter sampled from r. The more sensitive the parameter, the closer it is to the upper right corner of the graph. These results show a positive correlation between input and output uncertainties. Since GE and TA describe the amount of land transformation in a simulation, high values of these parameters will increase the model output. GE is the most sensitive parameter, followed

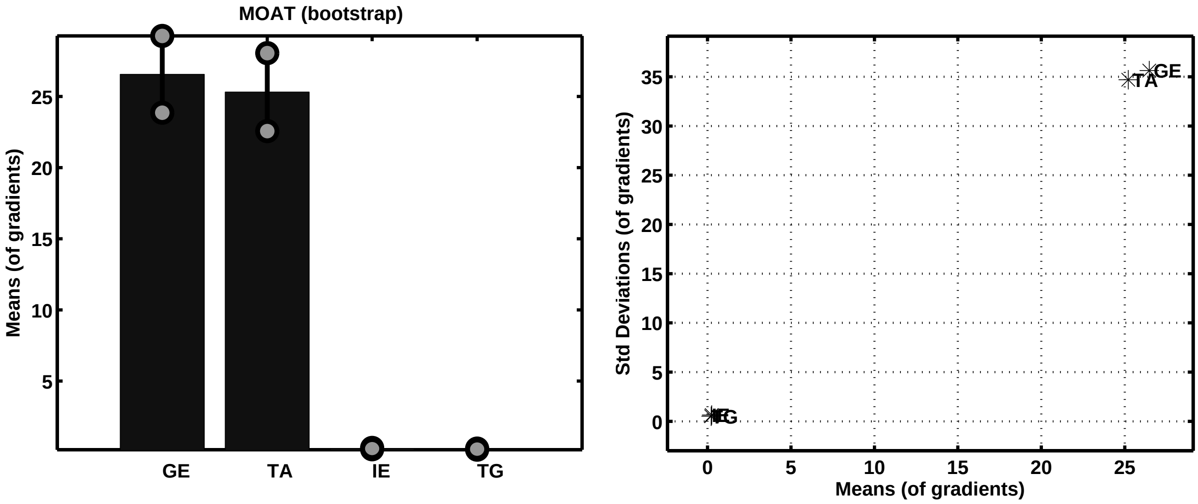

Figure 2.2: Parameter sensitivity rankings of MOAT method

by TA. To understand and to reduce uncertainty within this two variables will, therefore, reduce the uncertainty of the simulation as a whole.

GE represents the amount of land cover that is transformed by a group of human agents in a cell of the map. GE is a sensitive value for it indicates the voracity and velocity of the current land exploitation, what will directly affect the result of the simulation. GE is probably sensitive because the socio-economic groups responsible for large-scale cattle ranching and permanent agriculture are the principal driver of deforestation in Cerrado. Their rates of land change are more significant than the number of groups, what explain TG as an insensitive parameter to the output. As for TA, the more agents one instantiates in a simulation, the more land cover will be affected, the higher will be the land use transformation rates. Conversely, the potential of exploration of a single individual is less determinant than the number of single individuals acting on the land, with SA indicating TA a sensitive and IE as an insensitive parameter.

To investigate MOAT sensitivity results, we used different replications times r and different levels p to know for sure the relevance of the parameters as displayed in Fig. 2.3. It is possible to see that even within the same method, results may vary. The results for four replications are not very consistent with the other replication results, mainly with the mean. The results with $r = 56$, $r = 108$ and $r = 160$ present minor variations. We can infer that four replications are not enough to identify the parameters sensitivity in the MASE-BDI model successfully and therefore the number of replications should be higher to be effective.

Table 2.4 is a summary of the Basic Output Statistics of the MASE-BDI LUCC model. Each replication is assigned by $i = [1..5]$, the sample mean from the coefficient variation by CV_i, and the mean of all replications by $\bar{Z}$. We performed independent replications to

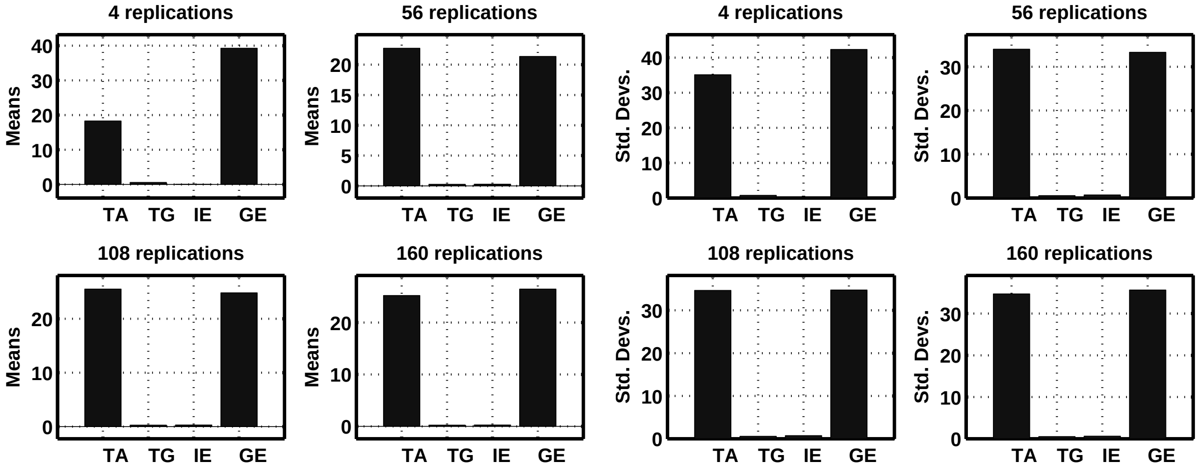

Figure 2.3: Sensitivity of parameters at different replication times r

verify the variation of the indicators, and for an initial analysis, we consider this variation as noise (uncertainty). Any impact conclusions in predictions can only be drawn if the changes in standards are greater than the uncertainty rate. Therefore, we have a first threshold to define if some result is valid, compared to the simulations behavior.

We also estimated the expected average FoM for simulations, using the five replication grouped results ($b = 5$). Considering the $\bar{Z}_{FoM} = 43.87$ and the estimated Variance $\hat{V}_R = 100.99$, we have an approximately $100(1-\alpha)\%$ two-sided CI for θ, according to the formalization proposed by Goldsman and Tokol (2000). For level $\alpha = 0.05$, we have $t_{0.025,4} = 2.78$, and gives $[31.39, 56.34]$ as a 95% CI for the expected FoM for MASE-BDI simulations.

Table 2.4: Coefficient of variation for MASE-BDI outputs

Output	CV_1	CV_2	CV_3	CV_4	CV_5	$\bar{Z}$
Time	0.300	0.130	0.250	0.260	0.200	**0.230**
Figure of Merit	0.015	0.011	0.008	0.007	0.090	**0.100**
Producer's Accuracy	0.015	0.011	0.008	0.007	0.009	**0.010**
User's Accuracy	0.006	0.005	0.004	0.004	0.003	**0.004**
Wrong Change	0.030	0.030	0.030	0.030	0.020	**0.030**
Wrong Persistance	0.007	0.007	0.008	0.008	0.013	**0.009**
Right Change	0.015	0.011	0.008	0.008	0.009	**0.010**

2.6 Conclusions

In this study, we first identified the most sensitive parameters for the MASE-BDI LUCC model using MOAT SA. We investigated some proper sampling design and sample size needed for MOAT screening the parameters effectively. Although these conclusions are model-specific, it corroborates possible variation among sampling techniques and SA methods.

This paper is the first exploratory study towards quantifying uncertainty within MASE-BDI simulations. Following experiments must be done to promote more standardization to this effort through the application of Design of Experiments. We look forward to investigating further on the model parameters, analyzing the remaining inputs besides the agent's quantities and their impacts.

The presented results allow us to understand the uncertainty when defining the parameters of the simulation of the LUCC model under study. Our feeling is that the uncertainty is very high which means that either model need to improve dramatically or LUCC policy need to be reevaluated. Most simulation tools fail to validate models and to state the uncertainty in simulation results. Consequently, policymakers and the general public develop opinions based on misleading research that fails to give them the appropriate interpretations required to make informed decisions. The efforts to assess ABMs through statistical methods are paramount to corroborate and improve the level of confidence of the research that has been made in LUCC simulation.

Chapter 3

An empirical workflow to integrate uncertainty and sensitivity analysis to evaluate agent-based simulation outputs

Full article published in Journal Environmental Modelling & Software, v. 107, p. 281-297, 2018.

3.1 Introduction

As cited in the literature, LUCC systems are dynamic, stochastic, and characterized by nonlinear and non-monotonic relationships between constant changing entities (Parker et al., 2003; Rindfuss, Entwisle, Walsh, An, et al., 2008; Verburg, 2006). Besides, ABMs have been used as a natural metaphor to model LUCC dynamics, since they capture emergent phenomena and provide an original description of the modeled system (Murray-Rust, Rieser, Robinson, Miličič, & Rounsevell, 2013; Ralha et al., 2013; Schreinemachers & Berger, 2011). However, ABMs are prone to uncertainty because they reflect the intrinsic randomness of environmental, physical, and social events. The uncertainty may also arise because of insufficient knowledge, lack of data, observation errors, measurements used to parametrize the model, or from vague premises of the model (Ligmann-Zielinska, Kramer, Cheruvelil, & Soranno, 2014; Lilburne & Tarantola, 2009). As a result, one could argue whether there is any quality in model predictions due to high uncertainty and the considerable number of assumptions imposed by ABMs models.

In this scenario, UA and SA are currently popular topics in ABMs as well as for many other complex systems (Pappenberger, Beven, Ratto, & Matgen, 2008). They are valuable tools in understanding LUCC models and deriving decisions on strategies to reduce model uncertainty. UA provides the variability of model results. SA presents which factors are responsible for this variability. This variability may be expressed quantitatively in terms of "elasticity" of performance concerning parameter levels. High sensitivities (elasticities) give cause for concern about the reliability of a model (Dayananda, Irons, Harrison, Herbohn, & Rowland, 2002). A factor is any source of uncertainty in the modeling process, including model structure, initial conditions, and input parameters. Using the terminology proposed by the National Research Council (2012), uncertainty quantification (UQ) is the process of quantifying uncertainties in a computed quantity of interest (QOI), with the goals of accounting for all sources of uncertainty and quantifying the contributions of specific sources to the overall uncertainty, i.e., UA and SA applied in tandem.

Although UA and SA applications are rising, most ABMs struggle with a shortage of testing in general, mainly due to time and other resource constraints (Kelly (Letcher) et al., 2013). Lee et al. (2015) argue that while a modeler invests a lot of time and effort in the development of ABMs, the output analysis is not always considered as deserving the same resource-intensive attention. According to a survey carried out by Heath et al. (2009), less than 5% of ABM publications present any statistical validation techniques. Angus and Hassani-Mahmooei (2015) argue that one possible cause for this "methodological anarchy" derives from the fact that, with so many possible degrees of freedom within an ABM, the responsibility to ensure and to demonstrate that a model is structurally sound and the prediction is reliable falls into each modeler.

We present a UQ workflow to integrate UA and SA in the evaluation of agent-based simulation outputs. We illustrate the use of this workflow in a particular spatial explicit LUCC case study in the framework Multi-Agent System for Environmental simulation, MASE-BDI Coelho et al. (2016). We apply general practices that should be a routine, to improve the level of confidence in results and to promote more rational and efficient use of ABMs. We may cite that broader and more complete workflows for the application of SA were already proposed, such as Pianosi et al. (2016) and Norton (2015). The UA-SA integrated proposal is what set our manuscript apart. We argue that UA should be used as an input to SA, in a broader process of UQ. Also, we noticed some conflicting results when we compared relevant studies on SA, mainly regarding the experimental setup. Table 3.1 summarizes the studies found in the literature (Vanrolleghem, Mannina, Cosenza, & Neumann, 2015)(1), Gan et al. (2014)(2), Wang, Li, Lu, and Fang (2013)(3), Yang (2011)(4), Pappenberger et al. (2008)(5), Y. Tang, Reed, Wagener, and van Werkhoven (2007)(6). Some authors have compared different SA methods and experimental setup, which are presented in the different lines of the table.

Table 3.1: Selected applications of sensitivity analysis approaches.

Reference	**Research Field**	**No. factors**	**Sampling**	**SA method**	**No. runs**
1	Urbain drainage	17	MOAT	MOAT	3000
			FAST	E-FAST	3000
			LH	SRC	2800
2	Watershed	13	MC	SPEA	3000
			MC	SRC	3000
			MOAT	MOAT	3000
			METIS	MARS	3000
			METIS	SOT	3000
			MC	DT	400
			LH	DT	400
			OA	DT	529
			OALH	DT	529
			LPTAU	DT	3000
			METIS	DT	3000
			METIS	GP	3000
			FAST	FAST	2777
			rLH	McKey	2890
			SOBOL-QR	SOBOL	3000
3	Crop growth	47	FAST	E-FAST	2049
4	Watershed	5	SOBOL-QR	SOBOL	18000
			MC	MOAT	3000
			MC	LR	3000
			MC	RSA	3000
			SOBOL-QR	SDP	500
5	Flood inundation	6	rLH	SOBOL	8192
			rLH	MOAT	12000
			rLH	Entropy-based	3000
			rLH	RSA	5000

Continued on next page

Reference	Research Field	No. factors	Sampling	SA method	No. runs
6	Watershed	18	SOBOL-QR	SOBOL	8192
			IFFD	ANOVA	1000
			LH	RSA	10000
			LP	PEST	10000

Where: MOAT = Morris screening One-at-A-Time; (E-)FAST = (Extended) Fourier Amplitude Sensitivity Testing; (r)LH = (replicated) Latin Hypercube; SRC= Standardized Regression Coefficient; MC = Monte-Carlo; LR = Linear Regression; SPEA = Spearman Correlation Coefficient; MARS = Multivariate Adaptive Regression Splines; SOT = Sum-of-Trees; DT = Delta δ Test; OA = Orthogonal Array; OALH = Orthogonal Array-based Latin Hypercube; IFFD = Iterated Fractional Factorial Design; SOBOL-QR = Sobol quasi-random; RSA = Regionalized Sensitivity Analysis; LP = Local Perturbation; PEST = Parameter Estimation Software.

Table 3.1 illustrates a glimpse of the myriad of possible combinations of strategies for sampling the model parameter space and SA methods, to quantify the impacts of sampled parameters on the model QOI. We understand that there is no combination of sampling and SA method that fits all applications. Thus, the work of Gan et al. (2014) shows that different sample strategies can even produce different outputs regarding the same SA method. Also, it seems that there isn't a clear relationship between the number of factors and the number of necessary runs to compute SA. Furthermore, in some cases, the number of runs used in the same sampling and SA method is not even in the same order of magnitude. For example, Pianosi et al. (2016) recommend $> 1000 \times M$ model runs to calculate variance-based SA, such as FAST, where M is the number of input factors subject to SA. Neither Wang et al. (2013) nor Vanrolleghem et al. (2015) nor Gan et al. (2014) executed this many number of runs. The first used a sample of size 2049 for a 47-factor problem (instead of $> 47,000$), while the second used a sample size of 3000 for a 17-factor problem (instead of $> 17,000$). The third used a sample size of 2777 for a 13-factor problem (instead of $> 13,000$). One could ask whether the number of runs should be based on something more than M.

In this manuscript, we will test different experimental strategies for a UQ workflow and discuss their relative benefits and limitations. A baseline scenario was developed, and we performed a comprehensive investigation of the impacts that differences in sample sizes, sample techniques, and SA methods may have on the QOI. In this work, we address the research question: how UA and SA may be applied to improve users' understanding of the uncertainty and relations among input and output responses in LUCC agent-based simulations? We are interested in finding which parameters are responsible for the most of the results' variability; if there is convergence when different SA techniques are applied; and finally, if there is a minimum sample size to achieve it. Although the statistical techniques are applied in a specific agent-based simulator, the methods described are quite general and may illustrate their application in another research.

In Section 3.2, we provide an overview of the different methods regarding variance stability, parameter space exploration, UA, and SA. We also present the proposed UQ workflow in Section 3.2. In Section 3.3, we describe the MASE-BDI framework and LUCC model used as a case study, followed by the experimental design. We present the results compared to related work. We discuss challenges and provide some assessment to extrapolate our finding into more general conclusions, to produce more robust or parsimonious models, as well as to make models more defensible in the face of scientific or technical controversy (Section 3.4). Finally, in Section 3.5, we summarize our findings and outline future work.

3.2 Materials and methods

The methods we applied in the case study are presented in this section alongside their experimental design. The UQ experiments have the objective to perform an output analysis on spatial stochastic models, to measure uncertainty and to reduce it. Ultimately, we want to understand better how the model behaves and expand our confidence in the response of a LUCC model.

3.2.1 Variance stability

Agent-based simulations are often stochastic, and therefore any analytical exercise requires an outcome pool drawn from a sufficient number of samples. It is only possible to draw conclusions if the output mean and variance reaches relative stability. Otherwise, the statistics could harbor too much uncertainty to be reliable (Lee et al., 2015). Moreover, some ABM simulations (MASE-BDI included) can take longer run times, which makes the execution of large samples prohibitive. Hence, knowing the minimum sample size to reach variance stability can be more compelling to modelers.

There are many methods to assess variance stability (Law & Kelton, 2000; Lee et al., 2015). We chose to apply the method proposed by Lorscheid et al. (2012), whose strategy is to assess stability from metrics on an outcome for a sequence of sample sizes. The proposed metric relies on the functional ratio between the variance and the sampled mean.

The coefficient of variation c_V is a dimensionless and normalized metric used to measure the uncertainty surrounding the variance, i.e., used for the analysis of experimental error variance. It is defined as the ratio of the standard deviation of a number of measurement s to the arithmetic mean μ

$$c_V = \frac{s}{\mu}. \tag{3.1}$$

If c_V is obtained from a small sample, e.g., it will vary more than if each sample contained far more runs. Lorscheid et al. (2012) propose a fixed epsilon (E) to limit c_V. This is done by calculating the c_V's of different sized set of simulation runs, in ascending order of size. The sample size at which the difference between consecutive c_Vs falls below the determined criterion E, and remains so, is considered a minimum sample size or the minimum number of simulation runs for ABMs. This is the point of variance stability. These points should be obtained for all ABM outputs, thereby the minimum number of runs for the ABMs is the maximum of these points (Lee et al., 2015)

$$n_{min} = argmax_n |c_V^{x,n} - c_V^{x,m}| < E, \forall x \text{ and } \forall m > n,$$

where n is the sample size; n_{min} is the estimated minimum number of required simulation runs; x is a distinct output; and m is some sample size for which the c_V is calculated. Thus, we apply the Lorscheid et al. (2012) method to establish the minimum sample size that guarantees that variance stability is achieved.

3.2.2 Parameter space exploration

Sampling methods provide a systematic exploration of the parameter space that guarantees the sample to have specific statistical or structural properties. The purpose of these methods is to actively reduce the number of parameter sets that are considered but still chose space-filling points in the design space (Thiele, Kurth, & Grimm, 2014a). For a complete revision of sampling methods, readers can refer to Gong et al. (2015); Kleijnen et al. (2005); Saltelli et al. (2008). In this manuscript, the most common sampling designs are illustrated and applied in the UQ process.

Since there are many methods to explore the parameter space, readers may have an overview of those sampling methods in Appendix B, including: Monte Carlo sampling (MC), Latin Hypercube (LH), Orthogonal Array (OA), Orthogonal Array-based Latin Hypercube (OALH), METIS sampling, Fourier sampling algorithm, LPτ (LPTAU), Sobol Extended (SOBOL), Morris one-at-a-time (MOAT).

3.2.3 Uncertainty analysis

UA evaluates and quantifies how the variability of input factors propagates through the model and affects the variability of output values (Ligmann-Zielinska et al., 2014). UA can also answer if there are any discontinuities associated with the distribution of results (Iman & Helton, 1988), plot the distribution itself, calculate the average output, the standard deviation, the quantiles of its distribution, and confidence bounds. An overview of the UA process can be found in Appendix C.

For the proposed UA, the only parameters considered relevant are the ones related to the QOI and previously selected as input factors of interest. All the other model factors and information fed into the model are disregarded, i.e., they do not vary, thereby they cannot cause variation in the output. However, the model outputs Y_j are non-deterministic because of the stochastic component derived from the emergence of the agent's behavior. Therefore, to ensure robustness, each vector $(\alpha^{(j)}, \beta^{(j)}, ...)$ of the output must be evaluated regarding the mean and the variance (Dosi, Pereira, & Virgillito, 2017). This confirmation is executed by a given number of model runs but with the same parameters configuration (ten Broeke et al., 2016).

After the UA quantified the magnitude of the resulting uncertainty in the model predictions due to uncertainties in model inputs, the next step in the UQ workflow would be to perform SA.

3.2.4 Sensitivity analysis

SA is the study of "how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input" (Saltelli et al., 2008). The authors show that each measure of sensitivity may produce its ranking of factors by importance.

There are different methods of SA, and each one has advantages and limitations. In the particular case of SA in spatial models, we incorporated the general guidelines provided by Lilburne and Tarantola (2009). It is clear from their work that each SA method has sampling and pre-processing technique requisites. Therefore, a careless combination of methods will result in inefficient and inappropriate results. Also, not all of the methods are capable of providing sensitivity index for non-monotonic input-output dependencies typically observed in ABMs (Fonoberova et al., 2013; ten Broeke et al., 2016). Therefore, we selected ten well-known methods of qualitative and quantitative SA. They were applied in MASE-BDI to verify if they were capable of providing those indexes for the LUCC model.

In general, gradient and linear-regression-based SA are known as qualitative methods, since they use some heuristic to represent the relative sensitivity of the parameters.

We will assess the Morris one-of-a-time screening method (MOAT) (Morris, 1991) and some correlation analysis, such as Spearman (SPEA) (Spearman, Con, & Page, 1904) and the standard regression coefficient (SRC). Variance-based methods are classified as quantitative methods because they tell the sensitivity of a parameter by calculating the impact of this parameter on the total variance of the model outputs (Saltelli, Tarantola, Campolongo, & Ratto, 2004). We will assess three variance-based SA techniques: SOBOL (Sobol', 1993), FAST (Cukier, Fortuin, Shuler, Petschek, & Schaibly, 1973), and McKay (McKay, Morrison, & Upton, 1999b). Also, we compare response-surface methods, such as Sum-of-Trees (SOT) (Breiman, Friedman, Olshen, & Stone, 1984; Chipman, George, & McCulloch, 2012), Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991), and Gaussian Process (GP) (Gibbs & MacKay, 1997). Other screening method such as the Delta δ Test (DT) (Pi & Peterson, 1994), are also assessed. The overall mechanisms of each method are discussed in Appendix D. The implementations of each technique are not provided due to space constraint, but readers may refer to Gan et al. (2014); Tong (2005) for details.

3.2.5 UA–SA integrated workflow

The integration of UA-SA has been applied to ABMs in a few relevant studies (Fonoberova et al., 2013; Ligmann-Zielinska et al., 2014; Ligmann-Zielinska & Sun, 2010b; Parry et al., 2013), that argue that a systematic evaluation of ABMs must comprise of an integrated approach to quantification of model output variability and its sensitivity to inputs. We followed the terminology of the National Research Council (2012) and called this process UQ: the process of quantifying uncertainties associated with a model QOI, to account for all sources of uncertainty (UA) and quantifying the contributions of specific sources to the overall uncertainty (SA). Figure 3.1 presents an overview of the UQ integrated workflow, with UA and SA as part of the modeling process, adapted from the original one proposed by Ligmann-Zielinska et al. (2014).

Analyzing the workflow, we argue that UA should be used as an input to SA, in a broader process of UQ. ABM input factors are often diverse, and the stochasticity makes multiple model runs a paramount step of the ABM's output evaluation. Once the modeler defines what is the QOI to be investigated, UA should be incorporated in the modeling process to indicate what is the variability of the QOI outcomes. The next step would be to test the sensitivity of model response to changes in the factors. This discovery could identify interactions among factors, factor fixing and prioritization that could lead to a model simplification, the reduction of output variance or the improvement of model accuracy.

This larger UQ process involves many smaller tasks, so a more detailed workflow is presented in Figure 3.2. Pianosi et al. (2016) proposed a practical workflow for the application of SA, with four fundamental group of activities: i) experimental setup; ii) input sampling; iii) model evaluation; and iv) post-processing. This work presents a state-of-the-art review and a very concise guide to good practices for readers. However, ABMs have specific characteristics, mainly due to stochasticity, uncertainties, equifinality, and because of the complex system applications. We took Pianosi et al. (2016) work as a guideline and tailored the level of effort and estimation to fit ABM needs. The main difference is the simplification of the SA tasks and the incorporation of the UA tasks.

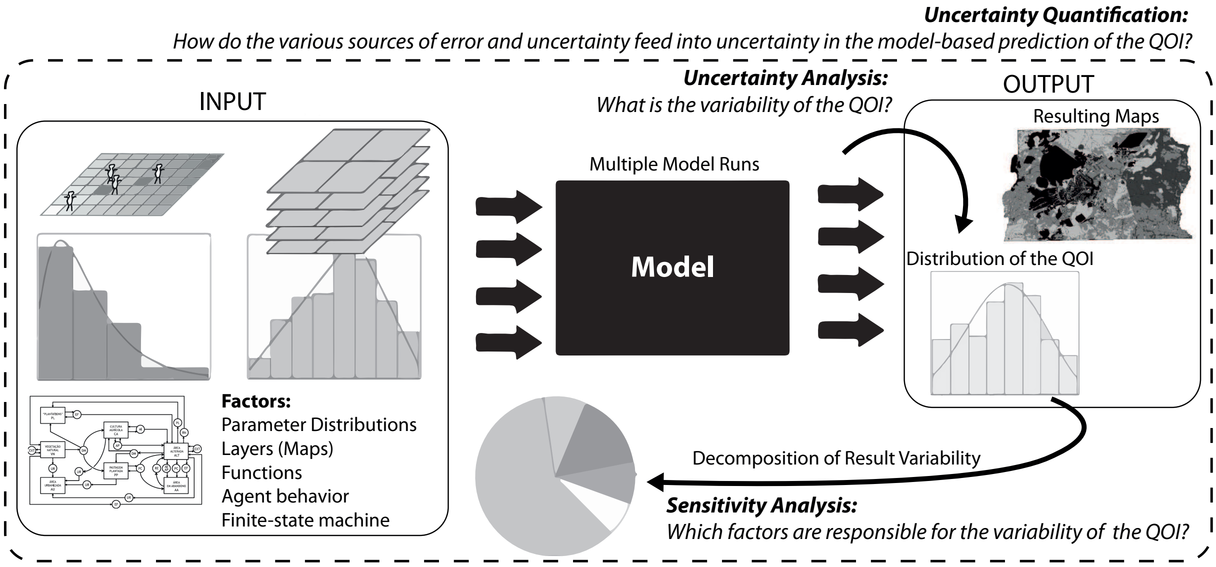

Figure 3.1: Overview of the modeling process, including UQ, UA and SA specific questions. Source: Adapted from (Ligmann-Zielinska et al., 2014).

Because a portion of ABM uncertainty is irreducible, a comprehensive evaluation of ABM uncertainty should assume that code verification, model-parameter calibration and validation have been successfully accomplished before UQ process begins for a QOI. The UQ workflow for ABMs (Figure 3.2) is composed of three basic steps: experimental setup, UA, and SA. We maintained the terminology proposed by Pianosi et al. (2016) as (*) in Figure 3.2. The first step of the workflow regards the experimental setup with basic choices: i) defining the QOI - the modeler must specify what the QOI for the problem at hand is; ii) select the input factors of interest; and iii) specify the range or distribution probability of each factor. The fourth task iv) is to determine variance stability - which represents the minimum number of simulation runs that accurately report the descriptive statistics.

In the second step of the workflow, the UA is composed of three tasks that summarize what is needed to discover what is the variability of the QOI in an ABM. After choosing the sampling strategy, the modeler would run multiple simulations (the minimum number of runs is provided by the last task of the experimental setup - define variance stability). ABM modelers usually choose factor values randomly from their respective range/distribution. As a result, UA produces a distribution of the QOI. The last task is to use this distribution to quantify the variability of the QOI, i.e., the use of descriptive statistics to analyze the model outputs.

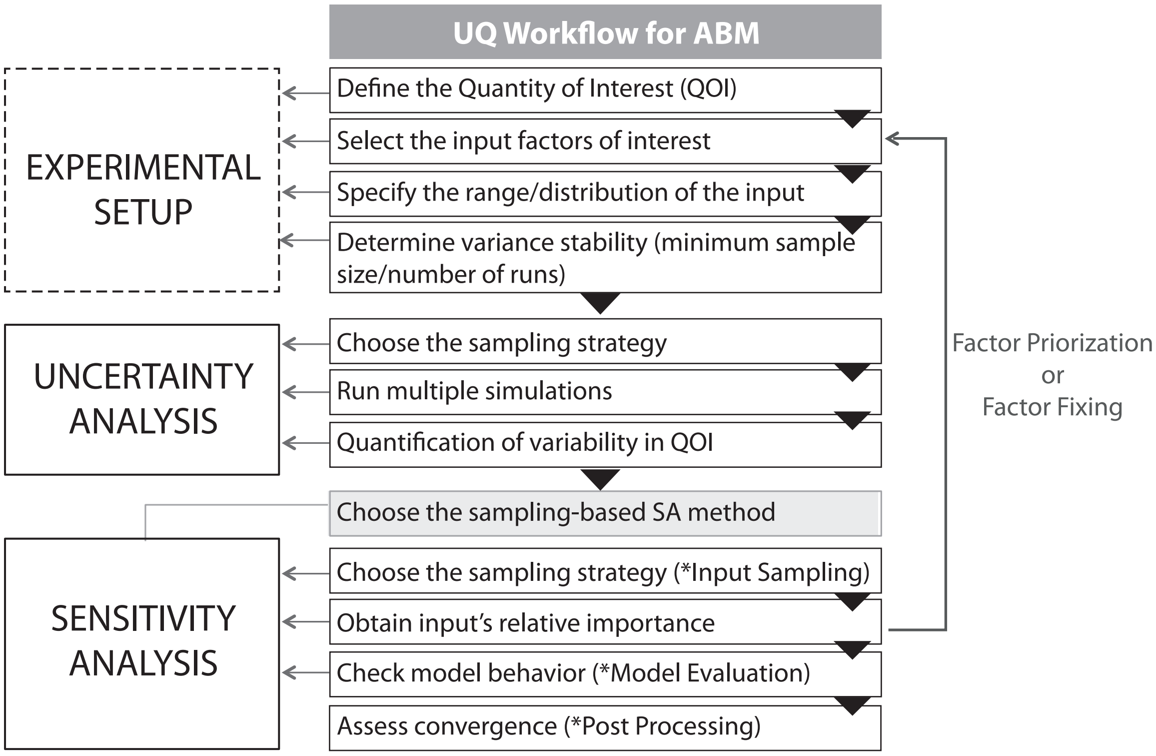

Figure 3.2: A UQ workflow for the application of UA + SA. Source: Adapted from (Pianosi et al., 2016).

The third step is a simplification of Pianosi et al. (2016) original workflow. It all begins with the selection of the SA method. Although the original work proposed a classification system based on the SA purpose, the literature shows that, for ABMs, this choice is somewhat model-specific. We decided to leave this decision to the modeler and tested many different methods to see the impacts of the SA method in our case study. The next task would be to define the input variability space by choosing the sampling strategy to be applied. There are several sampling methods, and although MC is still the most used sampling strategy, we tested different combinations of well-known techniques, such as MC and SOBOL, and also tailored sampling strategies to see if there would be an impact on the sensitivities outcomes.

The number of model runs required to perform SA is usually a rough estimation of a function of the number of factors subject to SA. We postulate that this minimum sample size should be equal or larger to the variance stability number of runs defined in the experimental setup step. We also test this empiric assumption and discuss it in later sections (Sections 3.3.4 and 3.4). This is what is necessary to obtain the factor's relative importance. The workflow's last two steps are checkpoints defined by Pianosi et al. (2016) to evaluate the model (check model behavior) and to assess convergence (check whether sensitivity estimates are independent of the size of the sample and if they would take similar values if we used independent samples). These steps inform us about the reliability of the results.

We applied the presented UQ workflow for ABMs in a case study. We tested several combinations of methods, sampling strategies, and sample sizes. In Section 3.3, we will present the application and the experimental setup designed for this application.

3.3 A land-use case study

MASE[1] is an agent-based simulation tool developed at the University of Brasília, Brazil. MASE enables modeling and simulations of LUCC dynamics using a configurable model and both top-down and bottom-up (Grimm, 1999) model structures simultaneously. MASE enables multiple types of agents with different behaviors to represent the interaction between agents with autonomy, the physical environment, and its relations (Ralha & Abreu, 2017). MASE has the overall goal of performing medium to long-term LUCC predictions. It also allows assisting decision-making processes related to LUCC.

We run the experiments in MASE-BDI, which is a freeware software extension of MASE that introduces cognitive reasoning-oriented agents through the implementation of the BDI rationality (Bratman, 1987). MASE-BDI was implemented in JADEX multi-agent platform (Braubach, Pokahr, & Lamersdorf, 2005). In the BDI model, agents have beliefs, a set of information about the world it inhabits, that changes both the perception and thinking about the world. Desires represent the motivational attitudes of agents, capturing the agent's wishes and driving the course of its actions. An agent can also make plans related to its intention to achieve its goals. This multi-agent reasoning model is defined as means-end-reasoning (Wooldridge, 2009).

The MASE-BDI architecture is composed of three layers (from top to bottom): a user interface, a utility layer, and an agent layer. The first provides an optional graphical interface (models and simulation parameters can also be defined directly in a configura-

[1]Project Website:`http://mase.cic.unb.br/`
Software availability:`https://gitlab.com/InfoKnow/MASE/MASE-BDI/SourceCode`

tion file) and a JADEX control center of the BDI model. The utility layer groups a set of modules to control the pre-processing of the maps and input of the geographic information. It also provides the simulation parameter automatic tuning, which is a complex and error-prone task in ABMs. The parameter adjustment is performed by employing efficient optimization algorithms to tune the simulation model parameters, concerning a user-defined single or multi-objective function of interest. Still, in the utility layer, a module of validation is responsible for evaluating the final simulation output maps and metrics (Coelho et al., 2016).

In the agent layer, we have an organization of hierarchical agents. The GRID Manager controls the general aspects of the simulation. The Spatial Manager controls the agents responsible for representing and updating the spatial environment. The Transformation Agents are computational entities accountable for moving, exploring, and reasoning about the space according to their internal goals and beliefs. The Transformation Manager rules and resolves the conflict due to the competition among transformation agents concurring for the same environmental resources. Readers who are interested in details of the MASE-BDI architecture, agent design and implementation may refer to Coelho et al. (2016).

3.3.1 The Cerrado Federal District study area

The Federal District of Brazil ($5,789km^2$) and its Cerrado (Brazilian savanna) coverage is the study area in this article. The simulations depict the land changes of the region (Figure 3.3), the most endangered biome in Brazil, and the second largest biome in South America harboring significant biodiversity. This area has been undergoing severe transformation due to the advance of cattle ranching and soy production, being an attractive study area for land use simulations. To allow replicability, the Cerrado LUCC simulation model was documented and described using the ODD protocol (Overview, Design concepts, and Details protocol) (Grimm et al., 2006). The characterization of agent behaviors and attributes in socio-ecological systems were applied by empirically grounding ABM mechanisms (Smajgl et al., 2011). A complete conceptual and methodological description of the model is available in Ralha et al. (2013).

The initialization data for the simulation is a couple of Landsat-derived grid raster maps consisting of the land cover of the region, from two different time periods (an initial and a final map). Furthermore, the user must adjust a set of initialization parameters of the multi-agent system, such as the number of agents that will explore the landscape (transformation agents), their typology (cattle ranchers and farmers), and characteristics of the initial behavior of those agents. The simulations are performed in steps, where each step corresponds to the measure of time defined by the user. In this example, one step equals to one week in chronological time. The user also determines the size of a

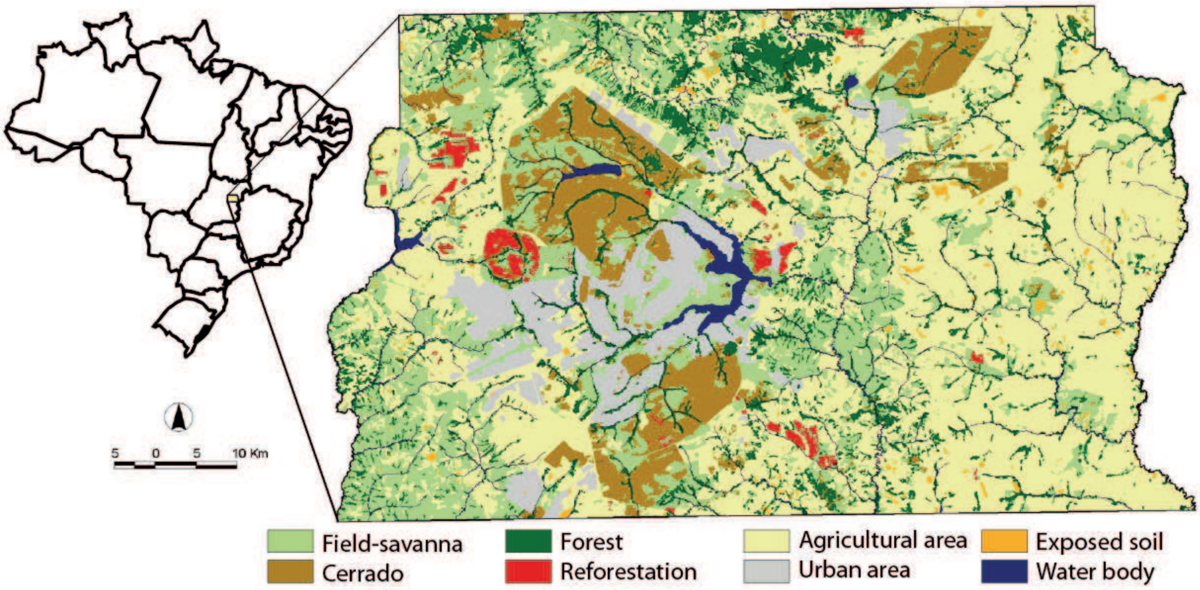

Figure 3.3: A land use map of the Cerrado study area in Federal District, Brazil. Data by (GDF, 2009).

plot or cell. Here, the total area of study was divided into plots of one hectare. The physical environment is spatially represented by a set of layers of geographical information data (shapes or raster files), such as rivers, lakes, slopes, building areas, highways, environmental protected areas, and regional zoning maps of the area. The aggregation of these geographical features determines the physical environment of any given point in the simulation grid. The transformation agents represent humans performing activities of cattle ranchers and farmers, with their behavior and beliefs, explicitly changing the natural landscape to achieve their internal goals (e.g., production expansion, sustainable exploration).

The simulations are calibrated by the simulation parameter automatic tuning tool, adjusting the parameters to best fit the observed change from the two initial maps. The outcome of the simulation is a result of the emergence of the agent's action within the duration of a simulation, determined by the user. The final landscape is a result of the emergence of the agent's effects on the land.

MASE-BDI is a spatially explicit framework because the results comprise of the quantity of land cover change and the spatial allocation of the change (which plots were chosen by the agents to initiate or expand their cattle ranching or farming business). The result of any MASE-BDI simulation is a couple of predicted maps with the spatial allocation of the land change, and the quantity of change - a set of metrics calculated during runtime, such as the total amount of land change. At the end of each simulation, the resulting image is submitted to a goodness-of-fit measurement, and the quality and errors of the quantity of change and allocation of land use change are calculated.

MASE-BDI produces stochastic simulations, which mean that the same input to the model may lead to a different result in the quantity and allocation of change. Therefore, the same set of parameters must be run several times to raise the confidence that the results are representative.

3.3.2 LUCC goodness-of-fit

According to Thiele et al. (2014a) there are two strategies for fitting model parameters to observational data: best-fit and categorical calibration. MASE-BDI applies the first strategy, in which we must find the parameter combination that best fit the data. The quality measure is one exact value obtained from the observational data, so it is easy to determine which parameter set leads to the lowest difference.

Pontius et al. (2008) define the most common quality measure for LUCC spatial explicit simulations; hence it is used in MASE-BDI. Although there is not a universally agreed-upon criterion to evaluate the goodness-of-fit of validation maps, the performance of the simulation model is done objectively by computing the sources of error of prediction maps.

A set of map comparisons is responsible for the evaluation of the model. Pontius, Huffaker, and Denman (2004) indicate that three maps are necessary: i) a reference map of the initial time t_0; ii) a reference map of a subsequent time t_1; and iii) a prediction map of the subsequent time t_1. There are three possible two-map comparisons, picking two maps at a time:

- Comparison between the reference map of time t_0 and the reference map of time t_1: characterizes the observed change in the maps, which reflects the dynamics of the landscape;

- Comparison between the reference map of time t_0 and the prediction map of time t_1: characterizes the model's predicted change, which reflects the behavior of the model;

- Comparison between the reference map of time t_1 and the prediction map of time t_1: characterizes the accuracy/error of the prediction's accuracy/error.

The total disagreement between any two maps that share a categorical variable is computed in terms of quantity disagreement and location disagreement (Pontius et al., 2004). Quantity disagreement derives from differences between the maps regarding the number of pixels for each category. Location disagreement is the difference that could be resolved by rearranging the pixels spatially within one map so that its agreement with the other map is as broad as possible. The sum of them both is the total disagreement.

To illustrate the methodology, we present the Brazilian Federal District Map with only two land cover categories: natural vegetation (the Cerrado), and developed (areas characterized by 30% or greater of constructed materials, e.g., asphalt, concrete, buildings). Considering these two categories, the comparison of pixels may result in the categories presented in Figure 3.4: error due to observed vegetation predicted as developed; correct due to observed developed predicted as developed; correct due to observed vegetation predicted as vegetation; and error due to observed developed predicted as vegetation.

According to Pontius et al. (2008), the most accurate applications are the ones where the amount of observed net change in the reference maps is larger. The Figure of Merit (FoM) is the ratio of the amount of correctly predicted pixels of change to the sum of all pixels

$$FoM = \frac{RightChange}{WrongPersistence + RightChange + WrongGaining + WrongChange},$$

where Wrong Persistence is the area of error due to observed change predicted as persistence; Right Change is the area of correct due to observed change predicted as change; Wrong Gaining is the area of error due to observed change predicted as wrong gaining category; and Wrong Change is the area of error due to observed persistence predicted as change.

FoM is a statistical measurement that can range from 0% - meaning no overlap between observed and predicted change, to 100% - meaning perfect overlap between observed and predicted change. When the amount of correctly predicted change is larger than the sum of the various types of error, FoM is greater than 50%. FoM is the best-fit quality measure of this manuscript. It is also the QOI chosen to illustrate the UQ workflow for ABM, as the first task of the experimental setup step.

It is worth mentioning that Pontius et al. (2008) set a testing benchmark, based on statistical methods for map comparison of 13 applications of different popular peer-reviewed land change models. The results show that in 12 of the 13 LUCC models predictive maps, the amount of error is more significant than the amount of correctly predicted change at the resolution of raw data. In contrast, MASE-BDI was able to surpass these statistics, presenting results that show high quality in the accuracy of their predictions (FoM$>$ 50). The complete explanation of the MASE simulation results using Pontius' statistical techniques of map comparison to land change models is presented in Ralha et al. (2013).

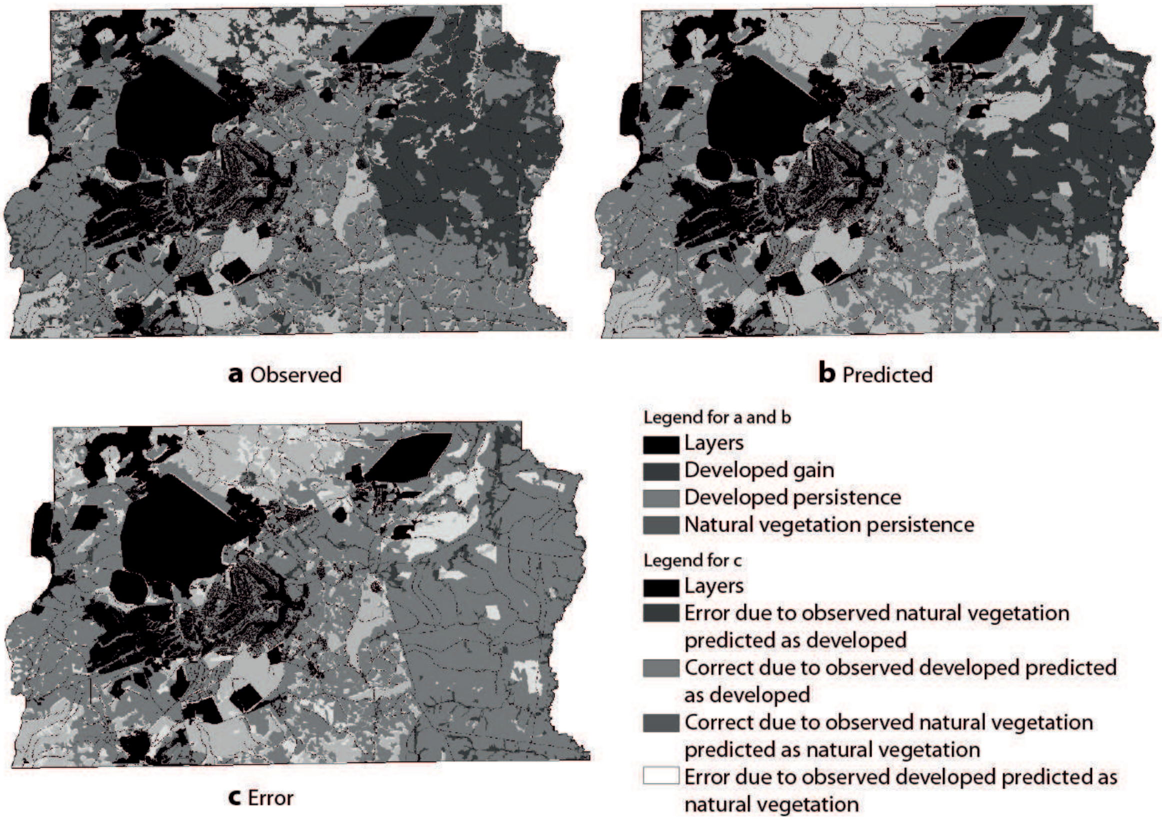

Figure 3.4: The Brazilian Federal District maps of: **a)** observed change 2002-2008, regarding the difference in observed land change within this period, produced from the input data itself; **b)** predicted chance 2002-2008, results produced by the simulated model; and **c)** prediction error 2008, generated when maps a) and b) are compared.

3.3.3 MASE-BDI and UQ tool integration

Previous work demonstrates that the initialization of the agents may have a substantial effect on the land dynamics and into the final simulation outcome (Lorscheid et al., 2012). Therefore, it was paramount to use a framework to control, calculate, trace, manage uncertainties, and finally make the output analysis feasible. The MASE-BDI framework itself does not provide the modeler with the means to statistically analyze the results. The difficulty to perform many different samplings, UA and SA analysis, may lead to a shortage of testing and finally to a perfunctory UQ. To avoid this pitfall, we chose a statistical platform that provided the tools needed to execute both UA and SA steps in the proposed UQ workflow for ABM (Figure 3.2).

Among the different UQ platforms available, we chose PSUADE[2] as the best fit to integrate with MASE-BDI, based on its smooth coupling with external models and variability

[2] `http://computation.llnl.gov/casc/uncertainty_quantification/`

and availability of UA and SA methods.

PSUADE is a software package composed of three main components: a *s*ample generator with the experimental design techniques; a *d*river to control the simulator execution environment; and an *a*nalysis toolset (Tong, 2005). The execution environment created by PSUADE allows sequential or parallel automatic simulation executions. We stylized the use of PSUADE by creating a Python driver to provide an interface for linking MASE-BDI' simulation executable code and PSUADE. Also, we created a graphical user interface (GUI) that clusters all PSUADE and MASE-BDI configurations, in a straightforward unified interface that encapsulates all the configuration complexity of both PSUADE and MASE-BDI. Users may edit the configurations of the model or the UQ analysis without having to handle directly the configuration files.

Figure 3.5 shows the flow of activities for MASE-BDI to work autonomously with the PSUADE tool, beginning with the configuration of the simulation and the UQ design of experiments, following through the generation of samples in PSUADE that are going to be the input of the multiple MASE-BDI simulations. All the MASE outputs are stored and compiled so the UA and SA chosen techniques would be applied. The UQ integration modules where designed to be model/framework independent, so that it can be coupled with PSUADE in any other model and platforms other than MASE-BDI. The codes of the implementation[3] are available to the research community.

3.3.4 Experimental setup

The application of the UQ workflow follows a sequence of steps that were presented in general terms in Section 3.2.5. Next, we describe the individual choices and methods used in a specific ABM application, the LUCC model simulated in MASE-BDI. We will present the choices we made at each step, and maybe help other modelers with our example.

Define the Quantity of Interest (QOI)

The first task of the experimental setup, the definition of the QOI, was determined as the output FoM, as described in Section 3.3.2. FoM was chosen as the QOI of our investigation as it represents the quality of our simulation predictions. The higher the FoM, the better fitted is the prediction.

Select the input factors of interest

Regarding the simulation data, a baseline scenario with fixed variables was selected for the LUCC model to investigate the initialization parameters of MASE-BDI. For this purpose,

[3] `https://gitlab.com/InfoKnow/MASE/MASE-BDI/SourceCode/tree/master/MASE-PSUADE`

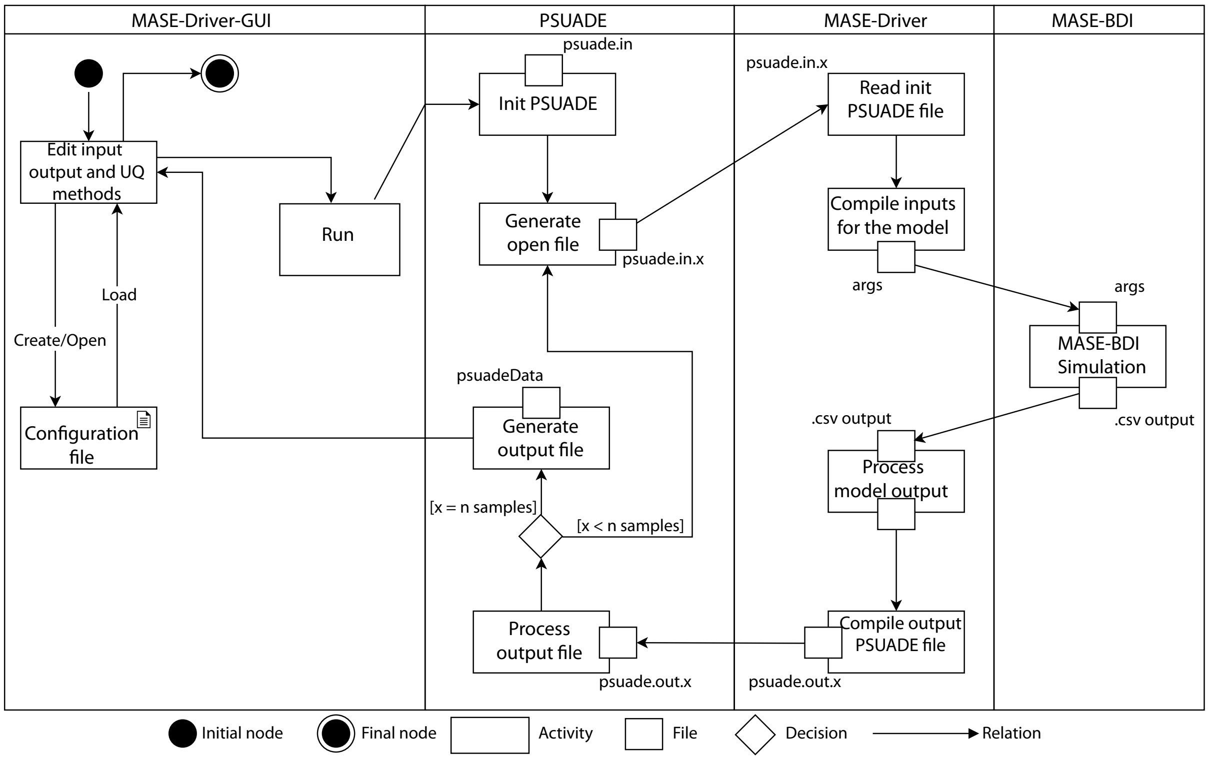

Figure 3.5: Activity diagram of MASE-Driver-GUI, PSUADE, MASE-Driver, and MASE-BDI tools.

there are no alterations in the geographic information in the simulated environment. All simulations were performed with only two types of transformation agents: cattle ranchers and farmers.

The input factors of interest refer to the number of agents initialized in a simulation, their initial state, and their behavior. These parameters characterize the instantiation of MASE-BDI agents, and therefore users may lack familiarity with those variables. The MASE-BDI provides a default value for the simulations, obtained through the calibration of the model. Therefore, these parameters are often a "black box" to users, and precisely because of this, can be an extra source of uncertainty.

The number of transformation agents (TA) is a parameter that reflects the number of computational agents (in the multi-agent system paradigm) instantiated in a simulation run. In this case study, one agent does not represent one single individual. TA was derived from data of the Brazilian Agricultural Census of 2006 and comprises a set of Producer legal status. The range of 1 to 100 is a percentage representation to the 3407 registered producers in the region. The MASE-BDI user must inform how many agents may be active or inactive in a given period. The details of those agent's characterization are thoroughly illustrated in Ralha et al. (2013). Likewise, the number of transformation group agents (TG) is an initial parameter which represents not an individual but an organization, cooperative, business, and so on. The range is an abstraction of the 548 group producers, ten of which have permanent exploration licenses.

The potential for exploration, individual or of a group, represents the impact an agent can produce in the natural vegetation cover of a cell during a step. In the Cerrado LUCC Model, considering the deforestation process, the potential of exploration is again an abstraction for the wood volume per hectare ($m^3.ha^{-1}$) of wood that can be obtained from a particular grid cell, until a nominal limit that represents resource depletion. The parameters of Table 3.2 will be the input for the UQ process.

Specify the range of the input

To illustrate the third task of the experimental setup step, Table 3.2 presents the four parameters that will vary in each run of the simulation. They were the selected input factors of interest, and the specification of the range of the input is presented in Table 3.2.

Table 3.2: MASE-BDI multi-agent initialization configuration parameters.

Parameter	Description	Distribution	Lower bound	Upper bound
TA	No. of Transformation Agents	Uniform	1	100
TG	No. of Transformation Group Agents	Uniform	10	100
IE	Potential of Individual Exploration	Uniform	1	500
GE	Potential of Group Exploration	Uniform	500	1500

In addition to the final LUCC maps, a MASE-BDI simulation generates 11 metrics as results. To evaluate the model response to the different parameters, FoM will be used as the objective function and the output to be analyzed in the UQ process. Nevertheless, another five variables were selected to observe the influence of the simulation input configurations on the model outputs. The experiments considered the outputs described in Table 3.3.

Table 3.3: MASE-BDI output parameters.

ID	Output	Description
1	FoM	Figure of Merit
2	IPA	Image Producer's Accuracy
3	IUA	Image User's Accuracy
4	WC	*Wrong Change*: observed change predicted as persistence
5	RC	*Right Change*: observed change predicted as change
6	WP	*Wrong Persistence*: observed persistence predicted as change

Variance stability determination

The last task in the experimental setup step of the UQ process is to define the minimum sample size through the determination of variance stability.

From a pool of over 138,800 model runs that were executed, $31,815$ runs represent the baseline scenario where only the four input variables vary (Factor Fixing of inputs presented in Table 3.2). The s and the μ for this fixed parameter set are already substantially smaller (Equation 3.1). We sampled from this fixed set to apply the variance stability methodology proposed by Lorscheid et al. (2012). In this multivariate setting, we compared the c_V (rounded to 1\1000) of differently sized set of runs (increased iteratively), $n \in \{10, 50, 100, 500, 800, 1000, 5000, 10000\}$.

The outcome drawn from runs of different sample techniques may affect variance stability. For clarification, we applied the proposed methodology with random (Monte Carlo) (Table 3.4) and quasi-random sampling (Table 3.5). We selected $E = 0.01$ as the limit of c_V.

Table 3.4: Coefficient of Variation at differently sized set of runs of Monte-Carlo samples

Output	n							
	10	50	100	500	800	1000	5000	10000
Figure of Merit	0.063	0.082	0.076	0.090	0.082	0.095	*0.091*	0.092
Producer's Accuracy	0.143	*0.138*	0.143	0.149	0.141	0.146	0.151	0.152
User's Accuracy	0.130	*0.122*	0.125	0.120	0.121	0.121	0.121	0.122
Wrong Change	0.602	0.485	0.593	*0.585*	0.575	0.578	0.568	0.572
Right Change	0.143	0.138	0.143	0.149	0.141	0.156	*0.151*	0.152
Wrong Persistence	0.229	0.242	*0.236*	0.242	0.244	0.254	0.250	0.250

Table 3.5: Coefficient of Variation at differently sized set of runs of Quasi-Random samples

Output	n							
	10	50	100	500	800	1000	5000	10000
Figure of Merit	0.018	0.117	0.075	0.093	*0.091*	0.087	0.091	0.094
Producer's Accuracy	0.089	0.184	0.135	0.152	*0.148*	0.147	0.151	0.152
User's Accuracy	0.124	*0.125*	0.122	0.123	0.119	0.123	0.122	0.121
Wrong Change	0.121	0.593	0.547	*0.557*	0.561	0.570	0.579	0.574
Right Change	0.089	0.184	0.135	0.152	*0.149*	0.147	0.152	0.153
Wrong Persistence	0.154	0.291	0.229	0.252	0.*248*	0.246	0.248	0.251

Although both means for FoM were roughly the same (MC FoM $\mu = 50.59$; QR FoM $\mu = 50.57$), the minimum number of runs were somewhat different for almost every output. For each outcome of interest {FoM, PA, UA, WC, RC, WP} the respective point of stability were {*5000*, 50, 50, 500, *5000*, 100} applying random sampling (Table 3.4), and {*800*, *800*, 50, 500, *800*, *800*} applying quasi-random sampling (Table 3.5). The highlighted values (*italic*) on Table 3.4 and 3.5 are the c_V that fall below the defined E. Therefore, the minimum number of runs for the Cerrado LUCC model would be 5000 MC random samples or 800 QR samples. Since we are looking for efficiency, 800 will be considered the minimum sample size (number of runs).

3.3.5 The methods for UA

In the second step of the UQ workflow, there are three tasks. The first one, to *choose a sampling strategy*, derives from the findings of the variance stability task. We chose the quasi-random sampling design since it was more effective in the definition of a minimum sample size. The second task of the UA step is to *run multiple simulations* of the model under study. Again, we used the findings of the experimental setup step as the minimum sample size. Therefore, 800 simulation runs were performed.

The third task is the *quantification of variability in QOI*. We performed descriptive statistics and statistics of dispersion of the outcomes to draw some UA conclusions for the second step of the UQ workflow. We will present the results only for the QOI: the FoM output. First, four initial moments of the sample are derived: the first moment ($\mu = 50.57$, standard error of $\mu = 0.16$), summarizing the central tendency of the stochastic model; the second moment (variance); the third moment (skewness); and the fourth moment (kurtosis). The results are summarized in Table 3.6. Also, the data set has $\sigma = 4.62$. To explore the variability of the simulation results, we performed UA by examining the observed distribution of the FoM of the sample resulting simulations. Figure 3.6 summarizes the empirical density and the cumulative distribution function of the experiment (800 model runs).

Table 3.6: Moments results of MASE-BDI model's objective function value - Figure of Merit.

Mean	Variance	Skewness	Kurtosis
50.57	21.33	-3.01	12.20

A Cullen and Frey graph (a squared skewness-kurtosis plot) is presented to illustrate whether the FoM followed a particular distribution. The data was bootstrapped using Monte Carlo samples to consider the uncertainty of the estimated values of kurtosis and skewness. Figure 3.7 is a plot with 1000 boot values. The diagram indicates that the skewness and kurtosis are consistent with a beta theoretical distribution, but the interval of FoM (not in the interval $[0, 1]$) disprove it. The data does not necessarily follow any

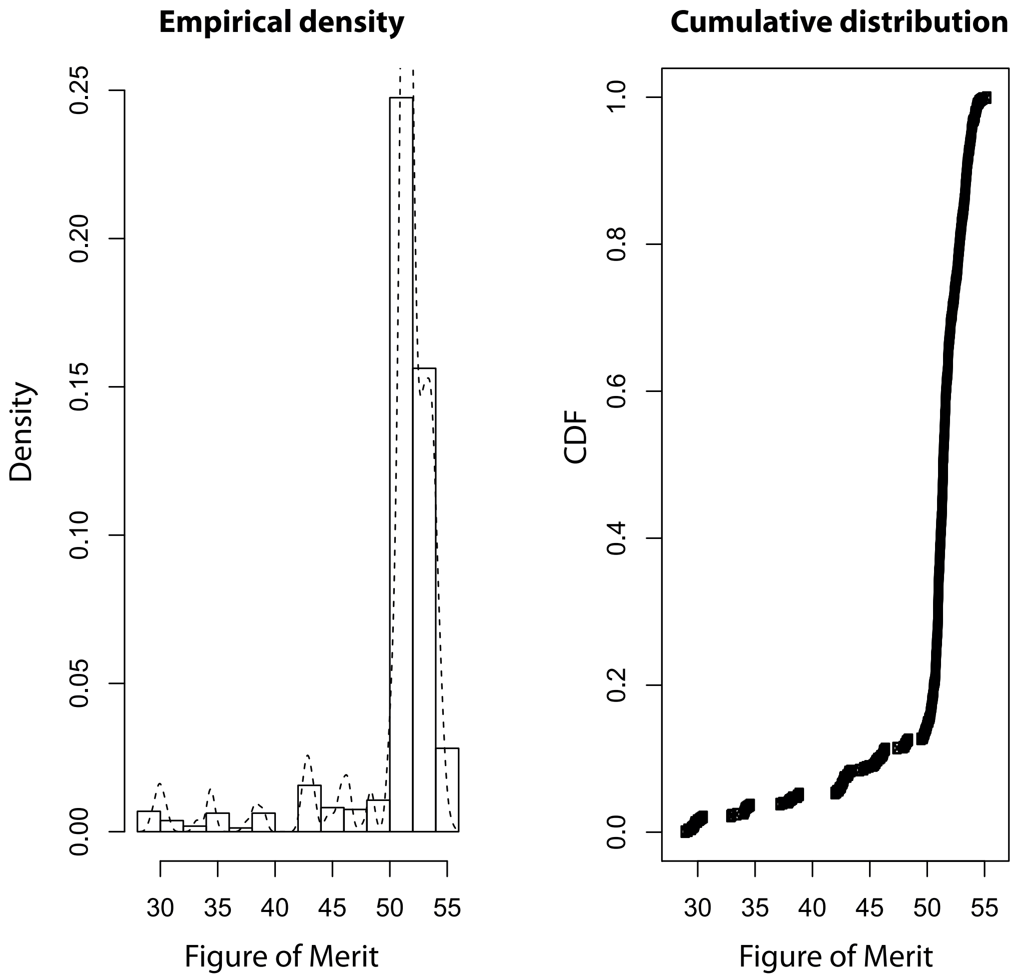

Figure 3.6: Observed distribution of Figure of Merit output. Histogram of empirical density of the data (left) and the cumulative distribution (right).

particular distribution, which means that the normality assumption and other known distributions do not refer to the observed data. Rather, the assumption is that the process that produces the data is a distributed process. So that process, likewise, can never be precisely normal because of asymmetries, discreteness, and boundness of the observable data.

3.3.6 SA experimental setup

For the last step of our proposed UQ workflow, multiple combinations of different sample strategies and sensitivity methods were tested to answer our research questions (Section 3.1) regarding SA. Instead of arbitrarily choosing an SA method (task 1: *choose the sampling-based SA method*) and the sampling strategy (task 2: *choose the sampling strategy*), we decided to test multiple combinations of techniques. The configuration of the experiments is presented in Table 3.7, following a similar experimental design of what was proposed by Fonoberova et al. (2013) and followed by Gan et al. (2014).

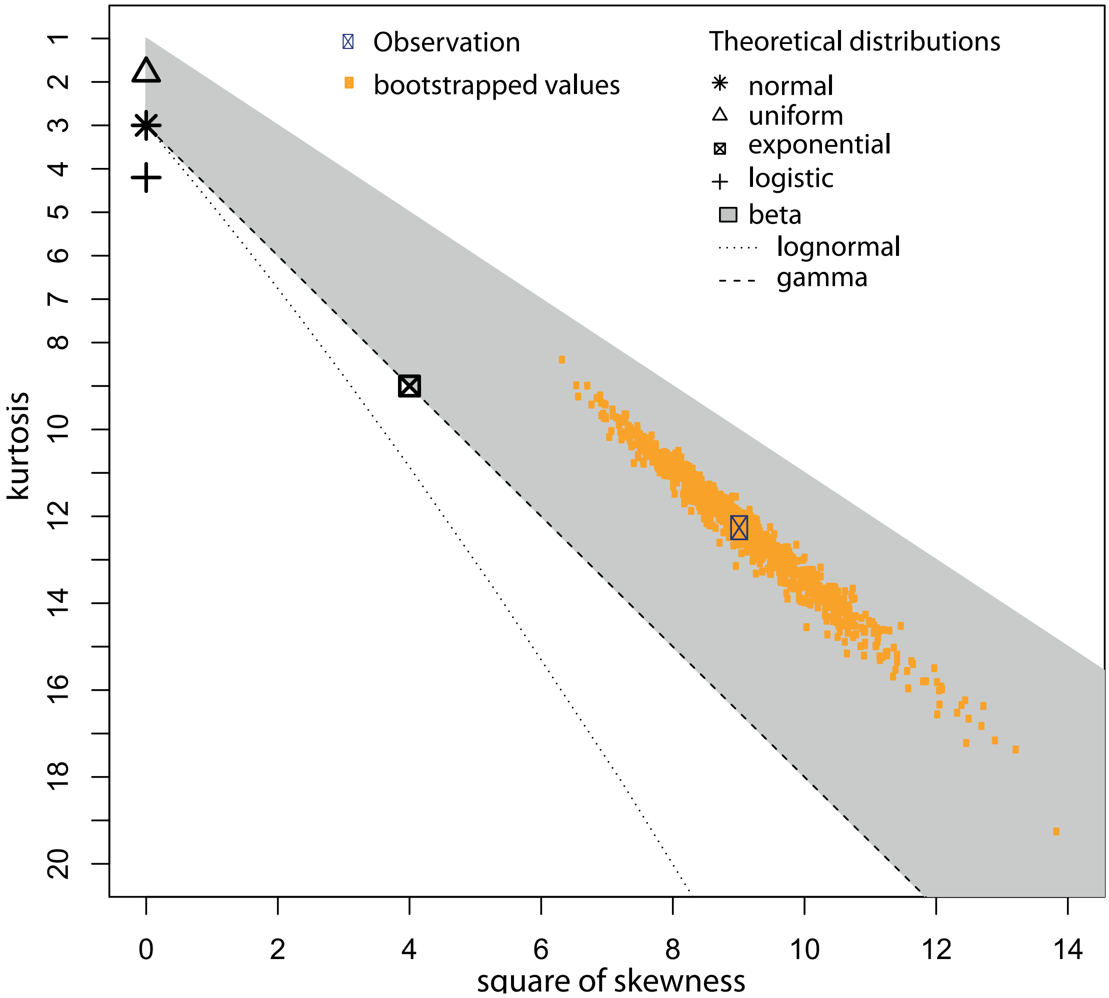

Figure 3.7: Bootstrapped Cullen and Frey graph of FoM results kurtosis and squared skewness.

We established the minimum quasi-random sample size of $N = 800$ runs as a guideline for the other sampling techniques. The differences among the sample size in Table 3.7 were due to the requisites of each sampling technique. The sample size for MC, METIS, and LH was assigned as 800 since there are no prerequisites for these techniques. The sample size of OA was set to $841 (= 1 \times 29^2)$.

For MOAT and SOBOL, 160 and 140 replications were used, resulting in samples of size 800 and 840, respectively. For the FAST technique, the maximum harmonic is $M_s = 6$ and the maximum frequency $\omega_{max} = 41$, when $n = 4$. Thus, the maximum size of the FAST sample for four inputs is 493. We decided to keep the FAST sample experiment, even though it disregards the variance stability calculation, as an open question of the experiment.

Table 3.7: Experimental configuration for the comparison of sensitivity analysis methods.

Sensitivity Analysis			Sampling	
Type	Method	Sensitivity measurement	Technique	Size
Gradient	MOAT	Modified Mean and Standard Deviation	MOAT	800
Linear-regression	CA	Spearman Correlation Coefficient (SPEA)	MC	800
	RA	Standardized Regression Coefficient (SRC)	MC	800
Response-surface	SOT	SOT score of sensitivity	METIS	800
	MARS	MARS score of sensitivity	METIS	800
	GP	GP score of sensitivity	METIS	800
Other	DT	Delta score of sensitivity	MC	800
Variance	Sobol	Sobol First and Total Indexes	SOBOL	840
	FAST	First order index	FAST	493
	McKay-1	First order correlation coefficient	LH	841
	McKay-2	Second order correlation coefficient	OA	841

To avoid an *ad hoc* definition on the sample size, we applied the same method presented in Section 3.2.1 by fixing all input parameters and choosing an $E = 0.001$. A quasi-random sample of 50 runs was determined as sufficient to qualify the model results for this given set of parameters. The next tasks of the SA step are to *obtain input's relative importance*, to *check model behavior* and to *assess convergence*. Those are presented and discussed in the following Section.

3.4 Output analysis results and discussion

To continue to execute the following tasks of our SA step, we must perform many tests and simulation. The global SA of all model outputs was performed using the MASE-Driver-PSUADE integration. The primary data obtained from the execution of each of the simulations are available for checking, reviewing, and replicating the experiments[4].

Input's relative importance

The method of global gradient SA is presented in Figure 3.8. Results from both methods of linear-regression-based SA are presented in Figure 3.9. Response surface SA methods are presented in Figure 3.10. The sensitivity scores represent the first-order indices, i.e., the contribution to the output variance by every single input alone. If the parameters are normalized $[0, 1]$, then the most sensitive parameters get a score next to 1 while the least sensitive ones get a score next to 0. The vertical axis in these figures denotes the

[4]Simulation results and UQ raw data: `https://gitlab.com/InfoKnow/MASE/MASE-BDI/SourceCode/tree/master/PSUADE%20Raw%20Data`

MASE-BDI input parameters used in the experiments. The simulations were performed according to the experiment design (Table 3.7). The color scale of each grid indicates the order of sensitivity from low to high; that is, light colors for low data values and dark colors for high data values.

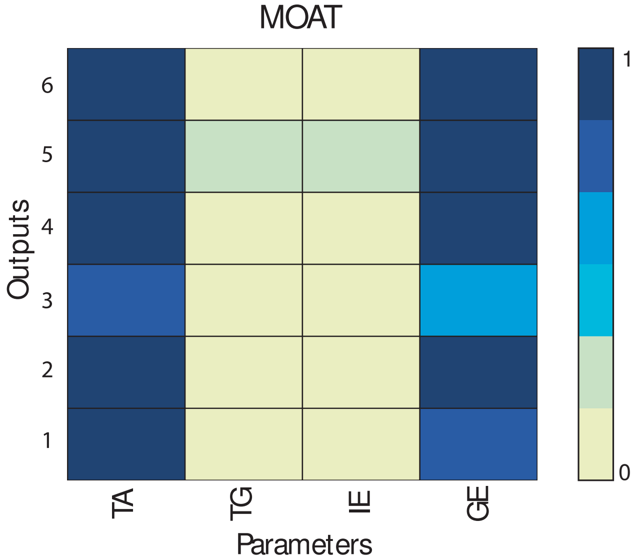

Figure 3.8: Heat map of MOAT gradient-based sensitivity analysis for MASE-BDI simulations, where TA - No. of Transformation Agents, TG - No. of Transformation Group Agents, IE - Potential of Individual Exploration, and GE - Potential of Group Exploration.

Figure 3.11 presents the compilation of all qualitative SA methods regarding one single output: FoM. FoM was chosen as the QOI of our investigation, as presented in Section 3.3.4. The results of the variance-based (quantitative) SA methods for the FoM output are summarized in Table 3.8.

To address the minimum sample size to detect the most sensitive variables efficiently, SA was calculated at different sample sizes for each SA method. We illustrate the application of MARS SA technique, exclusively for the FoM output, with different sampling methods and sampling sizes, as presented in Figure 3.12. The final result for minimum sample sizes and sampling methods are compiled in Table 3.9.

Check model behavior and assess convergence

The application of UA and SA offers a valuable complement to each other, and their close relation in ABMs has been proven by Fonoberova et al. (2013); Ligmann-Zielinska et al. (2014); Pianosi et al. (2016). Since the Cerrado LUCC model is stochastic, there is intrinsic uncertainty in the model even when all model parameters are fixed. One of the main concerns of our work was to find the minimum number of model evaluations,

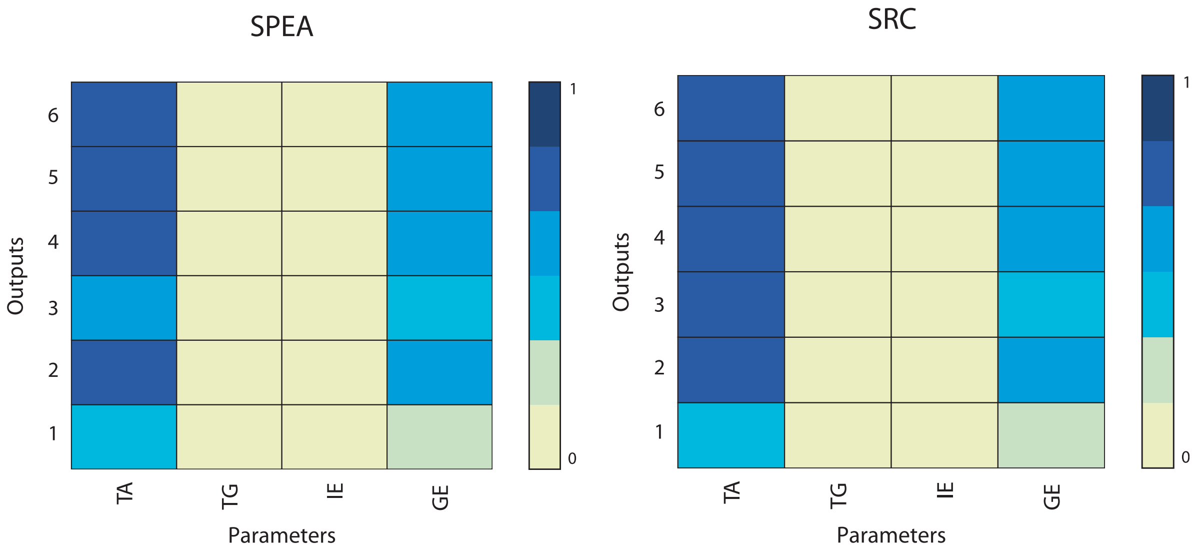

Figure 3.9: Heat map of linear-regression-based sensitivity analysis methods for MASE-BDI simulations, where TA - No. of Transformation Agents, TG - No. of Transformation Group Agents, IE - Potential of Individual Exploration, and GE - Potential of Group Exploration.

Table 3.8: Percentage of the variability of the results for each input based on variance-based SA results for FoM output.

Method	Sensitivity Measure	Input %			
		TA	TG	IE	GE
FAST	Total-effect index	61.72	0.17	0.12	37.99
McKay-1	First-order Correlation Coefficient	59.31	0.94	1.03	38.72
McKay-2	Second-order Correlation Coefficient	51.59	1.65	0.86	45.90
Sobol-1	First-order index	57.59	0	0	42.41
Sobol-t	Total-order index	54.89	0.03	0.01	45.07

i.e., the number of simulation runs that were required to secure the stability of output variance. We chose to apply the methodology brought by Lorscheid et al. (2012) and discussed by Lee et al. (2015).

Regarding the minimum number of runs in MASE-BDI, the found problem-specific point of stability was 800. This result stays in the middle of the typical find in the literature for a small number of inputs. The Gan et al. (2014) analysis is based on the $10 \cdot n$ rule, where n=number of input factor subject to SA. Pianosi et al. (2016) argue that the number of runs depends on the SA purpose, that should be around 1 to $1000 \cdot n$. When the purpose is screening the parameters through variance-based methods, the theoretical minimum number of runs should be $1000 \cdot n$.

From the results, it is clear that some statistical estimation must be done before

Table 3.9: Minimum sample sizes for each sampling technique.

Sensitivity Analysis		Sampling	
Type	Method	Technique	Size
Gradient	MOAT	MOAT	100
Response-Surface	SOT	MC	400
		LH	400
		LPTAU	400
		METIS	800
		OALH	361
	MARS	MC	200
		LH	200
		LPTAU	400
		METIS	800
		OA	361
		OALH	361
	GP	MC	200
		LH	200
		LPTAU	400
		METIS	400
		OA	361
		OALH	361
Variance	FAST	FAST	493
	SOBOL	SOBOL	400
	McKay	OA	400
		OALH	400

arbitrarily choosing a sample size and calculating descriptive and dispersion statistics. To neglect this previous analysis may lead to statistical pitfalls, such as results too uncertain to be reliable. Some other customary approach to determine minimum sample size may presuppose normality, and therefore its efficiency becomes sensitive to the shape of the distribution. This assumption is particularly relevant for the reason that ABMs and most real data often don't conform to parametric distributions. Moreover, as sample size increases, any theoretical distribution would likely be rejected.

Another interesting discovery found was that the definition of a sampling technique might alter the minimum sample size required to reach variance stability. The most common sampling approach involves a UA that summarizes the results of Monte Carlo simulation based on simple random sampling. We investigated one other scenario with quasi-random sampling and found that, for our particular case, the minimum sample size using random sampling is larger than the minimum found using a quasi-random sampling design. Similar findings were described in other areas of application, such as financial models (Niederreiter, Hellekalek, Larcher, & Zinterhof, 1998) and statistical circuit analysis (Singhee & Rutenbar, 2010). These results are in sync with the current trend of the use of quasi-random sampling in ABM (Ligmann-Zielinska et al., 2014; Saltelli et al., 2008), as it generates samples more uniformly over the parameter space.

Notwithstanding, in our investigation of SA techniques, we decided to test a broader combination of sampling techniques and sensitivity methods. This exercise is another guideline to be regarded, since there are sampling methods that best fit some SA methods and others that are inefficient or inappropriate. The design of the SA experiments must consider it to avoid perfunctory SA.

Very distinct results arise from the comparison of different SA methods in the Cerrado LUCC model. Not every method was able to identify the most sensitive parameters, such as the linear-regression-based techniques, SPEA and SRC, and the response-surface technique DT. For the most part, every other technique identified TA (Table 3.2) as the most critical parameter for all outputs, therefore answering the initial question of which parameters are responsible for most of the results' variability. Almost every technique also identified GE (input parameter 4) as an important parameter to most of the outputs. The most significant influence of GE is on the producers' accuracy, and in the pixel wrong change, right change, and wrong persistence. It is also clear across the different methods that TG and IE (input parameter 2 and 3) are entirely insensitive, hence not essential to explain the variability in the outputs.

These results show a positive correlation between input and output uncertainties and present consistency of the screening results and physical interpretations. Since GE and TA describe the amount of land transformation in a simulation, high values of these parameters will increase the model output values. GE is the most sensitive parameter, followed by TA. To understand and to reduce uncertainty within these two variables will, therefore, reduce the uncertainty of the simulation as a whole. GE represents the amount of land cover that is transformed by a group of human agents in a cell of the map. GE is a sensitive value as it indicates the voracity and velocity of the current land exploitation, which will directly affect the result of the simulation. GE was found as highly sensitive in every SA method. Therefore, this result proves that the model is coded in such a way that it behaves similarly to reality because the socio-economic groups responsible for large-scale cattle ranching and permanent agriculture are the principal driver of deforestation in the Cerrado (McAlpine, Etter, Fearnside, Seabrook, & Laurance, 2009; Smith, Winograd, Gallopín, & Pachico, 1998). SA is used to prove this similarity between our model and the observed drivers of change.

For qualitative SA methods, both linear-regression and gradient-based sensitivity were able to identify the non-significant parameters. Regarding the most important parameter, there are some discrepancies. We can highlight four findings. First, MOAT, MARS, SOT, and GP got similar results for most of the outputs. Second, SPEA and SRC presented very similar results, but differ from the other methods regarding TA and GE. We argue that traditional methods, such as correlation and regression analysis, are not suitable for nonlinear and non-monotonic problems like the MASE-BDI model. Third, the results from DT appear very different from that of other methods. The DT evaluation metrics were not able to screen the parameters correctly. Fourth, GP results were consistent in three of four input parameters. The divergences in the importance of GE may be attributed to the GP algorithm optimal configuration, but further investigation is required.

Regarding variance-based SA methods, the results were robust for all methods, indicating TA and GE the two parameters that explain almost all the output variation. Considering the FoM output, TA was responsible for over 57% of the output variation, followed by GE, that explains about 42% of the output variation. Both TG and IE combined are responsible for less than 1% of the variance. There is a consensus among variance-based results denoting that quantitative SA is more robust than qualitative SA. The divergences in qualitative SA may be explained by the use of heuristics to represent the relative sensitivity of the parameters.

For the SA comparison, the general finding on every approach is described. Moreover, the discrepancies and similarities of the related work (Table 3.1) are also summarized:

MOAT: The gradient-based SA technique was able to identify the elementary effects of the inputs correctly, and it seems to be ideal for screening purposes. The downside is that the interaction effects are not included. Gan et al. (2014) found similar results in a study case with three times more parameters. We were able to find consistent results with the minimum number of simulation runs, but Lilburne and Tarantola (2009) argue that the sample generation is not straightforward. A blind adoption of MOAT may not be representative since it is not a global SA practice.

Linear-regression: ten Broeke et al. (2016) and Lilburne and Tarantola (2009) agree that regression is a simple technique that can describe relationships, which yield insight into model behavior. The bad performance of the SPEA and SRC regression methods was also found by Gan et al. (2014), which may demonstrate that for these case studies, the regression model does not fit well to the particular ABMs.

Response-surface: These qualitative SA methods were very efficient to indicate the sensitive variables at a low computational cost (low number of runs). A discrepancy was found compared to the work of Gan et al. (2014). In the Cerrado LUCC model, the DT method performed poorly, while in the related work there were no such problems. On the contrary, Gan et al. (2014) discarded the use of GP because it was not able to find the sensitive parameters, a situation that did not happen in our study case. Response-surface methods are based on heuristics, and maybe these heuristics are more problem-specific, and a general guideline of use of any particular technique should not be endorsed before scrutiny.

Variance-based: The techniques with the higher computational cost were the ones with more consensus among them. They were all capable of finding the most sensitive parameters, and this result is corroborated by different works: (Gan et al., 2014; Lilburne & Tarantola, 2009; Saltelli et al., 2008; ten Broeke et al., 2016; Thiele et al., 2014a).

MC and LH were the sampling methods with better efficacy for qualitative SA methods, identifying the most sensitive parameters with a sample size of 200. All the quantitative SA achieved the same result with the sample size of 400. From the results, we can attest that qualitative methods are more efficient, i.e., find the sensitive parameters in fewer model evaluations. The main disadvantage is that there is no consensus among the methods, and in some cases, the resulting importance ranking of the parameters is quite the opposite. Fonoberova et al. (2013) argue that the use of surrogate models in ABMs may be an alternative to increase confidence in qualitative SA methods. Conversely, the results of all quantitative methods were broadly the same and the methods seemed more robust. They were all based on variance decomposition and were capable of computing parameter first-order effects, but it takes larger samples to do so. Quantitative methods, such as Sobol, are indeed more accurate, but at a higher computational cost, e.g.(Gan et al., 2014). For models with a larger number of parameters than the Cerrado LUCC model, one must evaluate the trade-off between accuracy and cost.

3.5 Conclusions

We investigated the various impacts that UA and SA experimental design have on ABM outputs. The results show that, although much of the analysis is problem-specific, there are known challenges that can be overcome by the use of statistical methods. Related work comparison illustrates general practices that should be a routine, both to improve the level of confidence in results derived from ABMs and to promote more rational and efficient use of ABMs. We suggest performing a specific investigation of the problem, aiming to test the robustness of the results. One should begin with an investigation of the number of simulation runs required to secure the stability of output variance, followed by a design of experiments selection (quasi-random sampling). It was clear that the quantity of samples has several ramifications to experimental design and the quality of the analysis. These steps must be done before UA. The results of UA should be explored in a global variance-based qualitative SA, such as Sobol.

We also investigated the impact that sampling techniques, sample sizes, and SA methods may have on the model output analysis. We identified the most significant and non-significant parameters of the MASE-BDI model. By applying gradient-based, variance-based, and linear-regression-based SA, we verified that TA is the parameter responsible for most of the variability of MASE-BDI results. Although the results were similar across the different SA approaches, they also showed that not any technique can be used without being tested and compared with others beforehand. Choice of analysis methods and sampling heavily impact model parameter sensitivities. Regarding ABMs, it seems that there is no single method able to embrace all models. The best-fit method is still dependable on the model and the goal of the experiment.

UA and SA were found to be essential tools for analyzing and evaluating ABMs, in particular in the LUCC context on the Cerrado LUCC model. Other than assuring the model predictions are correct, we believe those methods should be used for model corroboration to help researchers check, e.g., if the assumptions are fragile, if the inferences are robust, or if the variables are overly dependent. Regarding this matter, we implemented a comprehensive UQ through the integration of MASE-BDI and PSUADE. We were able to improve the Cerrado LUCC model factor prioritization setting, to identify which factor was most deserving of further analysis or measurement, and to assess the ABM parameter elasticity. As a future work, we are interested in identifying critical or otherwise interesting regions in the space of the input factors. Also, we search to uncover factors which interact, and which may therefore generate extreme values.

An ABM may be used for learning purposes, role-playing games, to understand the dynamics of a process, or to investigate different scenarios and configurations. Despite the research area, the number of parameters or the size of the model, there is room to apply UA and SA routinely, as a part of the modeling process or even in the model's operational use. It is time to make the methodology of agent-based modeling more robust and the analysis of results collected with ABMs more scientific. To this end, all expressions describing the systematic and methodological analysis of the responses and behaviors of the model, and the mapping between its inputs and its outputs (such as robustness checking, variability, UA or SA), are to be disseminated to the community and to be applied on a regular basis.

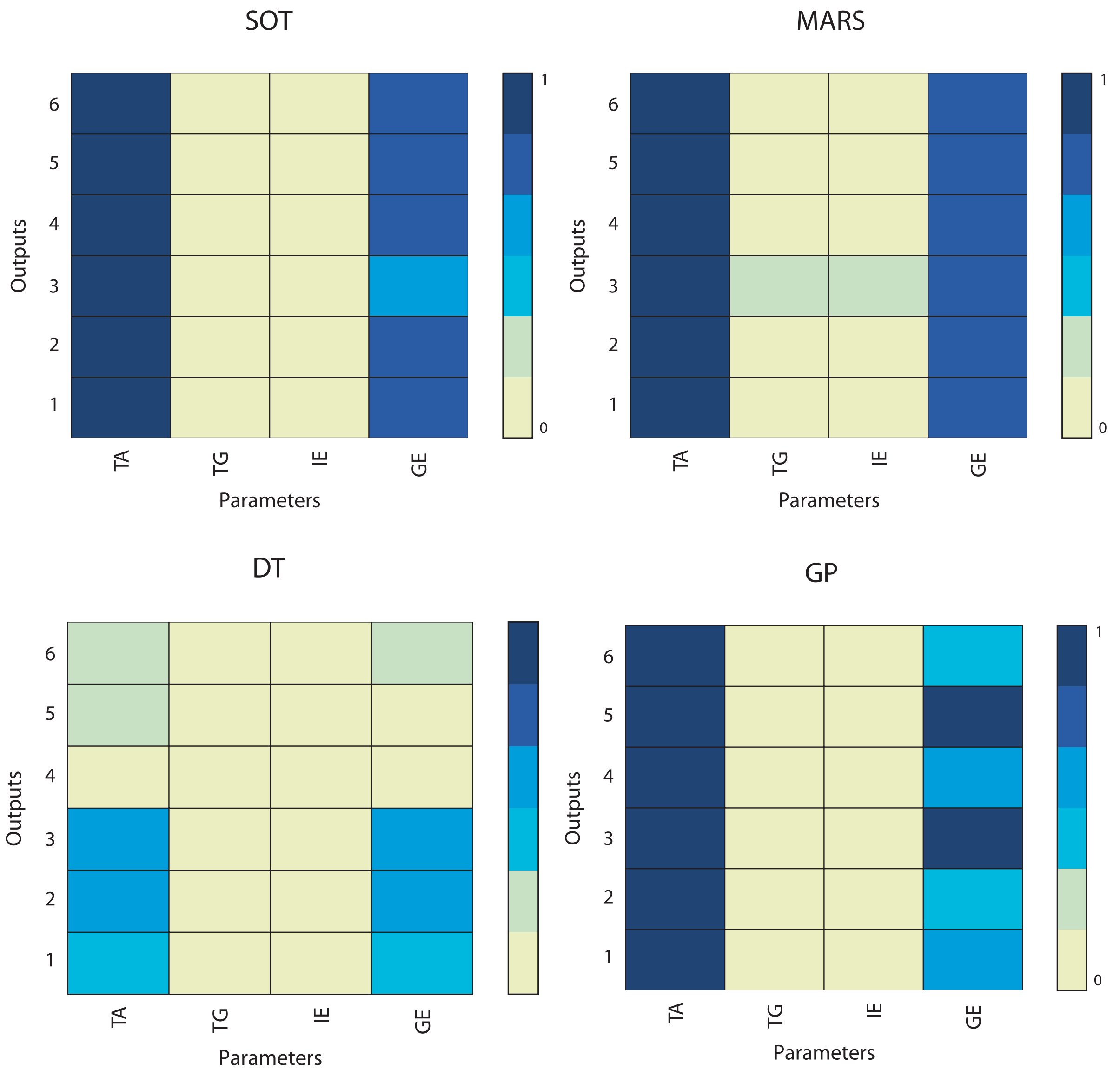

Figure 3.10: Heat map of response surface methods of sensitivity for MASE-BDI simulations, where TA - No. of Transformation Agents, TG - No. of Transformation Group Agents, IE - Potential of Individual Exploration, and GE - Potential of Group Exploration.

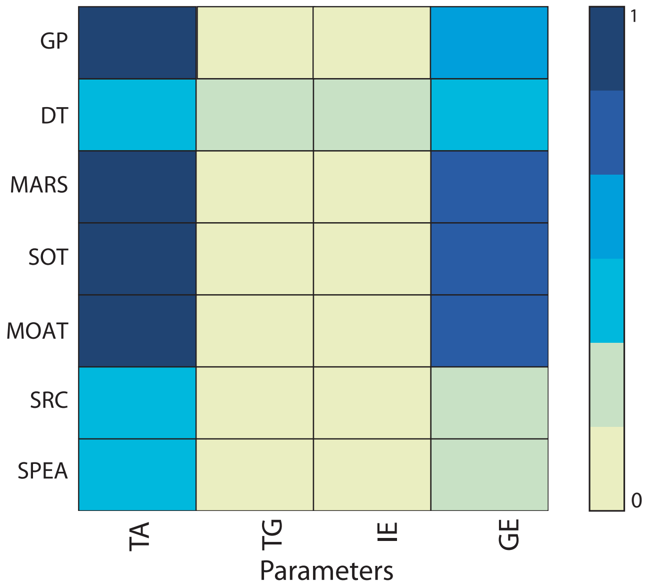

Figure 3.11: Heat map compilation of SA methods for FoM, where TA - No. of Transformation Agents, TG - No. of Transformation Group Agents, IE - Potential of Individual Exploration, and GE - Potential of Group Exploration.

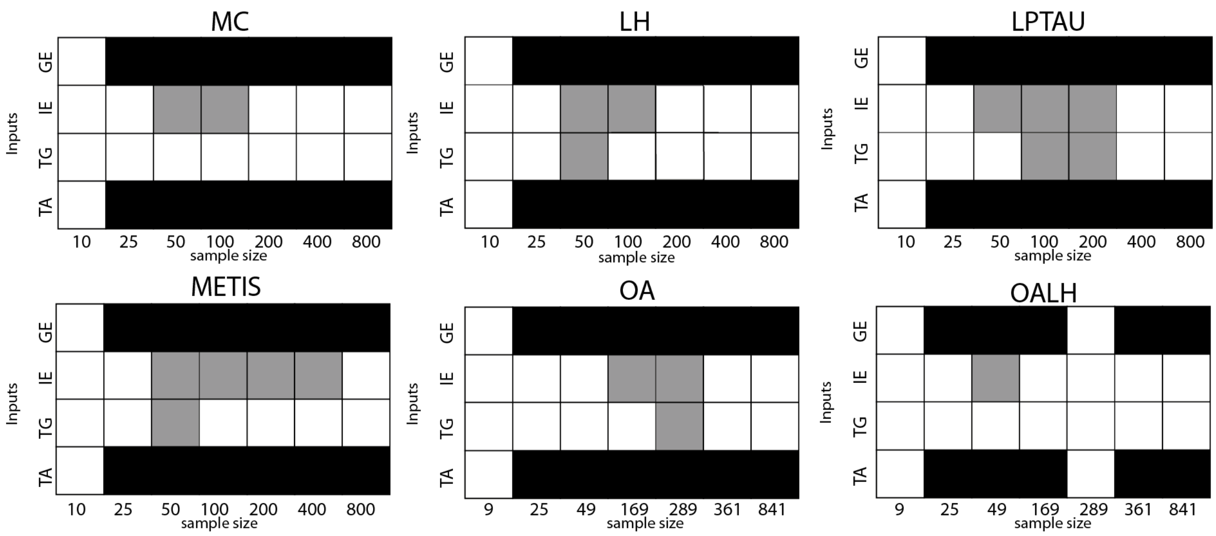

Figure 3.12: Comparison of different sampling methods and sampling sizes for MARS SA method.

Chapter 4

"Evaludation" of agent-based simulation output to improve analytical confidence

Full article under review in Journal Simulation Modelling Practice and Theory.

4.1 Introduction

ABMs are acknowledged for modeling complex systems, and simulations are commonly used to understand the dynamics and behavior of socio-ecological systems, such as LUCC. Realistic modeling and simulation of those systems must include the non-deterministic features of the system, i.e., the model must embrace the existence of uncertainty in the system or the environment, or human interaction with the system (Oberkampf, DeLand, Rutherford, Diegert, & Alvin, 2002).

Although ABMs provide a powerful tool for analyzing uncertain emergent phenomena, its utility is limited by difficulties in model analysis. ABMs simulations become rapidly complicated, what makes difficult to demonstrate the model is realistic and reliable. Significant drivers of this complexity are the number of factors, potential interactions between factors and possible non-linear effects (N. Gilbert & Troitzsch, 2005).

Rather often, ABMs are too complex and not at all appropriately validated to add value to informed decision making. Conversely, some ABMs are broadly applied without employing basic mechanisms of quality assurance(Grimm et al., 2014). These opposite realities stem from a not yet established culture of documentation, testing, replicability, and validation in ABMs. Even though almost all ABM and simulation review have expressed the need for statistical methods to evaluate the confidence of the results, these problems continue to be shortly tested and performed almost perfunctorily. One way to address these issues is standardization(Lorscheid et al., 2012).

Sensitivity analysis (SA) is referred to as a critical tool to help this type of model analysis because it quantifies the effects of changes in model parameters and inputs on the model predictions. However, existing methodologies of SA may be insufficient or not well-suited for a proper ABM analysis. Uncertainty Analysis (UA) is another set of methods that can be used to improve model legitimacy. Both SA and UA are closely related. Some authors such as Saltelli et al. (2008) suggest that the discrimination is that UA focuses on quantifying the uncertainty in the output of the model, while SA focuses on apportioning output uncertainty to the different sources of uncertainty (input factors).

UA and SA have been successfully used in tandem to simplify ABMs applications such as Ligmann-Zielinska et al. (2014),Fonoberova et al. (2013),Parry et al. (2013),Ligmann-Zielinska and Sun (2010b). The work of Ligmann-Zielinska et al. (2014) argue that any systematic evaluation of ABM uncertainty should meet three modeling objectives: i) the use of UA to evaluate the validity of simulation results; ii) the use of SA to generate a more parsimonious model; and iii) to prioritize input data refinement by identifying the ABM factors that are mostly responsible for model output variability.

The position paper of Hamilton, ElSawah, Guillaume, Jakeman, and Pierce (2015) describes a concrete advantage of this integrated assessment: to develop simplified or more computationally efficient versions of ABMs. Where the original model is complex, speeding up computation might allow more runs to be made to allow exploration of uncertainty, or might allow the model to be used in an interactive setting with stakeholders. The simplification might also help identify dominant characteristics of the system that are not otherwise obvious, or allow the efficient derivation of model properties, such as sensitivities to changes in inputs.

This work presents a systematic and standardized procedure for ABM research based on the model analysis workflow proposed by Abreu and Ralha (2018), composed of the design of experiments, UA and SA, focusing on their usefulness for the output analysis of LUCC ABMs. We applied those techniques in an LUCC ABM, on a particular case study of the Brazilian Cerrado. The results are simplified versions of the model, which can be used to explore model outcomes or conduct an exploratory analysis. Every step is documented for improving the effectiveness of communication, transparency, and reproducibility of our experiments.

For the sake of clarity, we do not imply that a simpler model is more likely to be true or get closer to the essence of the matter. In the interest of ABMs principles of model building, we seek model simplifications only if and when the model and evidence justify this. The simplification is grounded in objective principles such as the reduction of variability. We characterized this modeling approach as KIDS (*Keep It Descriptive, Stupid*) and defined by Edmonds and Moss (2005): we start with a straightforwardly descriptive model, based on evidence and resources, and then allows progressive development later (including simplification and abstraction).

Also, we contextualize the model analysis in a general framework for model "evaludation" (evaluation + validation) proposed by Augusiak et al. (2014), anchored on the modeling cycle. This new terminology describes the entire approach of assessing a model's quality and reliability. This framework proposes specific activities to document, check and verify each step of the design and simulation of a model. We focused on the last three stages of the evaludation process: model output verification, model analysis (based on the best-practices proposed by Abreu and Ralha (2018)), and model output corroboration. For each item, we provide the step-by-step of activities, applied to the case study model. We chose a framework of validation (catch-all term), so it is clear to decision-makers whether our model is a sufficiently good representation of our real system counterpart, and what criteria were used to answer this question. Therefore, we aim to provide enough information so that our model predictions could be more policy relevant.

In Section 4.2 we describe the UA and SA techniques, as well as the modeling and validation cycle considered in this manuscript. Also, we formulate and detail our integrated empirical proposal. Section 4.3 presents a portrait of the LUCC study-case. In Section 4.4, the evaludation of our framework is presented. In Section 4.5 we present a step-by-step view of the model simplification process and discuss our results. Finally, we conclude and present some future research work (Section 4.6).

4.2 Materials and methods

Every assessment of ABM output must begin with the definition of the quantity of interest, the output metric that provides insights about the model quality. In LUCC models, the metric is often related to the quality of the predicted maps generated through simulation. In this Section, we describe the methods that are going to be applied in the model analysis workflow, such as the output metric, the uncertainty analysis and the sensitivity strategies, and how they are integrated. We present our experimental design and provide context about how the model analysis should be understood as a task under a model evaludation framework.

4.2.1 LUCC goodness-of-fit metric

There are several verification techniques designed for spatial models. Social-ecological models need to be calibrated with spatially explicit data. Most spatial LUCC models use LUCC maps based on remote sensing as a starting point. We chose calibration tools that use aggregated values and spatial explicit validation methods, like the method proposed by Pontius et al. (2008). The authors developed several statistical LUCC indexes to determine accuracy (goodness-of-fit), including the *Null Model Hypothesis*, a reference for the LUCC model accuracy that corresponds to only persistence. Also a *Figure of Merit* (FoM), a ratio between correct predicted changes and the sum of observed and predicted changes. This methodology will be used in the steps four (model output verification) and six (model output corroboration) of the evaludation framework.

The underlying principle of those techniques (Pontius & Millones, 2011) (O'Neill & Niu, 2017) is the distinction between the quantity of change of a land use type and the location where these land use changes take place. The accuracy of the model is measured by the level of agreement between the reference (real) and the predicted (simulated) maps. The method compare: 1) a reference map of the initial time t_0; 2) a reference map of the subsequent time t_i; and 3) a prediction map of the subsequent time t_i. Those references and predicted maps are compared, pixel by pixel, and classified into percent correct and percent error.

These components allow the calculation of the FoM measurement that expresses the overlap between the observed and predicted change. This value ranges from 0 (no overlap) to 100 (perfect overlap).

4.2.2 Uncertainty analysis

From the modeling perspective, uncertainty is the lack of exact knowledge, regardless of what is the cause of this deficiency (Refsgaard, van der Sluijs, Højberg, & Vanrolleghem, 2007). One of the main sources of uncertainty are the model factors. Factors comprise various uncertain model components including variables, parameters, spatial data (maps) and functions, which often influence model behavior (Lorscheid et al., 2012). According to Saltelli et al. (2008), UA focuses on quantifying uncertainty in model output and usually precedes SA. Monte Carlo, based on random sampling, is the most common UA approach in ABMs.

In Abreu and Ralha (2018), we developed a baseline scenario of the same case study and performed a wide-ranging investigation of the impacts that differences in sample sizes, sample techniques, and SA methods may have on ABM model output. After a comprehensive study of the behavior of different sampling methods in the case study, we chose to use Sobol Extended (SOBOL) (Saltelli, 2002), which is a replicated version of low-discrepancy sequences (*quasi-random* samples). The SOBOL sampling strategy generates a uniform distribution in probability space, a qualitatively random distribution, filling previously unsampled regions of the probability function. This is done with two random $r \cdot n$ sample matrices M_0 and M_{n+1}, where r is the number of replications and n is the number of input factors. Therefore, the total number of sample points is $(n + 2) \cdot r$. The use of SOBOL is in sync with the current trend of use of quasi-random sampling in ABM (Ligmann-Zielinska et al., 2014) (Saltelli et al., 2008), because it generates samples more uniformly over the parameter space and comprises variation reduction techniques that artificially manipulate the sampling procedure.

ABMs are stochastic, and therefore the experimental error variance in estimation must be assessed as part of the model analysis. The stochasticity in model outcomes requires that any analytical exercise must be drawn from a sufficient number of samples. We adopted the concept of variance stability proposed byLorscheid et al. (2012) and Field and Hole (2003), where variance measures can determine the needed number of runs required per setting of a given simulation. We chose the coefficient of variation c_V as our measure and obtained 800 as the minimum sample size of our LUCC model for a determined quantity of interest (QOI) (Abreu & Ralha, 2018).

Selecting an appropriate sample design and the sample size is paramount since UA and SA are computationally expensive. Sampling methods provide a systematic exploration of the parameter space that guarantees the sample to have specific statistical or structural properties. The purpose of these methods is to reduce the number of parameter sets that are considered, but still chose space-filling points in the design space (Thiele, Kurth, & Grimm, 2014b).

4.2.3 Sensitivity analysis

SA consists of studying the effects of changes in the input on the output of a model. We adopted the application goals for SA which are common for ABM research, as proposed by Broeke, van Voorn, and Ligtenberg (2016): 1) to gain insight in how patterns and emergent properties are generated in the ABM; 2) to examine the robustness of emergent properties; and 3) to quantify the variability in ABM outcomes resulting from model factors.

Uusitalo, Lehikoinen, Helle, and Myrberg (2015) argue that the fundamental purpose of SA is to alter model input of the model and study the subsequent changes in model output. If the output values change little, the output is robust to changes in QOI within the model. It can indicate that the uncertainty about the QOI is relatively small. Conversely, if QOI changes markedly when factors change within their reasonable range, then it is a sign that there is substantial uncertainty about the variable's value (Uusitalo et al., 2015).

There are various methods of SA, and each one has advantages and limitations. In the particular case of SA in spatial ABMs, we incorporated the general guidelines provided by Lilburne and Tarantola (2009) and Fonoberova et al. (2013). We have already tested and compared different SA methods, as presented in Abreu and Ralha (2018). Following these results, we selected the SOBOL variance-based global SA method (Sobol', 1993).

The SOBOL method decompose the output variance $V(y)$ that assumes that the input factors are independent, hence, model free,

$$V(y) = \sum V(i)_i V(i) + \sum V(i,j)_{i,j} + \cdots + V_{i,j,\ldots,m}, \tag{4.1}$$

where the partial variance is defined as

$$V_i = V_{x_i}(E_{x_{-i}}(y|x)), \tag{4.2}$$

with x_i denoting all parameters except for x_i. If V_i is large, the expected model outcome strongly varies depending on x_i, indicating the factor to be sensitive. Sensitivity indices are defined by considering the partial variance relative to the total variance,

$$S_{s,i} = \frac{V_i}{V_y}. \tag{4.3}$$

The first-order index represents the main effect contribution of each input factor to the variance output. The total effect of a variable would be the total contribution to the output variation, that is its first-order effect plus all higher-order effects due to interaction. Higher-order sensitivity indices are defined by computing the partial variance over two or more parameters instead of a single parameter.

4.2.4 Integrated assessment of UA and SA

The coupled use of UA and SA has many objectives and has been successfully applied in different context in ABMs through the literature (Abreu and Ralha (2018), Abreu and Ralha (2017), O'Neill and Niu (2017), Fonoberova et al. (2013), DeJonge, Ascough, Ahmadi, Andales, and Arabi (2012), Ligmann-Zielinska and Sun (2010a), Crosetto, Tarantola, and Saltelli (2000)). We chose to employ Ligmann-Zielinska et al. (2014) quantitative UA-SA systematic evaluation of ABM uncertainty to meet three modeling objectives: 1) The use of UA to evaluate the validity of simulation results; 2) The use of SA to generate a more parsimonious model; and 3) to prioritize input data refinement by identifying the

ABM factors that are mostly responsible for model output variability (using both UA and SA).

In this framework, UA is applied to check the variability of the results in a stochastic baseline model (Figure 4.1). Therefore it is possible to improve model rightfulness, where the distribution of results informs the expected value validated against independent data, the variance around the mean and the extreme results. The SA is then applied to indicate which factors are responsible for the variability of results in two different set of experiments: *exploratory* and *explanatory*. Both are simpler versions of the baseline ABM.

In the *exploratory* experiment, the input space is restricted to the inputs that produced the most of the variance of the baseline ABM, creating a practical model with output distribution similar to the initial model. The benefit of this experiment is the possibility to simulate low-probability, but high-consequence events that may be of high policy relevance. In the *explanatory* experiment, the framework proposes the refinement of the most influential input value, resulting in a model that is less spread but preserve the mean of the output. Ligmann-Zielinska et al. (2014) argue that to improve model performance and provide a scientific explanation it is necessary to reduce output variability to achieve the necessary accuracy. This explanatory analysis would expose the smallest number of inputs influencing the steady state of the modeled system. To explain (different from predict) itself is a reason to model (Epstein, 2008), because it could bring to light the system-wide regularities which manifest themselves through the mean of the output of interest.

4.2.5 Evaludation of environmental models

Evaludation is the terminology proposed by Augusiak et al. (2014) to describe the entire process of assessing a model's quality and reliability. It is based on the modeling cycle, and it is composed of six fundamental steps: 1) data evaluation; 2) conceptual model evaluation; 3) implementation verification; 4) model output verification; 5) model analysis; and 6) model output corroboration. A simplified representation is presented in Figure 4.2.

Data evaluation is a critical step for scrutinizing the quality of numerical and qualitative data used for model development and testing. It includes the data used to parametrize the model via calibration, to define the conceptual model, to design the model structure, to formalize expert knowledge in probabilistic if-then rules, among others. Data is a significant source of uncertainty, and therefore data themselves do not always represent the real system sufficiently well. As Augusiak et al. (2014) evince, "a model cannot be expected to provide more accuracy and clarity than what has been used to develop it in the first place".

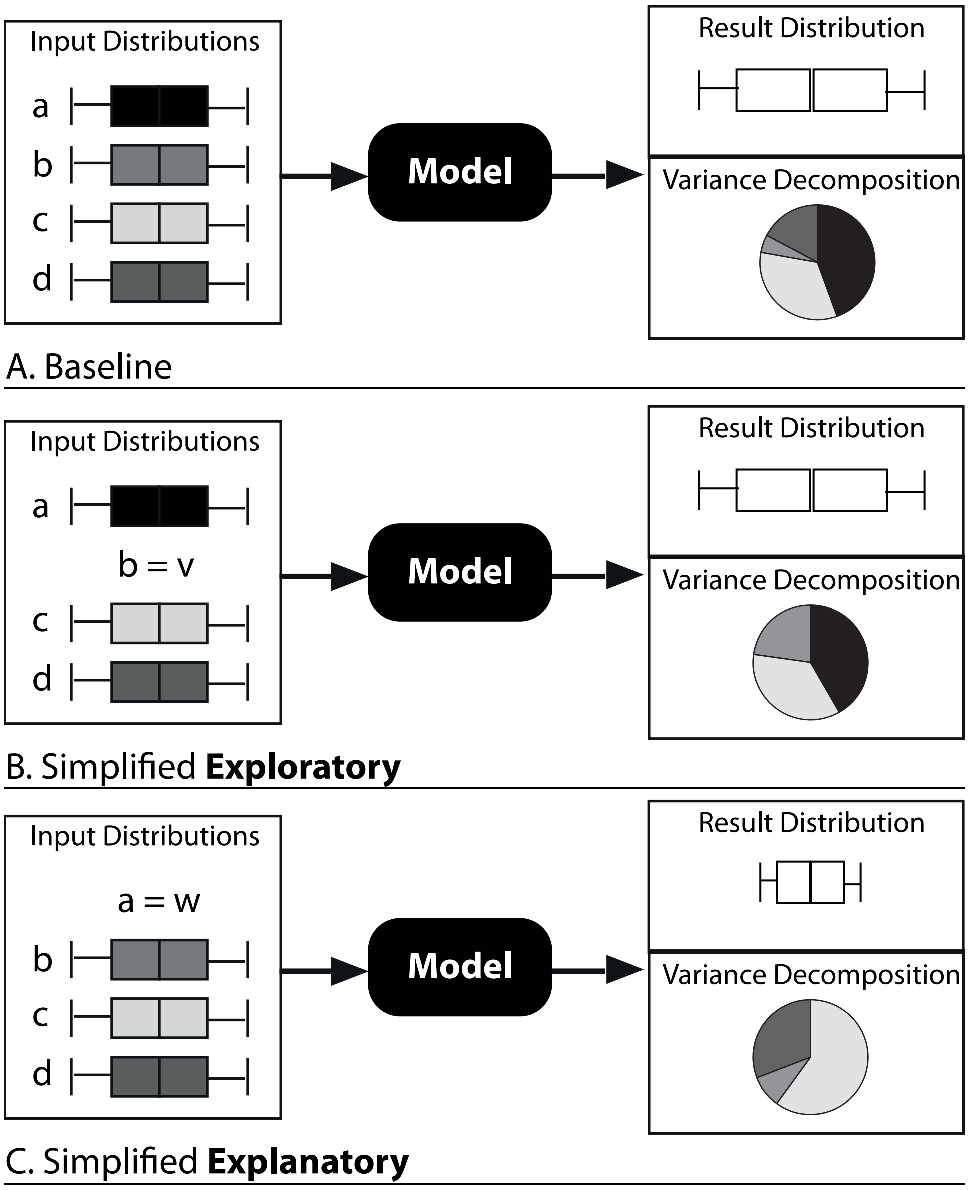

Figure 4.1: Framework for coupling uncertainty and sensitivity analysis of ABMs. Experiments to apply variance decomposition to (A) simplify a baseline stochastic model, and (B) to maintain its exploratory power embodied in outcome variability or (C) to improve its exploratory power by reducing its outcome variability. Source: (Ligmann-Zielinska et al., 2014).

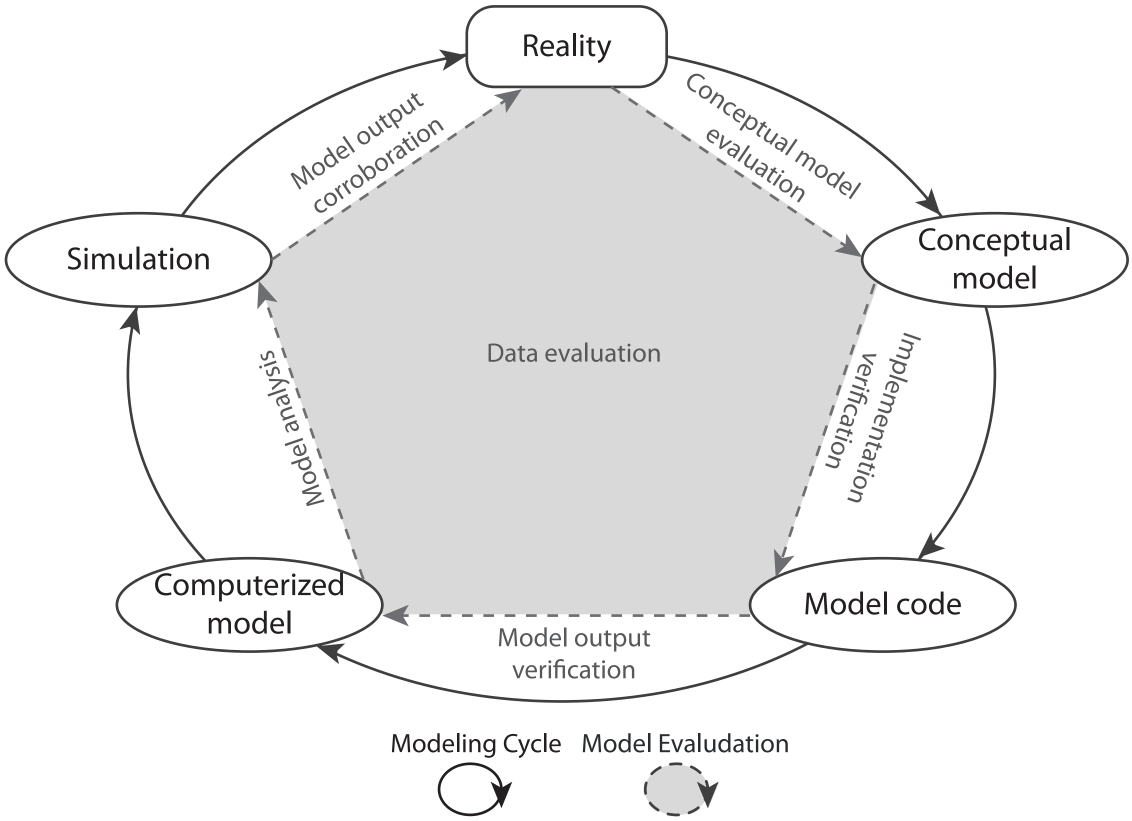

Figure 4.2: Representation of the evaludation steps of model development proposed by Augusiak et al. (2014). The modeling cycle presents the terminology for model quality assurance and it is an adaptation of the work of Refsgaard and Henriksen (2004) and Schlesinger (1979).

Conceptual model evaluation is the step created to examine the simplifying assumptions underlying a model's design. The assumptions include the spatial and temporal scales, the choice of environment, entities and processes to be represented, and even definitions about the stochasticity and interactions. The conceptual model is prone to bias due to the modeler subjectivity, judgment, and lack of awareness. The third evaludation step is the implementation verification. It concerns to test the model's implementation in equations and as a computer program. This element is concerned not only in checking for code errors and bugs but also for detachment due to vagueness in the model description.

The model output verification is an assessment of "how well model output matches observations" for a model is to be a good representation of the real system. However, researchers should be aware of what degree calibration, initial states of the model, and data sampling were involved in obtaining good fits of model output and data. Model analysis is the fifth step and regards the exploration of the sensitivity to changes in the computerized model parameters. Also to make sure that the emergence results, produced by the behaviors and processes of the model, were understood. Finally, model output corroboration is responsible for comparing the model outcome, often predictions, to independent data and patterns not used in the model conception and calibration.

Augusiak et al. (2014) propose this set of terms and "quality assessment" processes to ensure the reduction of avoidable uncertainties, to establish a control framework of the model, to improve communication (to peer researchers, decision-makers, non-technical audiences), to promote transparency of the capabilities/limitations of a model, and to raise the confidence of the model's results. However, the authors highlight that it is not possible to create a fool-proof protocol considering the complexity of environmental issues.

4.2.6 Proposal

We apply the general evaludation process proposed by Augusiak et al. (2014) to promote transparency and to improve the overall quality of the simulation results. We focused on presenting the details of the verification, model analysis and model output corroboration steps of the evaludation process. The overview of each step is presented in Figure 4.3.

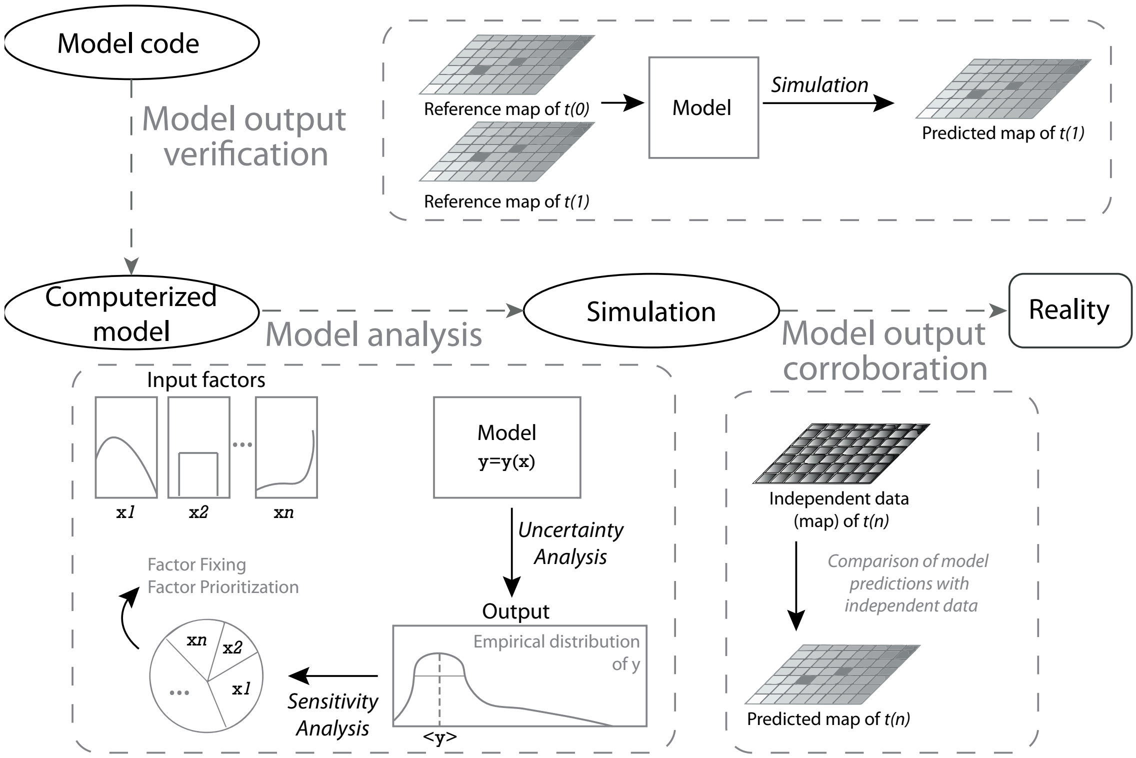

Figure 4.3: Integrated uncertainty and sensitivity assessment applied to the evaludation steps within the modeling cycle.

We used the most efficient sampling strategy, UA and SA methods for our specific land use study case (Abreu & Ralha, 2018). These methods were applied in a UA-SA integrated assessment of an LUCC case study. As proposed by Ligmann-Zielinska et al.

(2014), we seek to build two simplified and more computationally efficient versions of our ABM. The exploratory experiment provides the opportunity to investigate extreme system behavior. The explanatory experiment improves model performance and provides scientific explanation necessary to reduce output variability and improve analytical confidence.

Every step is applied aiming a more robust and concise model, focusing the reduction of variability within the ABM outputs. What sets apart this scientific contribution is that this simplification is focused on the reduction of variability of initialization configuration of ABM simulation. Each step of the evaludation is documented, as well as each step of the integrated assessment. This way we can demonstrate the robustness of the ABM simulation outputs.

4.3 ABM land use case study

An overview of the case study is provided so the reader can understand the ABM and its results. We will focus only on the ABM initialization variables as factors in the experiments. The description of the parameters and its impacts will be restricted to this dimension of uncertainty.

4.3.1 MASE-BDI computation modeling platform

Many environmental ABM simulation tools perform land change using the agent's approach, but few are using rational agents. Considering that agent's cognitive reasoning and decision making can be executed within the Belief-Desire-Intention (BDI) model (Bratman, 1987) the options are even fewer. Thus, this work uses the Multi-Agent System for Environmental (MASE)[1] simulation tool (Ralha et al., 2013) which was extended by introducing rationality to agents with the BDI model resulting in the MASE-BDI (Coelho et al., 2016). MASE-BDI allow multiple types of agents with different behaviors to represent the interactions and relations between agents and the physical environment considering spatially explicit models in the context of land change. A complete methodological description of MASE is available in Ralha et al. (2013). In Coelho et al. (2016), the MASE-BDI implemented architecture with the description of the agents' reasoning model and an auto-tuning module is presented.

[1]MASE Project Website: `http://mase.cic.unb.br/`

4.3.2 LUCC model description

The MASE-BDI LUCC model is a socio-ecological ABM with the purpose of exploring how the land cover is affected by external disturbances such as the individual behavior of agents and changes in land use policies and regulations. It is a spatially explicit model where the real landscape is represented by a set of geographic information system (GIS) derived maps. This model has a hybrid framework because it allows researchers and stakeholders to explore land change from the emergence of individual decision-making (farmers and ranchers will be based on the BDI mentalistic approach) and from a top-down perspective (regional spatial planning). The LUCC model presented herein is committed to the Transparency and Openness Promotion (TOP) guidelines (Nosek et al., 2015) and all the model code, maps and data are available for reproducibility[2]. This paper provides an overview of the conceptual model. For a full description of the model in the ODD protocol (Grimm et al., 2006) for ABM communication, readers can refer to Ralha et al. (2013).

Figure 4.4 presents the structure of MASE-BDI conceptual model using a UML Class diagram with properties/attributes sit at the top and methods/operations at the bottom. Note that the SimulationManager, SpatialManager, TransformationManager, and TransformationAgent inherit from BDIAgent through an implementation relationship. The FarmerAgent and the RancherAgent implement the TransformationAgent through a generalization relationship being an individualAgent or a GroupAgent. The SimulationManager instantiates the SpatialManager and TransformationManager. The SpatialManager manages the simulation GRID that contains Proximal Matrix. The GRID and Proximal Matrix contain Cell (composition - each simulation Cell has an instance of the GRID and Proximal Matrix). The GRID can call Proximal Matrix's properties or methods. The TransformationManager implements the conflict resolution of the TransformationAgent's, while the TransformationAgent checks the Proximal Matrix attributes before movement. The TransformationAgent occupies and transforms the Cell's (aggregation), while the TransformationManager instantiates and manages the TransformationAgents (composition).

The land cover change result from the emergence of the individual decision making of the ranchers and farmers. Each step of the simulation corresponds to a week in chronological time. The basic spatial unit is a plot, representing $1ha$ of the GIS map. During the model setup, the simulation GRID is loaded with the reference map of the initial time t_0, and a set of GIS layers, representing the environment such as hydrology (lakes and rivers), landscape, railways, highways, slope, streets and buildings, environment protected areas, territorial zoning maps, etc. The sum of the physical layers creates a proximal matrix that

[2]Model availability: `https://gitlab.com/InfoKnow/MASE/MASE-BDI/SourceCode`

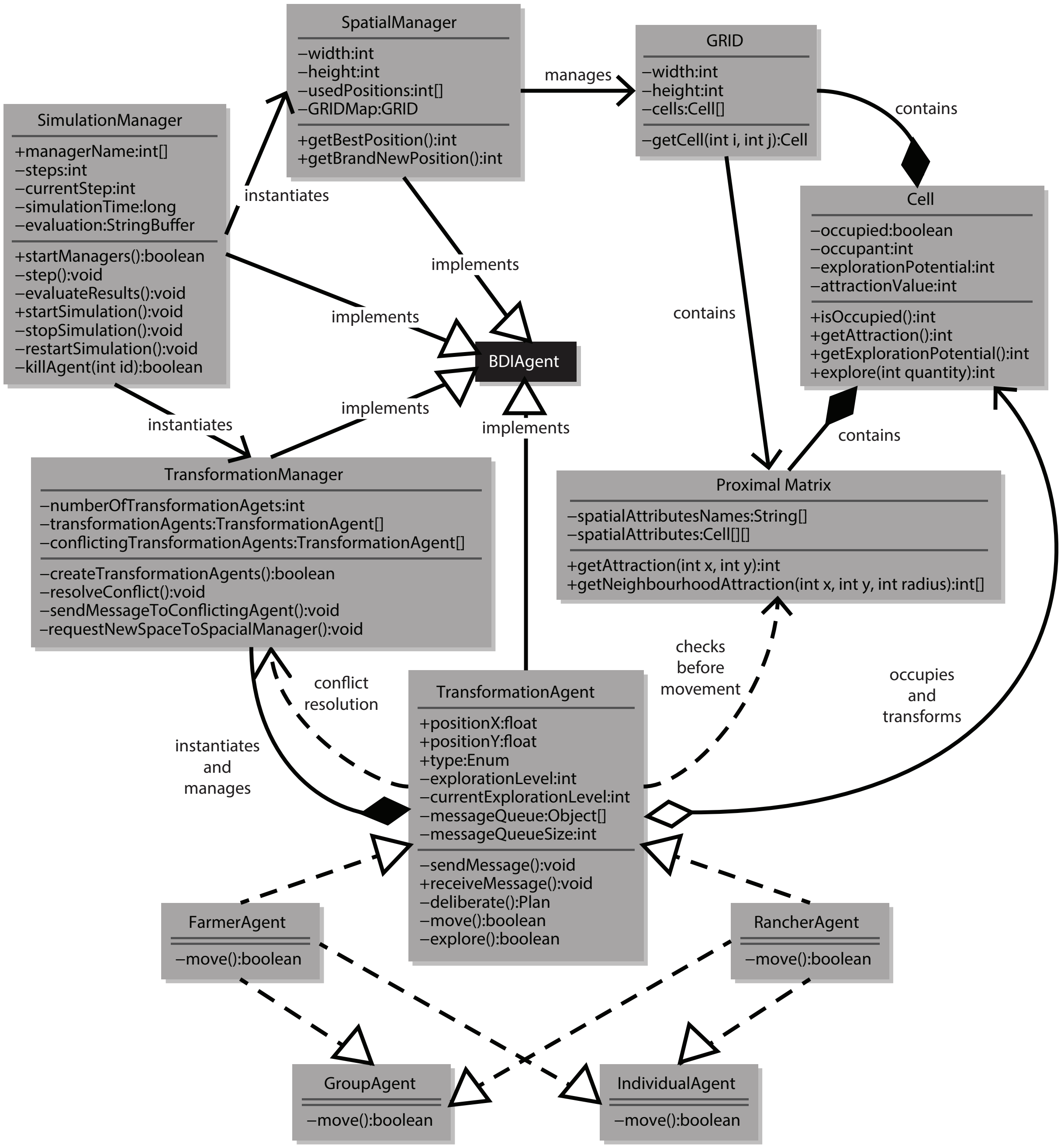

Figure 4.4: MASE Class Diagram

is perceived by the agents and is part of their beliefs. The farmer and rancher agents (TA - Transformation agents) are associated with various socio-demographic and economic factors (capacity of exploration, capacity of production, land tenure) and assigned to a plot.

A simple activity diagram for the TA is presented in Figure 4.5. A first step is to be assigned to a plot, where the agent may choose to explore the land or move to a more attractive plot of the neighborhood. If there is competition, the conflict is brought to a solution by a higher entity. TAs have their behavior and beliefs, explicitly changing the natural landscape to achieve their internal goals.

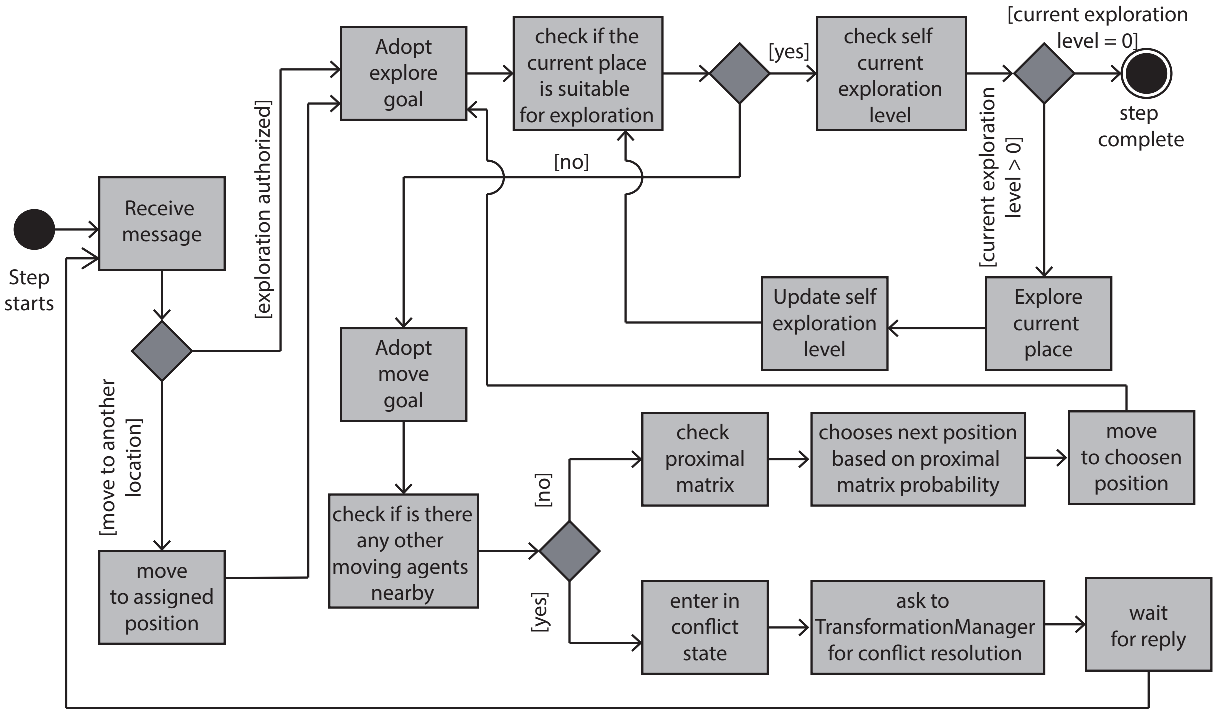

Figure 4.5: Transformation Agents Activity Diagram.

4.3.3 Case study: ABM of the Cerrado Federal District anthropic land use

Brazil's Cerrado is the country's second-largest biome, and the most bio-diverse and threatened savannah on the planet. This biome has already lost 48.2% of its original vegetation cover and is being affected by an intense process of habitat fragmentation. The high rates of vegetation loss and deforestation are attributed to unsustainable agricultural activities such as soy production and cattle ranching. This over-exploitation poses a continuous threat to numerous animal and plant species, especially to an estimated 20% of endemic species.

The Federal District is the only Brazilian state that has its territory entirely covered by the Cerrado biome. All of its $5,789 km^2$ territory is inserted in the Environmental Protected Area (EPA) of the Central Plateau, as presented in Figure 4.6. Therefore, the Federal District Spatial Plan must comply with various environmental management guidelines, from the federal, regional and local governments. This overlapping in attribution creates a peculiar scenario, in which farmers and ranchers receive multiples incentives and penalties depending on the land use, the specific area of the territory and the scale of the land exploration.

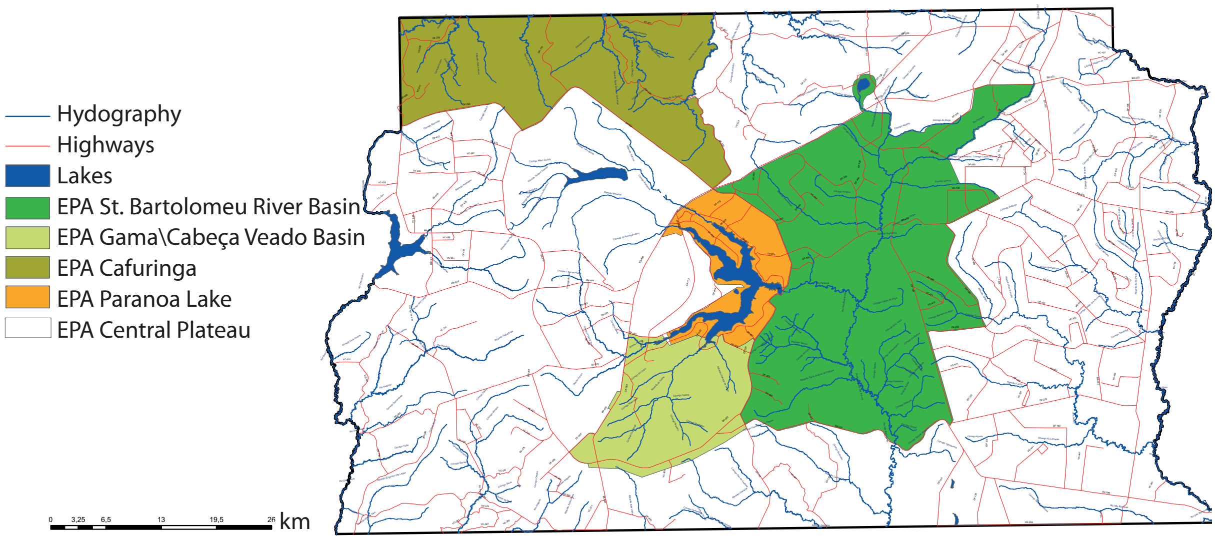

Figure 4.6: Federal District Environmental Protected Areas. Source: Adapted from (IBRAM, 2014).

The case study considers the participation of farmers and ranchers as land transformation agents that move and explore the land within the DF territory. The transformation agents have different beliefs, desires, and intentions and may comply with the given DF spatial plan.

A random distribution of agent behaviors usually initialized ABM simulations. In some cases, empirical data should be used to bring the model closer to reality. However, Grow and Van Bavel (2017) argue that one of the primary challenges of ABM initialization of the population of agents is that there is no set of data that contains every possible behavior. Therefore, researchers must elaborate a strategy to initialization, that can be purely random, utopian data-driven or, in most of the cases, something in between. In this case study, there are four initialization factors, presented in Table 4.1, that are responsible for a significant portion of output variability (Abreu & Ralha, 2018). We start with a simple random initialization and will adjust the range of the factors based on the feedback provided by the evaludation process. This way we can use empirical data not only for the verification but also for the initialization.

Table 4.1: Input factors of the MASE-BDI simulations of the land use case study.

Factor	Description	Distribution	Range
TA	No. of Transformation Agents	Uniform	1 - 100
TG	No. of Transformation Group Agents	Uniform	10 - 100
IE	Potential of Individual Exploration	Uniform	1 - 500
GE	Potential of Group Exploration	Uniform	500 - 1500

4.4 MASE-BDI LUCC model evaludation

We use model evaludation to improve the overall confidence of the models' results. The following items describe a summary of the steps that were performed in the evaludation process. The focus of this manuscript, model verification, analysis, and corroboration, will be detailed in Section 4.5 and the predictive modeling capability of the MASE-BDI LUCC model will be discussed.

Data evaluation: The LUCC model was calibrated to experimental data. Also, the available data for the parametrization of the model parts were taken from peer-reviewed literature and expert interviews. The empirical characterization of agent behavior was performed according to the Smajgl et al. (2011) methodology and is described in details in Ralha et al. (2013). We used parallel auto-tuning algorithms to evaluate the search space of over six million parameter combinations, and quickly tune the simulation model, regardless of the QOI used (Coelho et al., 2016). However, spatially explicit sets of data are scarce and are available in different temporal and spatial resolutions. We had to manually transform the spatial data into the same scale and group it in the same temporal window. Qualitative observed patterns were also used to design the overall model structure.

Conceptual model evaluation: The design and assumptions of the LUCC model simulated in MASE-BDI model are built over an existing model of the dynamics of the agricultural frontier in the Amazon and savannas of Brazil (Cerrado), proposed by Smith et al. (1998) and presented in Ralha et al. (2013). The conceptual model design is described on the ODD (Overview, Design concepts, and Details) Protocol, designed by Grimm et al. (2010) to standardize the published description of ABMs.

Implementation verification: In order to verify and guarantee that the model code works according to the ODD model description, we performed a series of code checks, unitary tests, and compilation tests. Moreover, visual testing through MASE-BDI interface was carried out. The computational efficiency was verified with stress tests, with extreme parameters values.

Model output verification: In this study, we performed calibration of the initialization parameters to optimize the FoM goodness-of-fit metric to our initial data set. We adopted the terminology proposed by Trucano, Swiler, Igusa, Oberkampf, and Pilch (2006), where calibration ultimately is an optimization under uncertainty problem. Therefore, we formulated the calibration problem to explicit acknowledges model uncertainty. We adjusted the set of parameters associated with the model code so that the model agreement is maximized to a set of experimental data. The spatial explicit index metrics such as the null hypothesis and FoM were our QOI, i.e., the calibration considered not only the quantity of land use change but also the allocation of change in the spatial grid. Each step of the model output verification is described in Section 4.5.1.

Model analysis: Although there is a relatively high computational time for each simulation, a comprehensive SA was performed. The sensitivity of the model outcomes was evaluated in a set of simulations covering different sampling strategies, sample sizes and SA methods (Abreu & Ralha, 2018). In this manuscript, we conducted an SA to explore the behavior of the model regarding the simulation initialization parameters, i.e., factors that were not directly determined from the literature. The model analysis is presented in Section 4.5.2 and has two main objectives: understand the emergence of model outputs to produce a simplified and computational efficient version of the model (exploratory and explanatory).

Model output corroboration: It is hard to find data that can be used to corroborate model results, given the spatiotemporal resolution of the LUCC simulations. Only recently the Brazilian Environmental Ministry published data from different years that could finally be used to corroborate the model. The experiments are presented in Section 4.5.3.

4.4.1 Design of experiment

We designed three sets of experiments to verify and analyze the model: a baseline experiment, an exploratory, and an explanatory experiment described as follows.

- In the baseline experiment, we run 138,800 runs using all four factors;

- In the exploratory experiment (1680 runs), we performed SOBOL SA and included the factors that highly impacts the FoM output;
- In the simplified explanatory experiment (1680 runs), we perform a variance reduction by fixing the most influential factor from the baseline experiment, leaving the remaining factors unchanged.

Finally, a corroboration experiment was performed. All simulations were run using high-performance computing at the University of Brasilia. Factor samples were produced using the quasi-random SOBOL experimental design. The sample size considers the variance stability for $E = 0.001$, as presented in Section 4.2.2. SOBOL S and ST indices, as long as all the UA were calculated using an integrated implementation of MASE-BDI and PSUADE software package (Tong, 2005).

4.5 Results and discussion

The results of our ABM simulations are land use maps and a set of calculated metrics, such as FoM, a goodness-of-fit metric. The spatially-explicit output is a simulated raster map of the predicted LUCC change, illustrated in Figure 4.7, an example result from the baseline experiment. The different colors show the predicted land use cover produced by simulation. Figure 4.8 represent a summary of the goodness-of-fit of the simulations runs. As proposed by Pontius et al. (2008), each bar is a rectangular Venn diagram where the solid and cross-hatched central segments represent the intersection of the observed change and the predicted change, while the central solid black segment is the change that the model predicts correctly. When $FoM > 50\%$, it means the amount of correctly predicted change is larger than the sum of the various types of error, and the model is more accurate than the null model.

Figure 4.8 exemplify some contrasting results due to the variation of initialization parameters. In Simulation $n^{o}55$ (Sim55), a result obtained from TA=55 agents, FoM is 55 whereas Simulation $n^{o}5$, a result obtained from TA=5 agents, FoM drops dramatically to 19. It is clear that this is still not a well-calibrated model. A brief analysis would show that variation in the initialization of the simulation may result in radical changes of significant consequence to the simulation results. We use UA-SA integration to clarify this results and focus on the causes of variability.

4.5.1 Model output verification

To investigate what is the variability of results we performed UA in our baseline sample to examine the distribution of the QOI. Table 4.2 summarizes descriptive statistics for FoM,

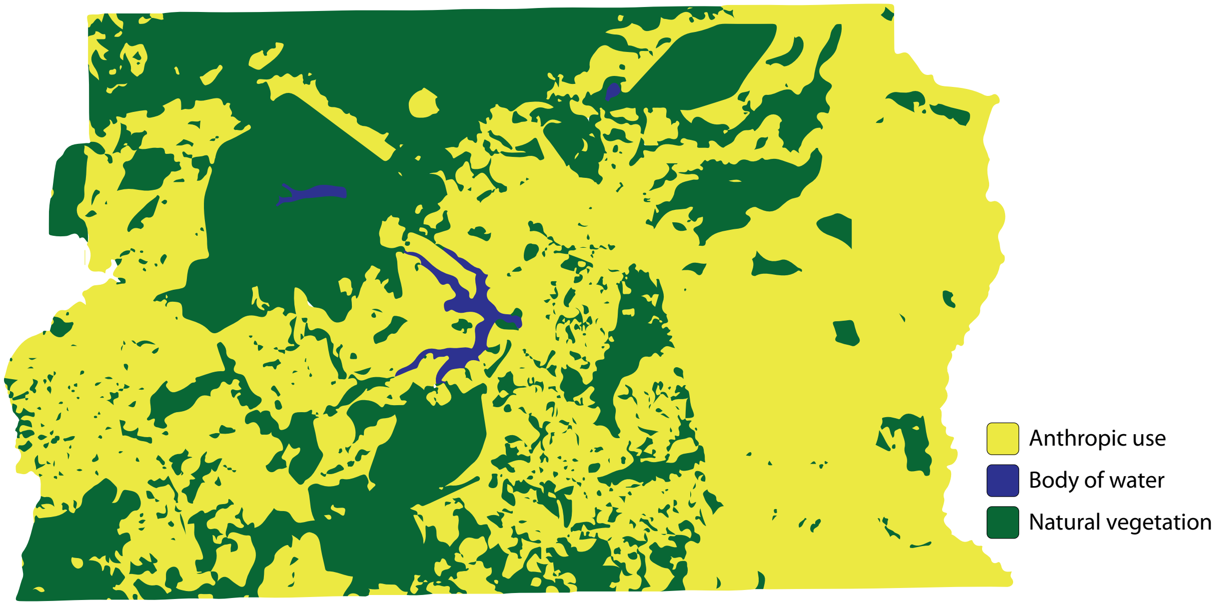

Figure 4.7: A land cover predicted map produced from a simulation on the Cerrado LUCC model in Brazilian Federal District, showing the changes from the year 2002 to 2008.

the moments and its errors. The first moment is the mean, denoted by $\mu = EX$. The second central moment is the variance. The third moment, or skewness (γ) is the measure of the lopsidedness of the distribution. Kurtosis, the fourth central moment is a measure of the heaviness of the tail of the distribution, compared to the normal distribution of the same variance. Figure 4.9 summarizes the empirical density of the FoM output on the baseline experiments simulations.

Table 4.2: Uncertainty Quantification: moments results of MASE-BDI model's FoM.

Mean (*Error*)	Variance	Skewness	Kurtosis
41.58 (*0.57*)	268.17	-1.34	3.25

Still, regarding the UA, the null hypothesis ($H_0 = \mu$ are equal) would confirm that all ABM representations are equivalent, but the experiments were considered being significantly different from any other (one way ANOVA ($F(15, 64125) = 625$, $p = 0$, $p < 0.05$)). This result refutes the assumption that all ABM representations were equivalent. However, UA alone does not provide the influence of the individual factors on the accuracy of the final map. We were interested in knowing the influence of each factor on the FoM variability. Figure 4.10 shows simple representations of pie charts of the S and ST indices for Sobol.

It is clear that TA is the most relevant factor that influences the output of the model, followed by TG. We needed to calibrate the input factors aiming to reduce the sampling

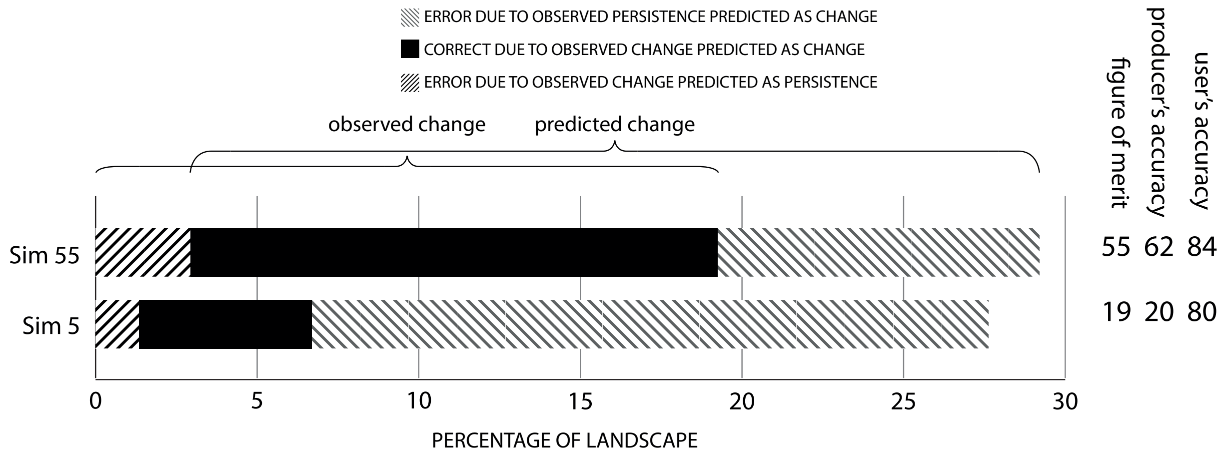

Figure 4.8: Sources of percent correct and percent error in different runs of the MASE-BDI simulations.

variance. Thus, we used an empirical approach to calibration as an optimization problem. We generated a quasi-random sample and performed a total of 138,800 simulation runs to adjust the range of the input factor. To further investigate the TA factor, we produced a scatter plot (Figure 4.11 for the visualization of the relationship between the FoM (x axis) and the number of agents, TA (y axis). It is possible to see that there is much noise in FoM when TA$<$ 40. The same observations were generated for the visualization of TG. The maximization of FoM was considered the optimization function, and we generated a set of range restrictions on the input factors, as presented in Table 4.3.

Table 4.3: Initialization configuration parameters post-calibration.

Parameter	Description	Distribution	Lower bound	Upper bound
TA	No. of Transformation Agents	Uniform	40	80
TG	No. of Transformation Group Agents	Uniform	10	100
IE	Potential of Individual Exploration	Uniform	1	500
GE	Potential of Group Exploration	Uniform	400	1000

Table 4.3 presents the experimental design henceforth referred to as the baseline experiment. The limitations imposed in the inputs initial range are presented in Figure 4.12. The distribution of the FoM after the calibration is presented in Figure 4.13. Table 4.4 summarizes the moments for the baseline experiment. Now the results of no simulation run is significantly different from any other (one way ANOVA ($F(1, 1185) = 0.746$, $p = 0.39$, $p > 0.05$)). We can confirm that in the baseline experiment all ABM representations are equivalent given the significance level of 0.05.

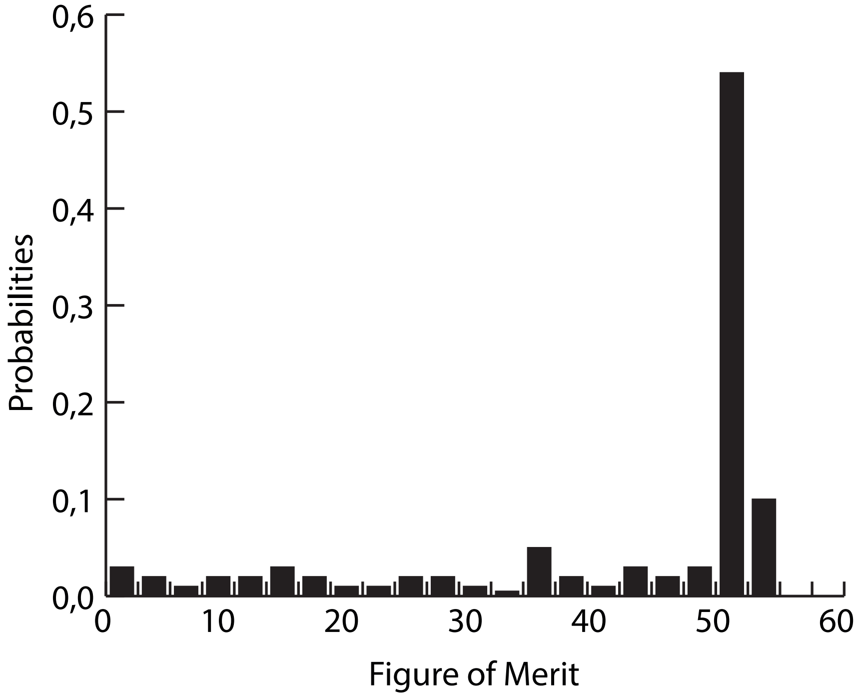

Figure 4.9: Empirical density of the baseline experiment simulation' results of the FoM distribution.

Table 4.4: Moments result for the baseline experiment, considering the calibrated input factors.

Mean (*Error*)	Variance	Skewness	Kurtosis
51.91 (*0.57*)	0.99	0.98	3.06

4.5.2 Model output analysis

The model analysis aims to explore the sensitivity to changes in the computerized model parameters and includes a description and a justification of the scenarios explored. We used the variance to evaluate FoM variability, and the results show that the variance of the second and third experiments are approximately equal (Table 4.5).

Table 4.5: Uncertainty Analysis: Means and Variance of Figure of Merit

	Mean	Variance
Experiment 1: Baseline	51.91	0.99
Experiment 2: Exploratory	51.98	0.97
Experiment 3: Explanatory	51.03	0.34

The first SA experiment is the simplified exploratory, in which the input factors with little or no influence on the variance decomposition are fixed. After our baseline SA analysis, we chose to fix the input parameter TG= 45 and IE= 250, the mean values for those factors. Due to their influence, TA and TG were not changed. Since FoM is

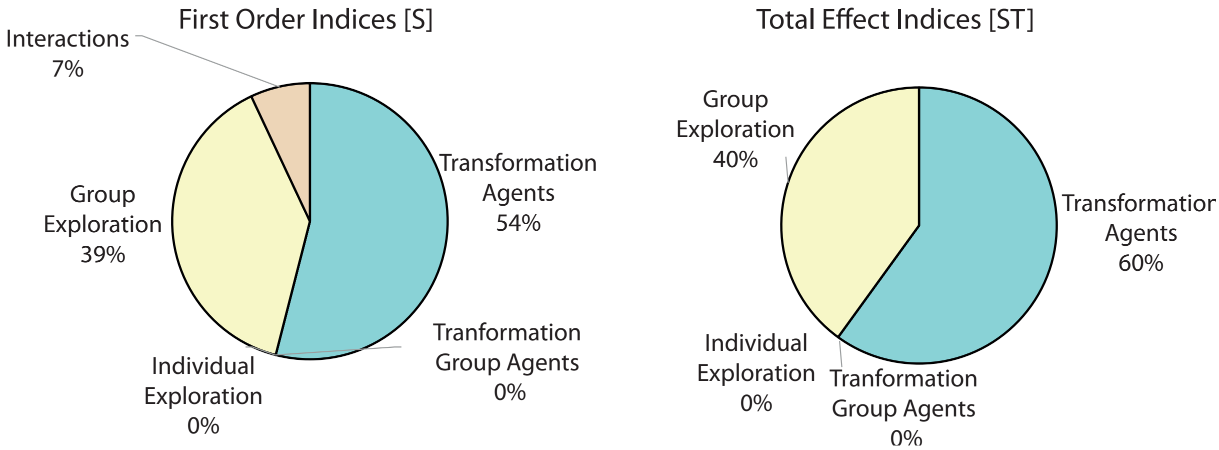

Figure 4.10: Sensitivity analysis of FoM output in the baseline experiment.

almost insensitive to variations in TG and IE, the fixation of those factors has almost zero influence in FoM output distribution. The results of the SA are presented in Figure 4.14. In fact, the baseline experiment and the simplified exploratory experiment distributions are nearly identical, including their means and variances. Also, the variance decomposition generated S and ST indices consistent with the baseline model.

The second scenario is the simplified explanatory, in which the most influential input factor is fixed. We set TA=55 (an arbitrary choice based on the best FoM). The results of the SA are presented in Figure 4.15. In this experiment, we want to explore how our ABM behaves when we fix the most sensitive initialization parameter. The results show that the mean is roughly the same, but the dispersion around the mean highly decreases. Also, because we fixed the most sensitive input factor, SA shows that only part of the variance decomposition can be apportioned to individual factors.

4.5.3 Model output corroboration

Model output corroboration is responsible for comparing the LUCC model predictions to independent data not used in the model conception and calibration. This step of the model evaludation is only possible because a new set of data of the Federal District land cover was released in 2017 by the Brazilian Environmental Ministry (Brasil, 2015), composed of maps and satellite images of each year varying from 2009 to 2015. Before this official release, the only available data was from 2002 to 2008, and it was what we used to test, verification and calibration.

We run our baseline experiment from 2009 to 2015. The results show FoM$= 51,84$, with a total of $462,76ha$ of new anthropic land cover change, added to the original map. The external data from 2009 to 2015 reported nearly $4km^2$ of anthropic land change in

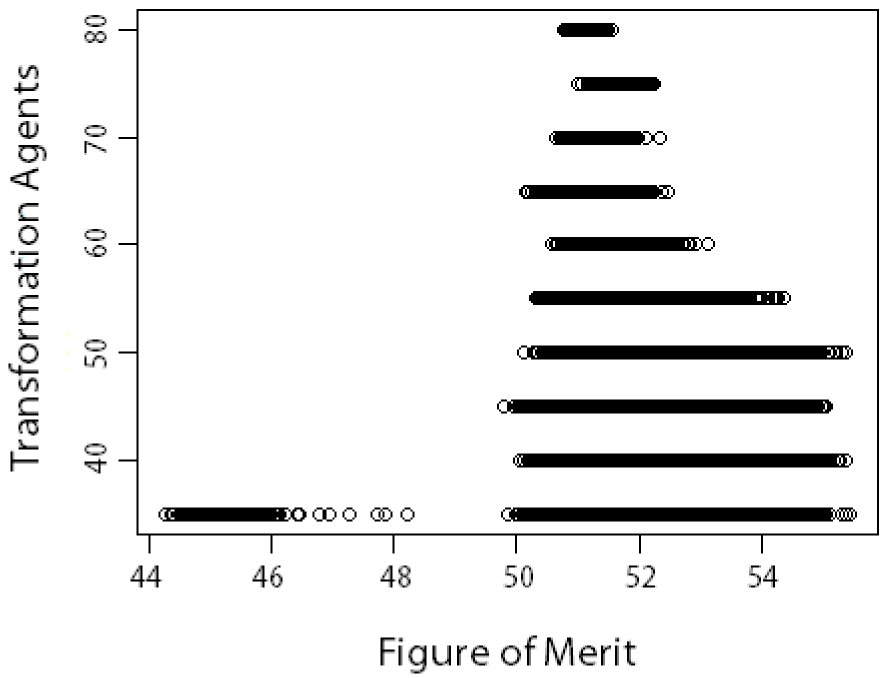

Figure 4.11: Scatter plot of the relationship between FoM and TA in 138,800 simulation runs.

the study area. Considering the simulation results of $\mu = 4.62 km^2$ of land cover change, we have that the simulated result is about 15% higher than the reported government land cover change. Since our experimental design uses a more uniform (quasi-random) sampling, we can infer that the calculated mean of land change is indeed the true (accurate) measure of central tendency. Therefore, we considered that the independent data promoted the model output corroboration.

Figure 4.16 illustrates the analytical procedure of map comparison proposed by Pontius et al. (2008) to corroborate the allocation of changes in the land use. In the map (a), we examined the difference between a reference map of 2009 and the reference map of the year 2015. In the map (b) we examined the difference between the reference map of 2009 and the prediction map for 2015. We wish to investigate whether the model predicts the land changes accurately. If the model were to predict the observed change correctly, then figures (a) and (b) would be equal. Finally, figure (c) examines the difference between the reference map of 2009 and the prediction map of 2009. Most of the error is location disagreement, which occurs primarily because the model predicts land change at the wrong locations.

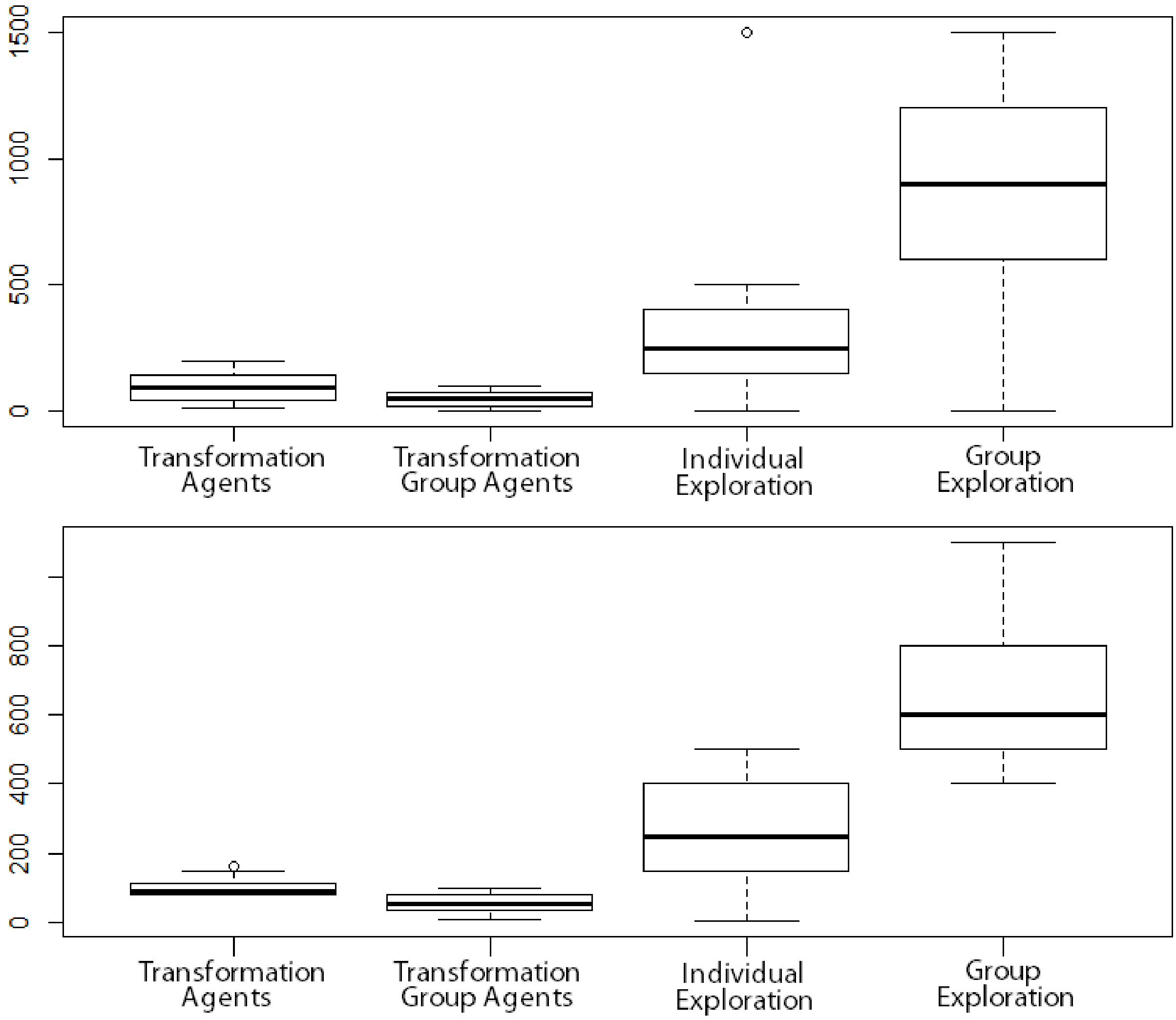

Figure 4.12: Distribution boxplots of the input factors (top) before the calibration and (bottom) after the calibration process.

4.5.4 Discussion

The UA results show that a change from factor TA=55 to TA=5 can raise the error due to observed persistence predicted as change in up to 290%. This radical changes of significant consequence to the simulation demonstrate the importance of assessing ABM initialization. Moreover, the choice of the objective function or QOI in an ABM can have a great impact on the identifiability of model parameters. i.e., the optimum QOI may not be given by a maximum in the parameter space, but rather by a complex interaction structure, in which many different combinations of the parameters are equally able to provide best fitting model simulations (Saltelli et al., 2004). Although the initialization in ABM can be tackled by gathering experimental data, the modeler cannot build a complete, exact image of a real system and has to simplify some processes and representations.

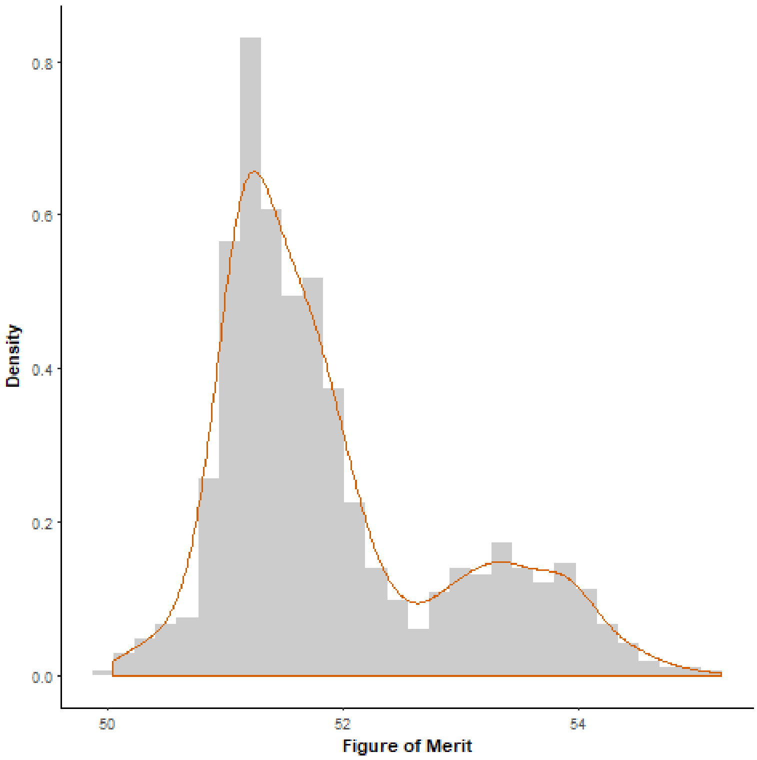

Figure 4.13: Density plot of the distribution of FoM outputs in the baseline experiment.

There will be underdetermination of the model due to epistemic uncertainty but also by the amount and quality of data. According to Cobelli and DiStefano (1980), the only way to uniquely identify model parameters, the number of conditional equations derived from applying a model to a dataset has to be higher than the number of parameters, and there must be sufficient variation in observations.

The UA results in the baseline experiment disclose another critical issue on ABMs: overfitting. In an overfitted model, the factors are chosen and calibrated to reproduce also the deviations present in the dataset, leading to an optimal fit in the calibration dataset, but deteriorating prediction in other situations (Forster, 2000). Modelers have to deal with the trade-off between aiming for a perfect fit and the risk of deteriorating predictive capacity for other samples. Zucchini (2000) argues that modelers should understand calibration as a problem of maximizing the expected accuracy of prediction for any sample, rather than finding an optimum fit to observed sample. It is necessary for the ABM for LUCC community of researchers to discuss what is an expected accuracy in a set of measurements and maps.

Regarding the SOBOL sensitive indices, S and ST, the variance decomposition results for the exploratory experiment is consistent with the baseline experiment. This simplification is more computationally efficient and could lead to a complete model analysis. It is also worth mentioning that variance remains almost the same between the baseline

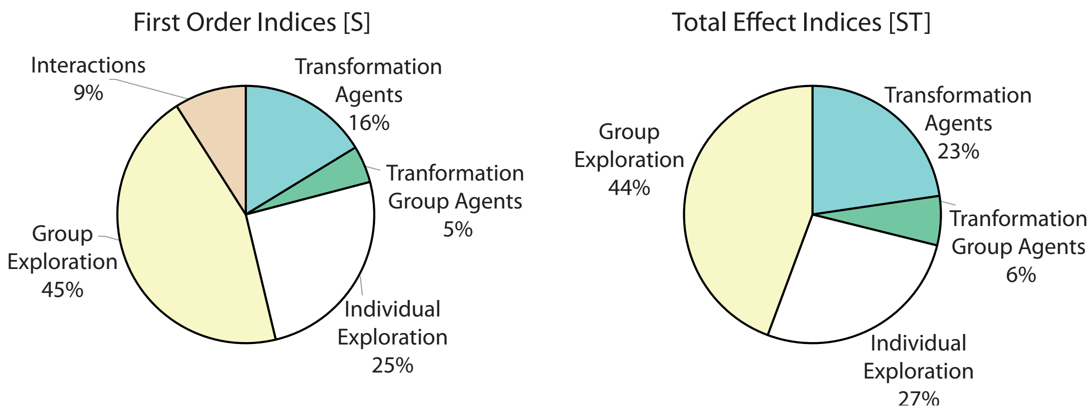

Figure 4.14: Sensitivity analysis of FoM output in the simplified exploratory experiment.

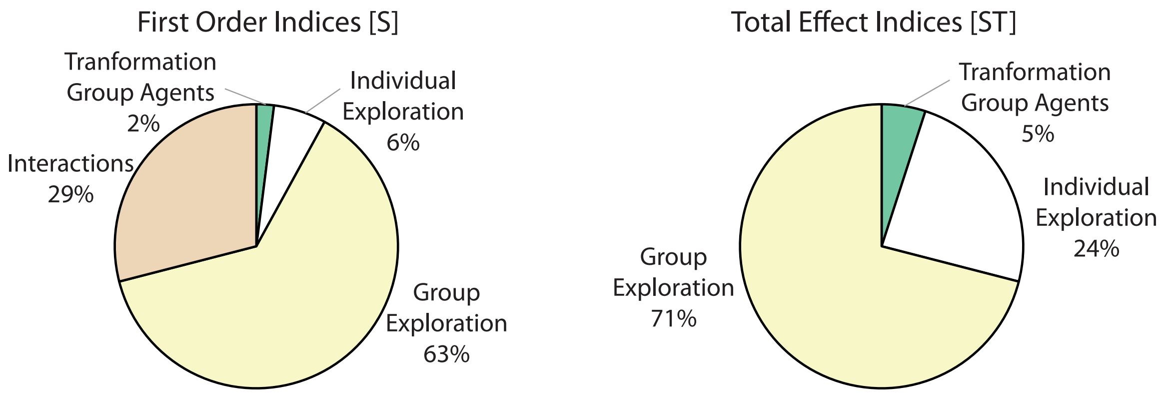

Figure 4.15: Sensitivity analysis of FoM output in the simplified explanatory experiment.

and the exploratory experiments. We can infer that the exploratory simplification of our model can be used in analysis without the loss of variability necessary when evaluating LUCC policies. This simplification maintains the resulting variability and therefore can be used to identify less probable but highly consequential policy scenarios, as shown by Ligmann-Zielinska et al. (2014).

On the contrary, the explanatory simplification version of the model maintains the same mean but reduces the variability. This is a consequence of the refinement of data because the most sensitive factor was fixed. The benefit of this approach is to mimic a scenario in which we obtain exceptionally accurate data for the most sensitive factor. It could be used to analyze the behavior and interactions of the other variables, and raise our understanding of other social and ecological processes of the LUCC region dynamics.

We agree with the assumption that says that it is best to reveal the complexity of a problem through the simulation instead of through the model structure. These simplifications may provide a more robust and concise model, focusing the reduction of variability

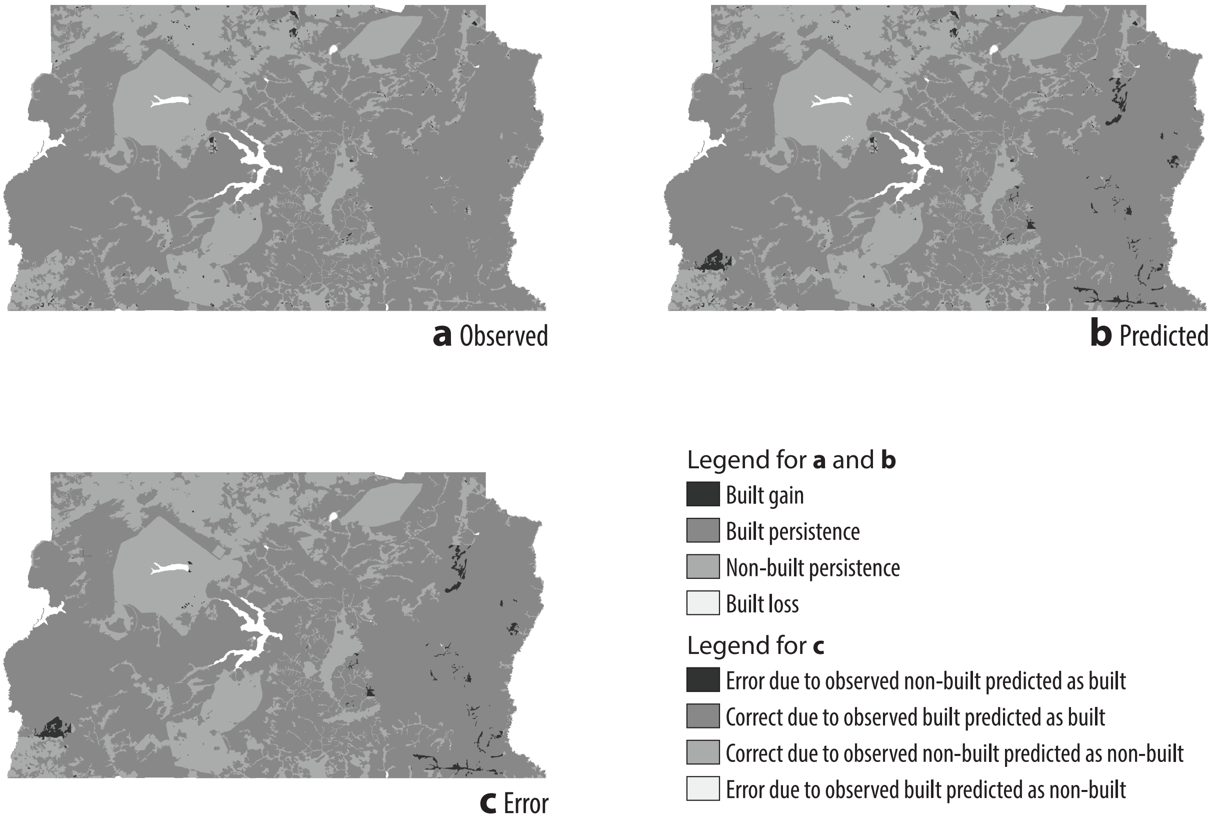

Figure 4.16: Sources of percent correct and percent error in different runs of the MASE-BDI simulations.

within the ABM outputs. One of the contributions of this manuscript is to show a real application of the use of objective principles such as the reduction of variability to simplify the model when data justify this approach. To focus on the initialization configuration of an ABM within an evaludation process may also help other researchers that face this common challenges.

Regarding model output corroboration, it is clear that just the mere fact of comparing model outputs to independent new data is neither sufficient nor necessary to make a model more useful to policymakers and to conclude to its validity. However, is one more step towards a more reliable model and predictions.

4.6 Conclusions

Despite the limitations and even though the presented analysis was done over a particular simulator, we conclude that important feedback can arise from the application of a broader evaluation process to improve the level of confidence in ABMs simulation outputs. The transparency of the sound statistic tests may contribute to a systematic treatment of uncertainty and better modeler-user communication.

Researchers for descriptive and predictive purposes have used ABMs but still, have limited use in policy-making. This may be explained by the lack of confidence in the accuracy of predictions. This UA-SA integrated assessment, applied within an evaludation framework is an effort to open the ABMs "black box", to make the predictions more transparent and to improve analytical confidence. This approach serves as a tool for better-informed ABM building and using of its results. Output uncertainty can be reduced if we can improve the quality of the data on the most sensitive factors. There are limitations to this approach, such as the choice of a QOI, in our case, the FoM goodness-of-fit metric. The investigation of another output could alter the results and the most influential factors. Also, changes in the distribution of the input factors may also result in different relative contributions to the outputs. This will lead to a future investigation of the output space. We are also interested if we can make our simplified versions of the model pass the"falsifiability" test. We will test different theories to see if there is any failure in the expected basic patterns of the model. Moreover, the study of changing landscape patterns involve calculating the indices for images of a landscape taken at several different times in history and then observing how these indices vary over time. As of today, the information on the landscape at each step of the simulation is not persisted in the MASE-BDI framework. LUCC ABMs based on cellular-automata usually persist this information. We will consider refactoring the code to gather this information for further analysis of the spatial and temporal complexity.

Chapter 5

General Discussion

ABMs are favorite for modeling complex phenomena. However, the credibility, and utility of ABMs are hampered by the lack of model analysis, transparency, and reproducibility of ABMs. It is partly due to the non-existence of a fit-all methodology for model validation. Also, there is a lack of experimentation, supported by a wide range of arguments. Some arguments suggest that experimentation is too difficult, useless, or that it cost too much. By the number of tests, simulations, experiments, and runs that were performed in this thesis, one can understand why there is so little experimentation in ABM community.

A first question to be answered is: ABMs have to be validated at all? Only if the answer is yes, we can argue that there is not enough of model analysis. Refsgaard et al. (2007) argue, regarding Popper's scientific, philosophical school, that models cannot be verified or validated. Despite the terminology, we do not seek for absolute certainty, but to consider the conclusions as admissible. Balci (1998) defines validation as substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives.

All of those insights must be the result of different experimentation. Concerning that topic, Tichy (1998) proposes an exciting discussion. The author states that no amount of experimentation provides proof with absolute certainty. However, experiments must be used for theory testing and exploration. To cite another computer scientist, Mr. Dijkstra (Dijkstra, 1970), an experiment can only show the presence of bugs in a theory, not their absence. Therefore we advocate that quantitative analysis of model outputs is mandatory to probe the influence of model assumptions, to understand model results, to ensure repeatability, and to raise the credibility of ABM as a science.

The usefulness of simulation models is limited by the ability of the modeler to demonstrate the robustness of the model results. OAT analysis is the most popular SA technique used in ABM. In Chapter 2 we were able to find the most influential factors but at a high uncertainty. The uncertainty was so high that the confidence intervals indicated that most of the results did not meet the minimum goodness-of-fit criteria. This experiment should suffice to indicate that the current level of experimentation on ABMs is not enough. This kind of shallow analysis does more harm than good. Modelers should be aware that even a widespread technique should not be applied without questions. The proposed UA-SA workflow register a sequence of steps that must be assessed in any model analysis: What is the point of variance stability? Which is the best sampling strategy? What is the variability of my results? Which SA measure should I apply? What factors are responsible for most of the variability of the output? These questions can help to disseminate the proposed workflow and evaluation guidelines.

Another common argument is that this type of comprehensive investigation cost too much, regarding time or computational resources. One could argue that more costly is to publish a paper with unvalidated claims. The review works (Angus & Hassani-Mahmooei, 2015; Heath et al., 2009) show us that most modelers are publishing untested frameworks, and this is one of the reasons ABMs are continuously criticized. The results of this thesis do not suggests that every ABM idea must be experimented, but testing can help build a reliable base of knowledge and reduce uncertainties. Also, testing can lead to unexpected insights and quickly eliminate fruitless approaches and erroneous assumptions (Tichy, 1998). Researchers should probe the importance of the research question. Besides, the insights gained from previous experiments are availed in the next iterations. All of the experiments that we had to undergo in Chapter 3 gave us the understanding of how MASE-BDI works regarding the factors under investigation. In Chapter 4, most of the initial investigation was reused, therefore reducing the so-called high cost.

The discourse that ABM complexity is so high that a researcher may lose track of how the model works can also be debated. If there are too many variables to control and too much uncertainty, is more of a reason to execute those disciplinary experiments. We agree with Tichy (1998) when he states that eschewing experimentation because of difficulties is not acceptable science. However, it is essential to have in mind that experiments are always be flawed in some way. Experiments may be based on unrealistic assumptions, researchers may manipulate the data, or it might be tough to quantify the QOI. Despite these problems, the flaws may be reduced by a description of robust experimentation.

5.1 Contributions

In this thesis, we achieved the proposed objective of evaluation of the application of several methodologies of uncertainty quantification in the ABM output analysis. We performed an integrated application of UA and SA techniques and evaluated the impacts that differences in sample sizes, sampling techniques, and SA methods may have on model output. To summarize, we highlight the following contributions of this thesis:

- Important feedback can arise from the application of a broader evaluation process to improve the level of confidence in ABM simulation outputs;
- The empirical workflow can promote transparency and sound statistics tests, that may contribute to a systematic treatment of uncertainty and better modeler-user communication;
- UA-SA integrated assessment as a communication tool can open the model "black-box";
- Parameter and Methodological uncertainty can effectively be reduced by the application of these guidelines;
- The validation and model output corroboration of the Cerrado LUCC MASE-BDI model is an important tool for understanding land-use dynamics and for policy decision-making in Brazil.

In our proposed empirical workflow to perform model output analysis, we organized a set of tasks under a macro prism of validation/evaluation of an ABM. The application of this workflow can be generalized and applied in all ABM, because the tasks and steps may be used as a guideline to assess the uncertainty of any kind of model. We advocate that the model may be evaludated, but only about site-specific applications and to pre-specified goodness-of-fit criteria, limited in terms of space, time, boundary conditions and types of application. The elaboration of this workflow aims to improve the quality of ABM studies by reducing the gap between the perceived need to improve ABMs credibility and the lack of commonly agreed modeling guidelines.

UA and SA were found to be essential tools for analyzing and evaluating ABMs, in particular in the LUCC context on the Cerrado LUCC model. Other than assuring the model predictions are correct, those methods were used for model corroboration to help researchers to check if the assumptions were fragile, if the inferences were robust, and if the variables were overly dependent. Regarding this matter, we implemented a comprehensive UQ through the integration of MASE-BDI and PSUADE. We were able to improve the Cerrado LUCC model factor prioritization setting and to create simplified scenarios to explore different parameter space regions. We also created a version of the model that helped us to explain the behavior of the system under pre-defined variance restrictions. All of the results were analyzed and validated by specialists.

We also reflected that most experiments relies on a single figure of merit. The investigation of another output could alter the results and the most influential factors. Also, changes in the distribution of the input factors may result in different relative contributions to the outputs. Although we are aware of the limitations, we think we benefit either way. An interesting definition is attributed to Enrico Fermi: "there are two possible outcomes: if the results confirm the hypothesis, then you have made a measurement. If the result is contrary to the hypothesis, then you have made a discovery". Following this logic, in either case a conclusion can be made. New conclusions come from the experiments on new ouput metrics.

Finally, an important property of good models is simplicity. A good model does not just define new useful quantities. It also leaves out many useless ones. Note that we are saying simple rather than simplistic. We agree with the KIDS methodology and will create simplified versions of the model only if the data or the results explicitly points in that direction.

Despite the limitations and even though the presented analysis was done over a particular simulator, we conclude that important feedback can arise from the application of a broader evaluation process to improve the level of confidence in ABMs simulation outputs. The transparency of the sound statistic tests may contribute to a systematic treatment of uncertainty and better modeler-user communication.

5.2 Future Work

As a future work, we are interested in identifying critical or otherwise interesting regions in the space of the input factors. Also, we would like to search to uncover factors which interact, and may therefore generate extreme values.

Also, calling a model validated does not make it valid. Researchers must continue to work toward finding ways to improve agent-based simulations. In this thesis we explored the parameter input space. In our future work, we look forward to investigate the output space and see if the model that we consider 'valid' in this manuscript would also pass the "fasifiability" test. On the same note, another interesting future work for MASE-BDI would be to test contrasting theories to see if there is fail in the expected basic patterns of the model.

We marginally assessed the spatial and temporal complexity of the model under a qualitative aspect, based on Agarwal et. al (2001) framework described in Ralha et al. Ralha et al. (2013). The Cerrado LUCC model objective is to assess the landscape dynamic having the mainly to human behavior drivers. It was out of the thesis' scope to characterize the dynamics of patchy spatiotemporal mosaics, but it is something that we would like to look into. Moreover, the study of changing landscape patterns involve calculating the indices for images of a landscape taken at several different times in history and then observing how these indices vary over time. As of today, the information of the landscape at each step of the simulation is not persisted in the MASE-BDI framework. We will consider re-factoring the code to gather this information for further analysis of the spatial and temporal complexity.

Until now, we have been working with parameter and methodological uncertainty. The next big breakthrough would be to extend our work to investigate the uncertainty within model structure. Another step would be an attempt to balance empirical validity, the base of this manuscript, with face validity, an approach that checks if processes and outcomes are reasonable and plausible within the frame of theoretic basis and implicit knowledge of system experts or stakeholders.

References

Abreu, C. G., & Ralha, C. G. (2017). Uncertainty Assessment in Agent-Based Simulation : An Exploratory Study. In G. Sukthankar & J. A. Rodriguez-Aguilar (Eds.), *Autonomous agents and multiagent systems* (pp. 36–50). São Paulo, Brazil: Springer International Publishing. doi: 10.1007/978-3-319-71682-4_3
62

Abreu, C. G., & Ralha, C. G. (2018). An empirical workflow to integrate uncertainty and sensitivity analysis to evaluate agent-based simulation outputs. *Environmental Modelling & Software*, *107*, 281–297. doi: https://doi.org/10.1016/j.envsoft.2018.06.013
58, 59, 60, 61, 62, 66, 71, 73

Albrecht, S. V., & Ramamoorthy, S. (2015). Are You Doing What I Think You Are Doing? Criticising Uncertain Agent Models. In *Proceedings of the 31st conference on uncertainty in artificial intelligence* (p. 10). Amsterdam, Netherlands.
13

Angus, S. D., & Hassani-Mahmooei, B. (2015). "Anarchy" Reigns: A Quantitative Analysis of Agent-Based Modelling Publication Practices in JASSS, 2001-2012. *Journal of Artificial Societies and Social Simulation*, *18*(4), 2001–2012. Retrieved from `http://jasss.soc.surrey.ac.uk/18/4/16.html` doi: 10.18564/jasss.2952
viii, xii, 4, 25, 86

Augusiak, J., Van den Brink, P. J., & Grimm, V. (2014, May). Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach. *Ecological Modelling*, *280*, 117–128. Retrieved from `http://dx.doi.org/10.1016/j.ecolmodel.2013.11.009` doi: 10.1016/j.ecolmodel.2013.11.009
x, xiii, xviii, 59, 63, 65, 66

Axelrod, R. (2003). Advancing the art of simulation in the social sciences - SSP. *Journal of the Japanese and International Economies*, *12*(3), 16–22. doi: 10.1002/(SICI)1099-0526(199711/12)3:2<16::AID-CPLX4>3.0.CO;2-K
2

Balci, O. (1998). Verification, validation, and accreditation. In *1998 winter simulation conference. proceedings (cat. no.98ch36274)* (Vol. 1, pp. 41–48). IEEE. Retrieved from `http://ieeexplore.ieee.org/document/744897/` doi: 10.1109/WSC.1998.744897
85

Bommel, P. (2017). Foreword. In D. F. Adamatti (Ed.), *Multi-agent based simulations applied to biological and environmental systems* (p. xv-xviii). Hershey, PA: IGI Global. Retrieved from `https://www.igi-global.com/pdf.aspx?tid=`

173208&ptid=164946&ctid=15&t=Foreword doi: 10.4018/978-1-5225-1756-6
2, 10

Brasil. (2015). *Mapeamento do Uso e Cobertura do Cerrado: Projeto TerraClass Cerrado 2013* (Tech. Rep.). Brasília, DF: Ministério do Meio Ambiente. Retrieved from http://www.dpi.inpe.br/tccerrado/index.php?mais=1
78

Bratley, P., Fox, B. L., Schrage, L. E., & Schouten, v. d. D. (1984). A Guide to Simulation. *Journal of the Operational Research Society*, *35*(4), 365–366. Retrieved from https://www.tandfonline.com/doi/abs/10.1057/jors.1984.73 doi: 10.1057/jors.1984.73
2

Bratman, M. E. (1987). *Intention, Plans, and Practical Reason.* Cambridge, MA: Harvard University Press. doi: 10.2307/2185304
34, 67

Braubach, L., Pokahr, A., & Lamersdorf, W. (2005). Jadex: A BDI-Agent System Combining Middleware and Reasoning. In R. Unland, M. Calisti, M. Klusch, M. Walliser, S. Brantschen, & T. Hempfling (Eds.), *Software agent-based applications, platforms and development kits* (pp. 143–168). Basel: Birkhäuser-Verlag. Retrieved from http://link.springer.com/10.1007/3-7643-7348-2_7 doi: 10.1007/3-7643-7348-2_7
34

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees.* Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.
31, 115

Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., & Mylopoulos, J. (2004). Tropos: An Agent-Oriented Software Development Methodology. *Autonomous Agents and MultiAgent Systems*, *8*(3), 203–236. Retrieved from http://www.springerlink.com/openurl.asp?id=doi:10.1023/B:AGNT.0000018806.20944.ef doi: 10.1023/B:AGNT.0000018806.20944.ef
2

Broeke, t. G., van Voorn, G., & Ligtenberg, A. (2016). Which Sensitivity Analysis Method Should I Use for My Agent-Based Model? *Journal of Artificial Societies and Social Simulation*, *19*(1), 1–35. doi: 10.18564/jasss.2857
61

Campolongo, F., & Braddock, R. (1999, Apr). The use of graph theory in the sensitivity analysis of the model output: a second order screening method. *Reliability Engineering & System Safety*, *64*(1), 1–12. Retrieved from http://dx.doi.org/10.1016/S0951-8320(98)00008-8http://linkinghub.elsevier.com/retrieve/pii/S0951832098000088 doi: 10.1016/S0951-8320(98)00008-8
19

Casti, J. L. (1995). *Complexification: Explaining a Paradoxical World through the Science of Surprise* (reprint ed.). HarperCollins.
10

Chipman, H. A., George, E. I., & McCulloch, R. E. (2012, Mar). BART: Bayesian

additive regression trees. *Annals of Applied Statistics*, *4*(1), 266–298. Retrieved from http://projecteuclid.org/euclid.aoas/1273584455 doi: 10.1214/09-AOAS285
31, 115

Cobelli, C., & DiStefano, J. J. (1980, Jul). Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *239*(1), R7–R24. doi: 10.1152/ajpregu.1980.239.1.R7
81

Coelho, C. G., Abreu, C. G., Ramos, R. M., D. Mendes, A. H., Teodoro, G., & Ralha, C. G. (2016, Jun). MASE-BDI: agent-based simulator for environmental land change with efficient and parallel auto-tuning. *Applied Intelligence*, *45*(3), 904–922. Retrieved from http://link.springer.com/10.1007/s10489-016-0797-8 doi: 10.1007/s10489-016-0797-8
vii, 10, 25, 35, 67, 72

Crosetto, M., Tarantola, S., & Saltelli, A. (2000). Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, Ecosystems & Environment*, *81*(1), 71–79. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167880900001699 doi: 10.1016/S0167-8809(00)00169-9
62, 113

Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973, Oct). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, *59*(8), 3873–3878. Retrieved from https://aip.scitation.org/doi/10.1063/1.1680571 doi: 10.1063/1.1680571
31, 110, 114

Dayananda, D., Irons, R., Harrison, S., Herbohn, J., & Rowland, P. (2002). *Capital Budgeting: Financial Appraisal of Investment Projects*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511753701
25

DeJonge, K. C., Ascough, J. C., Ahmadi, M., Andales, A. A., & Arabi, M. (2012). Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. *Ecological Modelling*, *231*, 113–125. Retrieved from http://dx.doi.org/10.1016/j.ecolmodel.2012.01.024 doi: 10.1016/j.ecolmodel.2012.01.024
62

Dijkstra, E. W. (1970, Apr). *Notes on Structured Programming* (Tech. Rep.). Technological University Eindhoven, The Netherlands.
85

Dosi, G., Pereira, M. C., & Virgillito, M. E. (2017, Mar). On the robustness of the fat-tailed distribution of firm growth rates: a global sensitivity analysis. *Journal of Economic Interaction and Coordination*, *13*(1), 173–193. Retrieved from https://link.springer.com/article/10.1007/s11403-017-0193-4 doi: 10.1007/s11403-017-0193-4
30

Duan, Q., Andréassian, V., Franks, S., Goteti, G., Gupta, H., Gusev, Y., ... Wood,

E. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, *320*(1), 3–17. Retrieved from `https://doi.org/10.1016/j.jhydrol.2005.07.031` doi: 10.1016/j.jhydrol.2005.07.031
3

Edmonds, B., & Moss, S. (2005). From KISS to KIDS - an 'anti-simplistic' modelling approach. In P. Davidsson, B. Logan, & K. Takadama (Eds.), *Multi-agent and multi-agent-based simulation* (Vol. 3415, pp. 130–144). Springer Berlin Heidelberg. doi: 10.1007/b106991
2, 59

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, *11*(4), 12. doi: 10.1080/01969720490426803
3, 63

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies : social science from the bottom up*. Brookings Institution Press.
2

Field, A. P., & Hole, a., Graham. (2003). *How to design and report experiments* [Book; Book/Illustrated]. London Thousand Oaks, Calif. Sage publications Ltd.
61

Fonoberova, M., Fonoberov, V. A., & Mezić, I. (2013, Oct). Global sensitivity/uncertainty analysis for agent-based models. *Reliability Engineering & System Safety*, *118*, 8–17. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S0951832013000999` doi: 10.1016/j.ress.2013.04.004
5, 30, 31, 46, 48, 53, 58, 62, 113

Forster, M. (2000). Key Concepts in Model Selection: Performance and Generalizability. *Journal of mathematical psychology*, *44*(1), 205–231. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/10733865` doi: 10.1006/jmps.1999.1284
81

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, *19*(1), 1–67. Retrieved from `http://projecteuclid.org/euclid.aos/1176347963` doi: 10.1214/aos/1176347963
31, 115

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., … Di, Z. (2014). A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling & Software*, *51*, 269–285. Retrieved from `http://doi.org/10.1016/j.envsoft.2013.09.031` doi: 10.1016/j.envsoft.2013.09.031
3, 13, 14, 15, 19, 25, 27, 31, 46, 49, 52, 53

Gao, L., & Hailu, A. (2012, May). Ranking management strategies with complex outcomes: An AHP-fuzzy evaluation of recreational fishing using an integrated agent-based model of a coral reef ecosystem. *Environmental Modelling & Software*, *31*, 3–18. doi: 10.1016/j.envsoft.2011.12.002
14

GDF. (2009). *Plano Diretor de Ordenamento Territorial do Distrito Federal: documento técnico* (Tech. Rep.). Brasília: Governo do Distrito Federal, Secretaria de Habitação, Regularização e Desenvolvimento Urbano - SEDHAB. Retrieved from `http://`

www.segeth.df.gov.br/plano-diretor-de-ordenamento-territorial/
36

Gibbs, M., & MacKay, D. (1997). *Efficient implementation of Gaussian processes* (Tech. Rep.). Cambridge University, UK: Cavendish Laboratory.
31, 116

Gilbert, G. (1996). Simulation As a Research Strategy. In *Social science microsimulation [dagstuhl seminar, may, 1995]* (pp. 448–454). Berlin, Heidelberg: Springer-Verlag. Retrieved from http://dl.acm.org/citation.cfm?id=647378.724486
2

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the social scientist* (2nd ed.). Maidenhead: Open University Press, McGraw-Hill Education.
57

Goldsman, D., & Tokol, G. (2000). Output Analysis Procedures for Computer Simulations. In J. Joines, R. R. Barton, K. Kang, & P. Fishwick (Eds.), *Proceedings of the 2000 winter simulation conference* (pp. 39–45). doi: 10.1145/510378.510388
20, 22

Gong, W., Duan, Q. Y., Li, J. D., Wang, C., Di, Z. H., Ye, A. Z., ... Dai, Y. J. (2015, Dec). An Intercomparison of Sampling Methods for Uncertainty Quantification of Environmental Dynamic Models. *Journal of Environmental Informatics*, *28*(11), 11–24. doi: 10.3808/jei.201500310
29, 109

Grimm, V. (1999). Ten years of individual-based modelling in ecology: What have we learned and what could we learn in the future? *Ecological Modelling*, *115*(2-3), 129–148. Retrieved from https://doi.org/10.1016/S0304-3800(98)00188-4 doi: 10.1016/S0304-3800(98)00188-4
34

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S. A., ... Railsback, S. F. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. *Ecological Modelling*, *280*, 129–139. Retrieved from http://dx.doi.org/10.1016/j.ecolmodel.2014.01.018 doi: 10.1016/j.ecolmodel.2014.01.018
57

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., ... Huse, G. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, *198*(1-2), 115–126. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0304380006002043 doi: 10.1016/j.ecolmodel.2006.04.023
16, 35, 68

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling*, *221*(23), 2760–2768. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S030438001000414X doi: 10.1016/j.ecolmodel.2010.08.019
16, 17, 18, 72

Grimm, V., & Railsback, S. F. (2012). Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1586), 298–310. Retrieved from http://rstb

.royalsocietypublishing.org/content/367/1586/298 doi: 10.1098/rstb.2011.0180
2

Groeneveld, J., Müller, B., Buchmann, C. M., Dressler, G., Guo, C., Hase, N., ... Schwarz, N. (2017). Theoretical foundations of human decision-making in agent-based land use models – A review. *Environmental Modelling and Software*, *87*, 39–48. Retrieved from http://dx.doi.org/10.1016/j.envsoft.2016.10.008 doi: 10.1016/j.envsoft.2016.10.008
3

Grow, A., & Van Bavel, J. (2017). *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications* (1st ed., Vol. 41). Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-32283-4
71

Hamilton, S. H., ElSawah, S., Guillaume, J. H. A., Jakeman, A. J., & Pierce, S. A. (2015). Integrated assessment and modelling: Overview and synthesis ofsalient dimensions. *Environmental Modelling & Software*, *64*, 215–229. Retrieved from http://dx.doi.org/10.1016/j.envsoft.2014.12.005 doi: 10.1016/j.envsoft.2014.12.005
58

Heath, B., Hill, R., & Ciarallo, F. (2009). A survey of agent-based modeling practices (January 1998 to July 2008). *Jasss*, *12*(4), 1–49. Retrieved from http://jasss.soc.surrey.ac.uk/12/4/9.html
10, 25, 86

Herd, B., Miles, S., McBurney, P., & Luck, M. (2015). A Monte Carlo Model Checker for Multiagent-Based Simulations. In *Multi-agent based simulation xvi* (Vol. 9568, pp. 37–54). Istanbul, Turkey. doi: 10.1007/978-3-319-31447-1
14

Houghton, J., Filho, L. M., Callander, B., Harris, N., Kattenberg, A., & Maskell, K. (Eds.). (1996). *Climate change 1995 the science of climate change.* IPCC, The Intergovernmental Panel on Climate Change.
11

IBRAM, B. (2014). *Guia de Unidades de Conservação do Distrito Federal* (Tech. Rep.). Instituto do Meio Ambiente e dos Recursos Hídricos do Distrito Federal - IBRAM.
71

Iman, R. L., & Helton, J. C. (1988, Mar). An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Analysis*, *8*(1), 71–90. Retrieved from http://doi.wiley.com/10.1111/j.1539-6924.1988.tb01155.x doi: 10.1111/j.1539-6924.1988.tb01155.x
30

Iooss, B., & Lemaître, P. (2015). A review on global sensitivity analysis methods. In *Uncertainty management in simulation-optimization of complex systems: Algorithms and applications.* Springer. doi: 10.1007/978-1-4899-7547-8_5
19

IPCC, I. P. o. C. C. (2014). *Climate change 2013 – the physical science basis: Working group i contribution to the fifth assessment report of the intergovernmental panel on climate change.* Cambridge University Press. doi: 10.1017/CBO9781107415324
11

Jain, R. (2011). *The art of computer systems performance analysis - techniques for experimental design, measurement, simulation, and modeling.* Wiley.
5

Karypis, G., & Kumar, V. (1998). *METIS* A Software Package for Partitioning Unstructured Graphs , Partitioning Meshes , and Computing Fill-Reducing Orderings of Sparse Matrices* (Tech. Rep.). Minneapolis, MN: University of Minnesota, Department of Computer Science / Army HPC Research Center.
110

Kelly (Letcher), R. A., Jakeman, A. J., Barreteau, O., Borsuk, M. E., ElSawah, S., Hamilton, S. H., . . . Voinov, A. A. (2013). Selecting among five common modelling approaches for integrated environmental assessment and management. *Environmental Modelling & Software*, *47*, 159–181. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S1364815213001151` doi: 10.1016/j.envsoft.2013.05.005
12, 25

Kleijnen, J. P., Sanchez, S. M., Lucas, T. W., & Cioppa, T. M. (2005). A user's guide to the brave new world of designing simulation experiments. *INFORMS Journal on Computing*, *17*(3), 263–289. Retrieved from `https://harvest.nps.edu/papers/UserGuideSimExpts.pdf` doi: 10.1287/ijoc.1050.0136
15, 29, 109

Klügl, F. (2008). A validation methodology for agent-based simulations. In *Proceedings of the 2008 acm symposium on applied computing* (pp. 39–43). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/1363686.1363696` doi: 10.1145/1363686.1363696
4

Law, A. M., & Kelton, D. (2000). *Simulation Modeling and Analysis* (3rd ed.) (No. 2). McGraw Hill.
28

Le, Q. B., Seidl, R., & Scholz, R. W. (2012, Jan). Feedback loops and types of adaptation in the modelling of land-use decisions in an agent-based simulation. *Environmental Modelling & Software*, *27-28*, 83–96. doi: 10.1016/j.envsoft.2011.09.002
14, 15

Lee, J. S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., . . . Parker, D. C. (2015). The complexities of agent-based modeling output analysis. *Jasss*, *18*(4), 1–25. Retrieved from `http://jasss.soc.surrey.ac.uk/18/4/4.html` doi: 10.18564/jasss.2897
10, 12, 13, 25, 28, 29, 49

Levy, S., & Steinberg, D. M. (2010, Dec). Computer experiments: A review. *AStA Advances in Statistical Analysis*, *94*(4), 311–324. Retrieved from `http://link.springer.com/10.1007/s10182-010-0147-9` doi: 10.1007/s10182-010-0147-9
20

Li, J. D., Duan, Q. Y., Gong, W., Ye, a. Z., Dai, Y. J., Miao, C. Y., . . . Sun, Y. W. (2013, Aug). Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. *Hydrology and Earth System Sciences Discussions*, *10*(2), 2243–2286. doi: 10.5194/hessd-10-2243-2013
14, 15

Li, Y., Brimicombe, A. J., & Chao, L. (2008, November). Agent-based services for

the validation and calibration of multi-agent models. *Computers, Environment and Urban Systems*, *32*(6), 464–473. Retrieved from http://roar.uel.ac.uk/327/
4

Ligmann-Zielinska, A. (2013). Spatially-explicit sensitivity analysis of an agent-based model of land use change. *International Journal of Geographical Information Science*, *27*(9), 1764–1781. Retrieved from http://dx.doi.org/10.1080/13658816.2013.782613{%}5Cnhttp://www.tandfonline.com/doi/abs/10.1080/13658816.2013.782613{%}5Cnhttp://www.tandfonline.com/doi/pdf/10.1080/13658816.2013.782613 doi: 10.1080/13658816.2013.782613
5

Ligmann-Zielinska, A., Kramer, D. B., Cheruvelil, K. S., & Soranno, P. A. (2014, Oct). Using uncertainty and sensitivity analyses in socioecological agent-based models to improve their analytical performance and policy relevance. *PLoS ONE*, *9*(10), e109779. Retrieved from http://dx.plos.org/10.1371/journal.pone.0109779 doi: 10.1371/journal.pone.0109779
24, 30, 31, 32, 48, 51, 58, 61, 62, 63, 64, 66, 82

Ligmann-Zielinska, A., & Sun, L. (2010a). Applying time-dependent variance-based global sensitivity analysis to represent the dynamics of an agent-based model of land use change. *International Journal of Geographical Information Science*, *24*(12), 1829–1850. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/13658816.2010.490533 doi: 10.1080/13658816.2010.490533
5, 62

Ligmann-Zielinska, A., & Sun, L. (2010b). Applying time-dependent variance-based global sensitivity analysis to represent the dynamics of an agent-based model of land use change. *International Journal of Geographical Information Science*, *24*(August 2016), 1829–1850. doi: 10.1080/13658816.2010.490533
31, 58

Lilburne, L., & Tarantola, S. (2009). Sensitivity analysis of spatial models. *International Journal of Geographical Information Science*, *23*(2), 151–168. doi: 10.1080/13658810802094995
24, 30, 52, 53, 62, 113

Lorscheid, I., Heine, B. O., & Meyer, M. (2012). Opening the 'Black Box' of Simulations: Increased Transparency and Effective Communication Through the Systematic Design of Experiments. *Computational and Mathematical Organization Theory*, *18*(1), 22–62. doi: 10.1007/s10588-011-9097-3
12, 15, 28, 29, 39, 43, 49, 58, 60, 61

Marks, R. E. (2007). Validating simulation models: A general framework and four applied examples. *Computational Economics*, *30*(3), 265–290. Retrieved from http://link.springer.com/10.1007/s10614-007-9101-7 doi: 10.1007/s10614-007-9101-7
12

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., ... Zwiers, F. W. (2010). Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. *Intergovernmental Panel on Climate Change (IPCC)*, 1–7.
11

Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, J. G., & Gotts, N. M. (2007). Agent-based land-use models : a review of applications. *Landscape Ecology*, *22*(10), 1447–1459. Retrieved from `https://link.springer.com/article/10.1007/s10980-007-9135-1` doi: 10.1007/s10980-007-9135-1
1, 3

McAlpine, C., Etter, A., Fearnside, P., Seabrook, L., & Laurance, W. (2009, Feb). Increasing world consumption of beef as a driver of regional and global change: A call for policy action based on evidence from Queensland (Australia), Colombia and Brazil. *Global Environmental Change*, *19*(1), 21–33. doi: 10.1016/j.gloenvcha.2008.10.008
51

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979, May). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, *21*(2), 239. doi: 10.2307/1268522
109

McKay, M. D., Morrison, J. D., & Upton, S. C. (1999a, Mar). Evaluating prediction uncertainty in simulation models. *Computer Physics Communications*, *117*(1-2), 44–51. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S0010465598001556` doi: 10.1016/S0010-4655(98)00155-6
11

McKay, M. D., Morrison, J. D., & Upton, S. C. (1999b, mar). Evaluating prediction uncertainty in simulation models. *Computer Physics Communications*, *117*(1-2), 44–51. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S0010465598001556` doi: 10.1016/S0010-4655(98)00155-6
31, 114

Metropolis, N., & Ulam, S. (1949, Sep). The Monte Carlo Method. *Journal of the American Statistical Association*, *44*(247), 335. doi: 10.2307/2280232
109

Morris, M. D. (1991, May). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, *33*(2), 161–174. Retrieved from `http://www.tandfonline.com/doi/abs/10.1080/00401706.1991.10484804` doi: 10.2307/1269043
19, 31, 111, 114

Murray-Rust, D., Rieser, V., Robinson, D. T., Miličič, V., & Rounsevell, M. (2013). Agent-based modelling of land use dynamics and residential quality of life for future scenarios. *Environmental Modelling & Software*, *46*, 75–89. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S1364815213000510` doi: 10.1016/j.envsoft.2013.02.011
24

National Research Council. (2012). *Assessing the Reliability of Complex Models: Mathematical and statistical foundations of verification, validation, and uncertainty quantification.* Washington, D.C.: National Academies Press. doi: 10.17226/13395
25, 31

Niederreiter, H., Hellekalek, P., Larcher, G., & Zinterhof, P. (Eds.). (1998). *Monte carlo and quasi-monte carlo methods 1996: Proceedings of a conference at the university of salzburg, austria, july 9–12, 1996* (1st ed., Vol. 127). Springer-Verlag New York.

Retrieved from https://www.springer.com/gp/book/9780387983356 doi: 07/978-1-4612-1690-2
51

Norton, J. (2015, Jul). An introduction to sensitivity assessment of simulation models. *Environmental Modelling & Software*, *69*, 166–174. doi: 10.1016/j.envsoft.2015.03.020
25

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Yarkoni, T. (2015, Jun). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. doi: 10.1126/science.aab2374
68

Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. V., & Alvin, K. F. (2002). Error and uncertainty in modeling and simulation. *Reliability Engineering & System Safety*, *75*(3), 333–357. doi: 10.1016/S0951-8320(01)00120-X
57

Olsen, M., & Raunak, M. (2016). Metamorphic validation for agent-based simulation models. In *Proceedings of the summer computer simulation conference* (pp. 33:1–33:8). San Diego, CA, USA: Society for Computer Simulation International. Retrieved from http://dl.acm.org/citation.cfm?id=3015574.3015607
4

O'Neill, Z., & Niu, F. (2017). Uncertainty and sensitivity analysis of spatio-temporal occupant behaviors on residential building energy usage utilizing Karhunen-Lo?ve expansion. *Building and Environment*, *115*, 157–172. doi: 10.1016/j.buildenv.2017.01.025
60, 62

Owen, a. B. (1992). Orthogonal Arrays for Computer Experiments, Integration and Visualization. *Statistica Sinica*, *2*(2), 439–452. Retrieved from https://www.jstor.org/stable/24304869
110

Padgham, L., & Winikoff, M. (2003). Prometheus: A methodology for developing intelligent agents. *Agent-Oriented Software Engineering III*, 174–185. doi: 10.1007/3-540-36540-0_14
2

Paegelow, M., Camacho Olmedo, M. T., Mas, J.-F., & Houet, T. (2014). Benchmarking of LUCC modelling tools by various validation techniques and error analysis. *Cybergeo*, *701*(online), 29. doi: 10.4000/cybergeo.26610
13, 15

Pappenberger, F., Beven, K. J., Ratto, M., & Matgen, P. (2008, Jan). Multi-method global sensitivity analysis of flood inundation models. *Advances in Water Resources*, *31*(1), 1–14. doi: 10.1016/j.advwatres.2007.04.009
25

Parker, D. C., Berger, T., & Manson, S. M. (2001). *Agent-Based Models of Land-Use and Land- Cover Change* (Tech. Rep.). Irvine, California, USA: Report and Review of an International Workshop. L. R. No.6.
1

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003,

Jun). Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review. *Annals of the Association of American Geographers*, *93*(2), 314–337. Retrieved from http://dx.doi.org/10.1111/1467-8306.9302004 doi: 10.1111/1467-8306.9302004
1, 24

Parry, H. R., Topping, C. J., Kennedy, M. C., Boatman, N. D., & Murray, A. W. (2013, Jul). A Bayesian sensitivity analysis applied to an Agent-based model of bird population response to landscape change. *Environmental Modelling & Software*, *45*, 104–115. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S1364815212002265 doi: 10.1016/j.envsoft.2012.08.006
5, 31, 58

Pi, H., & Peterson, C. (1994). Finding the Embedding Dimension and Variable Dependencies in Time Series. *Neural Computation*, *6*(3), 509–520. Retrieved from https://doi.org/10.1162/neco.1994.6.3.509 doi: 10.1162/neco.1994.6.3.509
31, 116

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016, May). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. Retrieved from http://dx.doi.org/10.1016/j.envsoft.2016.02.008 doi: 10.1016/j.envsoft.2016.02.008
ix, xii, 5, 25, 27, 32, 33, 34, 48, 49

Pontius, R. G. (2000). Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing*, *66*(8), 1011–1016. Retrieved from https://www.clarku.edu/{~}rpontius/pontius{_}2000{_}pers.pdf
1

Pontius, R. G., Boersma, W., Castella, J.-C., Clarke, K., de Nijs, T., Dietzel, C., ... Verburg, P. H. (2008, Mar). Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, *42*(1), 11–37. Retrieved from http://link.springer.com/10.1007/s00168-007-0138-2 doi: 10.1007/s00168-007-0138-2
viii, 10, 18, 37, 38, 60, 74, 79

Pontius, R. G., Huffaker, D., & Denman, K. (2004). Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling*, *179*(4), 445–461. doi: 10.1016/j.ecolmodel.2004.05.010
37

Pontius, R. G., & Millones, M. (2011, Aug). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, *32*(15), 4407–4429. doi: 10.1080/01431161.2011.552923
60

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

Railsback, S. F., & Volker, G. (2011). *Agent-Based and Individual-Based Modeling:*

A Practical Introduction. Princeton University Press. Retrieved from http://gen.lib.rus.ec/book/index.php?md5=950D915C046EF0ACBF81687F9D167308
1

Ralha, C. G., & Abreu, C. G. (2017). A multi-agent-based environmental simulator. In D. F. Adamatti (Ed.), *Multi-agent-based simulations applied to biological and environmental systems* (pp. 106–127). Hershey, PA: IGI Global. Retrieved from http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-5225-1756-6.ch005 doi: 10.4018/978-1-5225-1756-6.ch005
9, 16, 34

Ralha, C. G., Abreu, C. G., Coelho, C. G. C., Zaghetto, A., Macchiavello, B., & Machado, R. B. (2013, Apr). A multi-agent model system for land-use change simulation. *Environmental Modelling & Software*, *42*, 30–46. Retrieved from http://dx.doi.org/10.1016/j.envsoft.2012.12.003 doi: 10.1016/j.envsoft.2012.12.003
10, 15, 16, 18, 24, 35, 38, 41, 67, 68, 72, 88

Refsgaard, J. C., & Henriksen, H. J. (2004). Modelling guidelines - Terminology and guiding principles. *Advances in Water Resources*, *27*(1), 71–82. Retrieved from https://doi.org/10.1016/j.advwatres.2003.08.006 doi: 10.1016/j.advwatres.2003.08.006
xviii, 65

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process - A framework and guidance. *Environmental Modelling & Software*, *22*(11), 1543–1556. doi: 10.1016/j.envsoft.2007.02.004
60, 85

Richiardi, M., Leombruni, R., Saam, N., & Sonnessa, M. (2006, jan). A Common Protocol for Agent-Based Social Simulation. Retrieved from http://jasss.soc.surrey.ac.uk/9/1/15.html
4

Rindfuss, R., Entwisle, B., Walsh, S., Li, A., Badenoch, N., Brown, D., ... Verburg, P. (2008). Land use change: Complexity and comparisons. *Journal of Land Use Science*, *3*(1), 1–10. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/17474230802047955 doi: 10.1080/17474230802047955
1

Rindfuss, R., Entwisle, B., Walsh, S. J., An, L., Badenoch, N., Brown, D. G., ... Verburg, P. H. (2008, Jul). Land use change: complexity and comparisons. *Journal of Land Use Science*, *3*(1), 1–10. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/17474230802047955 doi: 10.1080/17474230802047955
24

Saltelli, A. (2002, May). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, *145*(2), 280–297. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0010465502002801 doi: 10.1016/S0010-4655(02)00280-1
61, 111

Saltelli, A., & Annoni, P. (2010, Dec). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, *25*(12), 1508–1517. Retrieved from http://dx

.doi.org/10.1016/j.envsoft.2010.04.012 doi: 10.1016/j.envsoft.2010.04.012
5, 114

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer.* Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/9780470725184
12, 29, 30, 51, 53, 58, 60, 61, 109, 112, 113, 114

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A Guide to Assessing Scientific Models.* Ispra, Italy: John Wiley & Sons, Ltd.
31, 80

Schlesinger, S. (1979, mar). Terminology for model credibility. *SIMULATION*, *32*(3), 103–104. Retrieved from http://journals.sagepub.com/doi/10.1177/003754977903200304 doi: 10.1177/003754977903200304
xviii, 65

Schreinemachers, P., & Berger, T. (2011, Jul). An agent-based simulation model of human–environment interactions in agricultural systems. *Environmental Modelling & Software*, *26*(7), 845–859. Retrieved from http://dx.doi.org/10.1016/j.envsoft.2011.02.004 doi: 10.1016/j.envsoft.2011.02.004
4, 14, 24

Schulze, J., Müller, B., Groeneveld, J., & Grimm, V. (2017). Agent-Based Modelling of Social-Ecological Systems: Achievements, Challenges, and a Way Forward. *Journal of Artificial Societies and Social Simulation*, *20*(2). Retrieved from http://jasss.soc.surrey.ac.uk/20/2/8.html doi: 10.18564/jasss.3423
1, 3

Shin, M. J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2013). Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R. *Journal of Hydrology*, *503*, 135–152. Retrieved from http://dx.doi.org/10.1016/j.jhydrol.2013.08.047 doi: 10.1016/j.jhydrol.2013.08.047
5

Singhee, A., & Rutenbar, R. A. (2010, Nov). Why Quasi-Monte Carlo is Better Than Monte Carlo or Latin Hypercube Sampling for Statistical Circuit Analysis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *29*(11), 1763–1776. Retrieved from https://ieeexplore.ieee.org/document/5605333/ doi: 10.1109/TCAD.2010.2062750
51

Smajgl, A., Brown, D. G., Valbuena, D., & Huigen, M. G. A. (2011). Environmental Modelling & Software Empirical characterisation of agent behaviours in socio-ecological systems. *Environmental Modelling and Software*, *26*(7), 837–844. Retrieved from http://dx.doi.org/10.1016/j.envsoft.2011.02.011 doi: 10.1016/j.envsoft.2011.02.011
16, 18, 35, 72

Smith, J., Winograd, M., Gallopín, G. C., & Pachico, D. H. (1998). Dynamics of the agricultural frontier in the Amazon and savannas of Brazil: analyzing the impact of policy and technology. *Environmental Modeling and Assessment*, *3*(1), 31–46. Retrieved from http://link.springer.com/10.1023/A:1019094218552 doi: 10.1023/A:1019094218552

51, 72

Sobol', I. (1993). Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical Modelling and Computational Experiments*, *1*(4), 407—-414.
31, 62, 114

Sobol, I. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, *55*(1), 271 - 280. (The Second IMACS Seminar on Monte Carlo Methods) doi: https://doi.org/10.1016/S0378-4754(00)00270-6
111

Spearman, C., Con, C. S., & Page, N. (1904, Jan). The Proof and Measurement of Association between Two Things. *Source: The American Journal of Psychology*, *15*(1), 72–101. Retrieved from `http://www.jstor.org/stable/1412159` doi: 10.2307/1412159
31, 115

Statnikov, R. B., & Matusov, J. B. (2002). *Multicriteria Analysis in Engineering.* Dordrecht: Springer Netherlands. doi: 10.1007/978-94-015-9968-9
110

Sterman, J. D. (1991). A Skeptic's Guide to Computer Models. *Managing a Nation: The Microcomputer Software Catalog*, 209–229. Retrieved from `http://web.mit.edu/jsterman/www/Skeptic's{_}Guide.pdf`
2

Tang, B. (1993, Dec). Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association*, *88*(424), 1392–1397. doi: 10.1080/01621459.1993.10476423
110

Tang, Y., Reed, P., Wagener, T., & van Werkhoven, K. (2007, Feb). Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences*, *11*(2), 793–817. doi: 10.5194/hess-11-793-2007
25

ten Broeke, G., van Voorn, G., & Ligtenberg, A. (2016). Which Sensitivity Analysis Method Should I Use for My Agent-Based Model? *Journal of Artificial Societies and Social Simulation*, *19*(1), 1–35. Retrieved from `http://jasss.soc.surrey.ac.uk/19/1/5.html` doi: 10.18564/jasss.2857
5, 30, 52, 53

Thiele, J. C., Kurth, W., & Grimm, V. (2014a). Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models: A Cookbook Using NetLogo and 'R'. *Journal of Artificial Societies and Social Simulation*, *17*(3), 1–45. Retrieved from `http://jasss.soc.surrey.ac.uk/17/3/11.html` doi: 10.18564/jasss.2503
29, 37, 53, 109

Thiele, J. C., Kurth, W., & Grimm, V. (2014b). Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models: A Cookbook Using NetLogo and 'R'. *Journal of Artificial Societies and Social Simulation*, *17*(3), 1–45. doi: 10.18564/jasss.2503
61

Tichy, W. F. (1998). Should computer scientists experiment more? *Computer*, *31*(5),

32–40. Retrieved from https://ieeexplore.ieee.org/document/675631/ doi: 10.1109/2.675631
85, 86

Tong, C. (2005). *PSUADE User's Manual* (Tech. Rep.). Livermore, CA: Lawrence Livermore National Laboratory (LLNL).
ix, xiii, 31, 40, 74

Tong, C. (2015). *PSUADE Reference Manual (Version 1.7)* (Tech. Rep.). Livermore, CA: Lawrence Livermore National Laboratory (LLNL).
15, 19

Trucano, T., Swiler, L., Igusa, T., Oberkampf, W., & Pilch, M. (2006, Oct). Calibration, validation, and sensitivity analysis: What's what. *Reliability Engineering & System Safety*, *91*(10-11), 1331–1357. Retrieved from https://doi.org/10.1016/j.ress.2005.11.031 doi: 10.1016/j.ress.2005.11.031
73

Uusitalo, L., Lehikoinen, A., Helle, I., & Myrberg, K. (2015, Jan). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling & Software*, *63*, 24–31. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364815214002813 doi: 10.1016/j.envsoft.2014.09.017
61, 62

Valbuena, D., Verburg, P. H., Veldkamp, A., Bregt, A. K., & Ligtenberg, A. (2010). Effects of farmers' decisions on the landscape structure of a Dutch rural region: An agent-based approach. *Landscape and Urban Planning*, *97*(2), 98–110. Retrieved from http://dx.doi.org/10.1016/j.landurbplan.2010.05.001http://linkinghub.elsevier.com/retrieve/pii/S0169204610000915 doi: 10.1016/j.landurbplan.2010.05.001
4

Vanrolleghem, P. A., Mannina, G., Cosenza, A., & Neumann, M. B. (2015, Mar). Global sensitivity analysis for urban water quality modelling: Terminology, convergence and comparison of different methods. *Journal of Hydrology*, *522*, 339–352. doi: 10.1016/j.jhydrol.2014.12.056
25, 27

Verburg, P. H. (2006, Nov). Simulating feedbacks in land use and land cover change models. *Landscape Ecology*, *21*(8), 1171–1183. Retrieved from http://link.springer.com/10.1007/s10980-006-0029-4 doi: 10.1007/s10980-006-0029-4
1, 24

Walter, E. (1987). *Identifiability of Parametric Models* (Vol. 63). Cham: Elsevier Science Ltd.
4

Wang, J., Li, X., Lu, L., & Fang, F. (2013). Parameter sensitivity analysis of crop growth models based on the extended Fourier Amplitude Sensitivity Test method. *Environmental Modelling and Software*, *48*, 171–182. doi: 10.1016/j.envsoft.2013.06.007
25, 27

Wooldridge, M. (2009). *An introduction to multiagent systems* (2nd ed.). Wiley Publishing.

34
Wooldridge, M., Jennings, N. R., & Kinny, D. C. N. W.-.-. (2000). The Gaia Methodology for Agnet-Oriented Analysis and Design. *Journal of Autonomous Agents and Multi-Agent Systems*, *3*(3), 285–312. Retrieved from `https://doi.org/10.1023/A:1010071910869` doi: https://doi.org/10.1023/A:1010071910869
2
Yang, J. (2011, Apr). Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. *Environmental Modelling & Software*, *26*(4), 444–457. doi: 10.1016/j.envsoft.2010.10.007
25
Zimmermann, A., Heckelei, T., & Domínguez, I. P. (2009). Modelling farm structural change for integrated ex-ante assessment: review of methods and determinants. *Environmental Science and Policy*, *12*(5), 601–618. Retrieved from `https://doi.org/10.1016/j.envsci.2009.01.014` doi: 10.1016/j.envsci.2009.01.014
4
Zucchini, W. (2000, Mar). An Introduction to Model Selection. *Journal of Mathematical Psychology*, *44*(1), 41–61. Retrieved from `https://doi.org/10.1006/jmps.1999.1276` doi: 10.1006/jmps.1999.1276
81

Appendix A

List of Publications

Journals

- Abreu, C.G.; Ralha, C.G. (2018) An empirical workflow to integrate uncertainty and sensitivity analysis to evaluate agent-based simulation outputs. *Environmental Modelling & Software*, v. 107, p. 281-297, 2018. `https://doi.org/10.1016/j.envsoft.2018.06.013`

- Abreu, C.G.; Ralha, C.G. (2018) 'Evaludation' of agent-based simulation output to improve analytical confidence. *Simulation Modelling Practice and Theory*, p.1-34. Under review.

- Coelho, C.G.C.; Abreu, C.G.; Ramos, R.M.; Mendes, A.H.D.; Teodoro, G.L.M.; Ralha, C.G.(2016) MASE-BDI: agent-based simulator for environmental land change with efficient and parallel auto-tuning. *Applied Intelligence (Boston)*, 45:904. p. 904-922, 2016. `https://doi.org/10.1007/s10489-016-0797-8`

- Ralha, C.G.; Abreu, C.G.; Coelho, C.G.C; Zaghetto, A.; Macchiavello, B.; Machado, R.B. (2013) A multi-agent model system for land-use change simulation. *Environmental Modelling & Software*, v. 42, p. 30-46, 2013. `http://dx.doi.org/10.1016/j.envsoft.2012.12.003`

Book Chapters

- Abreu, C.G.; Ralha, C.G. (2018) Uncertainty Assessment in Agent-Based Simulation: An Exploratory Study. In: Dimuro G., Antunes L. (eds). *Multi-Agent Based Simulation XVIII*. MABS 2017. Lecture Notes in Computer Science, vol 10798, p. 74-88. Springer, Cham. `https://doi.org/10.1007/978-3-319-91587-6_6`

- Ralha, C.G. ; Abreu, C.G. (2017) Mase: A Multi-Agent-Based Environmental Simulator. In: D. Adamatti. (Ed.). *Multi-Agent-Based Simulations Applied to Biological and Environmental Systems* (pp. 106-127). Hershey, Pennsylvania,USA: IGI Global. `https://doi.org/10.4018/978-1-5225-1756-6.ch005`

- Abreu C.G.; Ralha C.G. (2017) Uncertainty Assessment in Agent-Based Simulation: An Exploratory Study. In: Sukthankar G., Rodriguez-Aguilar J. (eds) *Autonomous Agents and Multiagent Systems*. AAMAS 2017. Lecture Notes in Computer Science, vol 10642. Springer, Cham. `https://doi.org/10.1007/978-3-319-71682-4_3`

Conferences

- Alencar, F.L.; Monsalve-Herrera, D.; Abreu, C.G.; Coelho, C.G.C.; Alves, C.M.A.; Ralha, C.G. Hydric-Agent: Ferramenta de Simulação Baseada em Agentes para Gestão da Água em Áreas Residenciais. In: 12º Brazilian e-Science Workshop (BreSci), 2018, Natal. Anais do XXXVIII CSBC - XII BreSci. Porto Alegre: SBC, 2018. v. 1. p. 1-18.

- Abreu, C.G.; Ralha, C.G. Uncertainty Assessment in Agent-Based Simulation: An Exploratory Study. In: 18th Workshop on Multi-agent-based Simulation MABS / AAMAS, 2017, São Paulo. Proceedings of the MABS / AAMAS'17. NY: IFAAMAS, 2017. v. 1. p. 1-12.

- Abreu, C.G.; Coelho, C.G.C.; Ralha, C.G. MASE-BDI: Agents with Practical Reasoning for Land Useand Cover Change Simulation. In: WCAMA 2015. VI Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais, 2015, Recife. Anais do XXXV Congresso da Sociedade Brasileira de Computação. Porto Alegre: SBC, 2015. p. 1-10.

- Abreu, C.G.; Coelho, C.G.C. ; Ralha, C. G. ; Macchiavello, B.L. A Model and Simulation Framework for Exploring Potential Impacts of Land Use Policies: the Brazilian Cerrado Case. In: 47th Hawaii International Conference on System Science, 2014, Hawaii. Proceedings of the Forty-Seventh Annual Hawaii International Conference on System Sciences (HICSS), 2014. p. 847-856.

- Ralha, C.G. ; Abreu, C.G. ; Coelho, C.G.C. MASE: Multi-Agent System for Environmental Simulation. In: IV Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais, 2013, Maceió. Anais do IV WCAMA. Porto Alegre: SBC, 2013. v. 1. p. 1-5.

- Ralha, C. G. ; Abreu, C.G. ; Coelho, C.G.C. ; Macchiavello, B.L. . The MASE design experience. In: 20th International Congress on Modelling and Simulation (MODSIM2013), 2013, Adelaide, Australia. Proceedings of MODSIM'13. Australia: MSSANZ, 2013. v. 1. p. 1-7.

- Abreu, C.G.; Costa, F.; Santos, L.; Monteiro, L.; Oliveira, L.F.P.; Lustosa, P.; Weigang, L. Entity Extraction within Plain-Text Collections WISE 2013 Challenge - T1: Entity Linking Track. In: The 14th International Conference on Web Information System Engineering (WISE 2013), 2013, Nanjing. Proceedings of the Web Information System Engineering (WISE 2013). Berlin: Springer-Verlag Berlin Heidelb: LNCS, 2013. v. 8180. p. 491-496.

Awards

- Best Paper-First Place 18th Workshop on Multi-agent-based Simulation (MABS / AAMAS 2017) (2017), Institute of Smart Cities - project Ayudas a Talentos by Caixa and Fundación Caja Navarra (Spain).

- Best Paper Nomination, 47th Hawaii International Conference on System Sciences (HICSS) (2014), IEEE Computer Society.

- Championship on T1- Entity Linking Track: Precision in the WISE 2013 Challenge, WISE (2013), Springer.

Appendix B

Parameter space exploration

Sampling methods provide a systematic exploration of the parameter space that guarantees the sample to have specific statistical or structural properties. The purpose of these methods is to actively reduce the number of parameter sets that are considered but still chose space-filling points in the design space (Thiele et al., 2014a). For a complete revision of sampling methods, readers can refer to Gong et al. (2015); Kleijnen et al. (2005); Saltelli et al. (2008). In this manuscript, the most common sampling designs are illustrated and applied in the UQ process.

Monte Carlo

Monte Carlo sampling (MC) (Metropolis & Ulam, 1949) method is the most common class of computational techniques based on repeated random sampling to obtain N numerical approximations of a specified distribution function of an unknown probabilistic entity. However, larger sample sizes are required to explore the parameter space fully.

Latin Hypercube

Latin Hypercube (LH) (McKay, Beckman, & Conover, 1979) is a 1-dimensionally space-filling method, also known as stratified sampling method without replacement. When sampling a function of n variables, the range of each variable is divided into p equally probable intervals, with a total of p sample points. Therefore, each sample point is the only one in each interval. LH method selects sample points in the interior of the hypercube of p levels. LH can capture more variability in the sample space than simple random sampling.

Orthogonal Array

Orthogonal Array (OA) (Owen, 1992) is a 2-dimensionally space-filling method that uses a general fractional factorial design to improve LH. The OA design extends to t dimensional margins the univariate stratification properties of LH. That is, for a n-dimension, p-level parameter space, a t-strength OA sampling generates p^t sample points, when $t < n$.

Orthogonal Array-based Latin Hypercube

Orthogonal Array-based Latin Hypercube (OALH) (B. Tang, 1993) uses orthogonal arrays to construct Latin hypercubes. In other words, the samples go through a stratification process to produce samples that have been both orthogonalized and stratified. This sampling scheme provides more suitable designs for computer experiments and numerical integration than general LH sampling.

METIS

METIS sampling (Karypis & Kumar, 1998) is an m-directional space-filling method that is a part of a set of multilevel partitioning algorithms designed for partitioning irregular graphs, partitioning large meshes and computing fill-reducing ordering of sparse matrices. METIS can partition an unstructured graph into a user-specified number k of parts.

Fourier

Fourier sampling algorithm (Cukier et al., 1973) was designed specifically for the Fourier Amplitude Sensitivity Test (FAST). In this method, the parameter space is explored periodically with interference-free frequencies. It takes a small number of correlated random samples from a signal and processes them efficiently to produce an approximation of the discrete Fourier transform (DFT) of the signal. The minimum sample size of FAST is $N = 2 \cdot M_s \cdot \omega_{max} + 1$, where M_s is the maximum harmonic (in general 4 or 6) and ω_{max} is the maximum frequency which is determined by the number of inputs.

LPτ

LPτ (LPTAU) (Statnikov & Matusov, 2002) is a quasi-random (QR) sampling method, i.e., the samples are generated from a finite subset of low-discrepancy sequence of points. These samples are not random, in the sense of being completely unpredictable. However, they are like random points in the sense that they are uniformly distributed across a n-dimensional space. LPTAU explores the parameter space using partitions of the parameter ranges on the base of two.

Sobol Extended

Sobol Extended (SOBOL) (Saltelli, 2002; Sobol, 2001) is a replicated version of low-discrepancy sequences (quasi-random). SOBOL generates a uniform distribution in probability space, a qualitatively random distribution, filling previously unsampled regions of the probability function. This is done with two random $r \cdot n$ sample matrices M_0 and M_{n+1}, therefore, the total number of sample points is $(n+2) \cdot r$.

Morris one-at-a-time

Morris one-at-a-time (MOAT) (Morris, 1991) sampling was designed specifically for MOAT SA and is similar to SOBOL. The range of each parameter is divided into $p - 1$ equal intervals. Next, r points are generated from the n-dimension, $p - 1$-orthogonal grid. For each one, other sample points are generated by perturbing one dimension at a time, until all dimensions have been varied for only one time, with a $(n+1) \cdot r$ total number of sample points.

Appendix C

Uncertainty Analysis

The uncertainty analysis assesses a confidence bound on the output estimation by quantifying the uncertainty associated with the model response due to uncertainties in the model input. To achieve this results we follow the necessary steps of UA summarized by Saltelli et al. (2008):

1. Start from a model parameter α $N(\overline{\alpha}, \sigma_\alpha)$, which reads: after estimation, the distribution of α is known, with mean $\overline{\alpha}$ and standard deviation σ_α;

2. Assume that all the parameters (β, γ, ...) are independent of each other;

3. Draw a sample from the respective distributions of each parameter. In other words, produce a set of row vectors $(\alpha^{(j)}, \beta^{(j)}, ...)$ in a way that $(\alpha^1, \alpha^2, ..., \alpha^{(N)})$ is a sample from $N(\overline{\alpha}, \sigma_\alpha)$. Likewise for all parameters

$$\begin{bmatrix} \alpha^{(1)} & \beta^{(1)} & \gamma^{(1)} & ... \\ \alpha^{(2)} & \beta^{(2)} & \gamma^{(2)} & ... \\ ... & ... & ... & ... \\ \alpha^{(N-1)} & \beta^{(N-1)} & \gamma^{(N-1)} & ... \\ \alpha^{(N)} & \beta^{(N)} & \gamma^{(N)} & ... \end{bmatrix};$$

4. Run the model for all vectors $(\alpha^{(j)}, \beta^{(j)}, ...)$ thereby producing a set of N values of a model output Y_j

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ ... \\ y^{(N-1)} \\ y^{(N)} \end{bmatrix}.$$

By executing these steps, it is possible to quantify the impact of input uncertainties on the model response and assess whether or not the response meets the required standards of precision. Although Monte Carlo is the most used method, there are many other methods available to generate the samples and estimations required by UA. Some interesting UA applications and experimental design are described in the literature: (Fonoberova et al., 2013), (Saltelli et al., 2008),(Lilburne & Tarantola, 2009), (Crosetto et al., 2000). The expected means and variance are quantified to each parameter. Additionally, a histogram of the output variable can be displayed, thus thoroughly describing the stochastic features of the model output. The overall computational cost of UA depends basically on the cost of the model evaluations, which is linked to the complexity of the model itself.

Appendix D

Sensitivity Analysis Methods

Many techniques for SA have been proposed, and a thorough description of the techniques can be found in Saltelli et al. (2008). Regardless of the technique, Saltelli and Annoni (2010) present a guideline on how to avoid perfunctory SA, which we applied throughout the manuscript. A brief description of the methods applied is found next.

Morris one-at-a-time

The Morris one-of-a-time screening method (MOAT) (Morris, 1991) may be regarded as a gradient-based global SA as the final measure is obtained by averaging local measures, the elementary effects (EE). It is composed of individually randomized one-at-a-time experiments that calculate two sensitivity measures of the gradients of each parameter sampled from r local changes. The mean μ assesses the overall influence of the factor on the output. The standard deviation σ estimates the ensemble of the factor's effects, whether nonlinear or due to interactions with other factors. EE provides the information that the effects for a given parameter may be: i) negligible, ii) linear and additive, or iii) nonlinear or involved with interactions with other factors. MOAT can be much faster than other variance-based SA techniques.

Variance-based SA techniques

We assessed three variance-based SA techniques: SOBOL (Sobol', 1993), FAST (Cukier et al., 1973), and McKay (McKay et al., 1999b). In general, they have higher computational cost than qualitative SA, but some exciting features to ABMs are that variance-based SA measures are model independent, and provide the investigation of interaction effects. The first-order index represents the main effect contribution of each input factor to the variance output. The total effect of a variable would be the total contribution to the output variation, that is its first-order effect plus all higher-order effects due to interaction. In the SOBOL method, the variance may be attributed to a single input (first-order/main-effect) or by the interaction of two or more inputs (second-order-effect). The sum of those contributions is the total effect of a parameter. To decompose the variance, FAST varies different parameters at different frequencies and applies a Fourier transformation to measure each parameter contribution. McKay uses analysis of variance (ANOVA) to calculate a correlation ratio, that is a ratio of the variance of a parameter and the total variance of the output. The significance of the parameter increases with the correlation ratio.

Linear-regression-based SA techniques

Linear-regression-based SA decomposes the variance of the model outcomes by fitting a regression function of the input parameters to these outcomes. Therefore, the simulation outcomes are described concerning input-output relationships, which can be validated using standard statistical measures such as R^2. Correlation Analysis (CA) measures the parameter sensitivity through correlations coefficients, such as Spearman et al. (1904). Regression Analysis (RA) makes the same measures using the standard regression coefficient (SRC), to estimate the result from a regression analysis that has been normalized so that the variances of the dependent and independent variables are equal to one. The efficacy of this methods relies on the input-output being somewhat linear or monotonic.

Response-surface SA techniques

The methods Sum-of-Trees (SOT), Gaussian Process (GP), and Multivariate Adaptive Regression Splines (MARS), are considered response-surface or surrogate models, from which it is possible to obtain relative scores of the total effects of a parameter. Those methods provide a mapping from parameters to outputs. SOT (Breiman et al., 1984; Chipman et al., 2012) is a tree-based Bayesian method. A single regression tree model is obtained by the use of a recursive binary partition of the parameter space. The created balanced binary tree, in which the variables are split to cause the maximum decrease in the residual sum of squares, has each terminal node with a minimum number of sample points. The variable with the larger number of splits is considered the most sensitive one.

Non-parametric regression SA techniques

MARS (Friedman, 1991) is a non-parametric regression able to model nonlinearities and interactions between parameters. It is considered an extension of the tree method because after partitioning the space it builds localized regressions (first and second-order). For each model, a score (generalizes cross-validation) is computed. It will remove each parameter and recalculate the model score. The larger the score, the more important is the removed parameter.

Gaussian SA techniques

GP is an implementation of the Tpros algorithm, proposed by (Gibbs & MacKay, 1997). GP is a method for regression using Gaussian process priors which allow exact Bayesian analysis using matrix manipulations. The theory behind the method states that points that are close on parameter space give rise to similar response values. Thus, it is possible to identify the influence of the parameters on the model response.

Tailored SA techniques

The DT (Pi & Peterson, 1994) is a method that establishes dependencies in continuous functions given a sequence of measurements δ, an estimate of noise variance when a subset of variables in the sample are selected for regression. The approach is based on calculating conditional probabilities from vector component distances. It has been proved that adding unrelated variables or withdrawing related ones will increase δ. Hence, the subset of all variables that minimize noise variance is considered the most sensitive.

Appendix E

PSUADE: Configuration Files

Input file

Example of a *psuade.in* configuration file. The INPUT section defines the inputs, their ranges, and distributions. When the distribution is not informed, the uniform distribution is chosen as default. The OUTPUT section defines the order and name of the output variables. In the METHOD section, it is possible to observe that the MOAT sampling technique is set to 800 samples. The APPLICATION section specifies the direct call to the MASE-Driver, which will control the interface between the PSUADE and the MASE-BDI executions. The last section of the PSUADE file, the ANALYSIS section, does not present any method in this example. This allows the generated sample results to be saved to a *psuadeData* file that can be stored. Multiple analysis can be performed by command line or graphical interface.

```
PSUADE
INPUT
dimension 4
variable 1 transformationAgentQty = 1 100
variable 2 transformationAgentGroupPercentage = 10 100
variable 3 individualExploration = 1 500
variable 4 groupExploration = 1 1500
END
OUTPUT
dimension 11
variable 1 time
variable 2 qtyAgents
variable 3 percentageAgents
```

```
variable 4 figureOfMerit
variable 5 producersaccuracy
variable 6 usersaccuracy
variable 7 wrongchange
variable 8 rightchange
variable 9 wrongpersistance
variable 10 nullModel
variable 11 simulatedNullModel
END
METHOD
sampling = MOAT
num_samples = 800
num_replications = 1
num_refinements = 0
refinement_size = 10000000
reference_num_refinements = 0
refinement_type =
randomize
random_seed = 12504321
END
APPLICATION
driver = ./MASE-Driver.py
END
ANALYSIS
analyzer output_id = 1
printlevel 4
END
END
```

Call to MASE-Driver

The following code is an example of a call from PSUADE to the *MASE-driver*. Each of the 800 samples is sequentially passed to MASE-BDI framework to be simulated. Each line of the call is a parameter that corresponds to a simulation input. The order of the parameters are the ones specified in the *psuade.in* file.

```
4
  4.4077203315252991e+00
```

```
  2.3059819407323289e+01
  3.3939929566939020e+02
  5.3074622200728061e+02
```

Output file

This is an example of the *psuadeData* output file, which contains the simulation outputs for each of the samples that were generated in the configuration file. The *psuadeData* output file has an additional section, when compared to *psuade.in*: PSUADE IO. This section contains all sample points and their results. The first line of this section consists of three numbers, respectively: the number of input parameters, the number of output parameters and the number of runs of the simulation.

In our example, 800 samples were generated. The following file was edited to show only the results of the first two samples, to illustrate how the integration of PSUADE and MASE-BDI works. Eleven values are displayed in each simulation, concerning the simulation results for each one of the outputs. The order of the factors is the order defined in the *psuade.in* file.

```
PSUADE_IO (Note : inputs not true inputs if pdf ~=U)
4 11 800
1 1
  3.4000000000000000e+01
  7.0000000000000000e+01
  3.3366666666666663e+02
  5.0066666666666663e+02
  2.9891000000000000e+04
  6.8000000000000000e+01
  7.0000000000000000e+01
  5.1069961502892660e+01
  5.5050423090297905e+01
  8.7597757119669467e+01
  2.0478709634191286e+00
  1.4464230784458840e+01
  1.1810282602538216e+01
  2.6274513386997056e-01
  1.3858153565957343e-01
2 1
  3.4000000000000000e+01
```

```
  1.0000000000000000e+01
  3.3366666666666663e+02
  5.0066666666666663e+02
  2.5760000000000000e+04
  6.8000000000000000e+01
  1.0000000000000000e+01
  5.0855483774199428e+01
  5.5022603454271469e+01
  8.7038157581091738e+01
  2.1529446520538098e+00
  1.4456921310466861e+01
  1.1817592076530193e+01
  2.6274513386997056e-01
  1.3970536728584002e-01
>>>Edited file. Only two results were presented to illustrate the output
 file.<<<

PSUADE_IO
PSUADE
INPUT
   dimension = 4
   variable 1 transformationAgentQty = 1.0000000000000000e+00
1.0000000000000000e+02
   variable 2 transformationAgentGroupPercentage = 1.0000000000000000e+01
1.0000000000000000e+02
   variable 3 individualExploration = 1.0000000000000000e+00
5.0000000000000000e+02
   variable 4 groupExploration = 1.0000000000000000e+00
1.5000000000000000e+03
END
OUTPUT
   dimension = 11
   variable 1 time
   variable 2 qtyAgents
   variable 3 percentageAgents
   variable 4 figureOfMerit
   variable 5 producersaccuracy
```

```
   variable 6 usersaccuracy
   variable 7 wrongchange
   variable 8 rightchange
   variable 9 wrongpersistance
   variable 10 nullModel
   variable 11 simulatedNullModel
END
METHOD
#  sampling = MC
#  sampling = FACT
#  sampling = LH
#  sampling = OA
#  sampling = OALH
   sampling = MOAT
#  sampling = SOBOL
#  sampling = LPTAU
#  sampling = METIS
#  sampling = FAST
#  sampling = BBD
#  sampling = PBD
#  sampling = FF4
#  sampling = FF5
#  sampling = CCI4
#  sampling = CCI5
#  sampling = CCIF
#  sampling = CCF4
#  sampling = CCF5
#  sampling = CCFF
#  sampling = CCC4
#  sampling = CCC5
#  sampling = CCCF
#  sampling = SFAST
#  sampling = UMETIS
#  sampling = GMOAT
#  sampling = GMETIS
#  sampling = SPARSEGRID
#  sampling = LSA
```

```
#  sampling = RFF4
#  sampling = RFF5
   num_samples = 400
   num_replications = 1
   num_refinements = 0
   refinement_size = 10000000
   reference_num_refinements = 0
#  refinement_type = adaptive
   randomize
#  randomize_more
   random_seed = 12504321
END
APPLICATION
   driver = ./MASE-Driver.py
   opt_driver = NONE
   aux_opt_driver = NONE
   ensemble_driver = NONE
   ensemble_opt_driver = NONE
#  max_parallel_jobs = 1
#  min_job_wait_time = 1
   max_job_wait_time = 1000000
#  nondeterministic
#  launch_only
#  limited_launch_only
#  gen_inputfile_only
#  ensemble_run_mode
#  launch_interval = 1
#  save_frequency = 1000000
END
ANALYSIS
#  analyzer method = Moment
#  analyzer method = MainEffect
#  analyzer method = TwoParamEffect
#  analyzer method = ANOVA
#  analyzer method = GLSA
#  analyzer method = RSFA
#  analyzer method = MOAT
```

```
#  analyzer method = Sobol
#  analyzer method = Correlation
#  analyzer method = Integration
#  analyzer method = FAST
#  analyzer method = FF
#  analyzer method = PCA
#  analyzer method = ARSMGP
#  analyzer method = FORM
#  analyzer method = RSMSobol1
#  analyzer method = RSMSobol2
#  analyzer method = RSMSobolTSI
#  analyzer method = Bootstrap
#  analyzer method = RSMSobolG
#  analyzer method = ARSMNN
#  analyzer method = ARSM
#  analyzer method = REL
#  analyzer method = AOPT
#  analyzer method = GOWER
#  analyzer method = DELTA
#  analyzer method = ETA
#  analyzer method = ARSM
#  analyzer method = LSA
   analyzer output_id  = 1
   analyzer rstype = MARS
#  analyzer rstype = linear
#  analyzer rstype = quadratic
#  analyzer rstype = cubic
#  analyzer rstype = quartic
#  analyzer rstype = selective_regression
#  analyzer rstype = GP1
#  analyzer rstype = GP2
#  analyzer rstype = SVM
#  analyzer rstype = PWL
#  analyzer rstype = TGP
#  analyzer rstype = MARSBag
#  analyzer rstype = EARTH
#  analyzer rstype = sum_of_trees
```

```
#  analyzer rstype = Legendre
#  analyzer rstype = user_regression
#  analyzer rstype = sparse_grid_regression
#  analyzer rstype = Kriging
#  analyzer rstype = splines
#  analyzer rstype = KNN
#  analyzer rstype = RBF
#  analyzer rstype = Acosso
#  analyzer rstype = Bssanova
#  analyzer rstype = psuade_regression
#  analyzer rstype = RBFBag
#  analyzer rs_legendre_order = -1
#  analyzer rs_mars_num_bases = -1
#  analyzer rs_mars_interaction = -1
#  analyzer rs_num_mars = -1
#  analyzer rs_kriging_mode = -1
#  analyzer rs_kriging_tol = -1
#  analyzer opt_save_history
#  analyzer opt_use_history
#  analyzer regression_wgt_id = -1
#  graphics
#  sample_graphics
   analyzer threshold = 1.000000e+00
   rs_max_pts = 10000
#  analyzer rs_constraint = psData indexFile Lbnd Ubnd
#  analyzer moat_constraint = psData indexFile Lbnd Ubnd
#  analyzer rs_index_file = indexFile
#  optimization method = crude
#  optimization method = txmath
#  optimization method = appspack
#  optimization method = minpack
#  optimization method = sm
#  optimization method = mm
#  optimization method = mm_adaptive
#  optimization method = bobyqa
#  optimization method = sce
#  optimization method = moo
```

```
#  optimization method = ouu
#  optimization method = ouu1
#  optimization method = ouu2
#  optimization num_local_minima = 0
#  optimization use_response_surface
#  optimization print_level = 0
#  optimization num_fmin = 0
#  optimization output_id = 0
#  optimization max_feval = 10000
#  optimization deltax = 1.0e-6
#  optimization fmin = not defined
#  optimization cutoff = not defined
#  optimization tolerance = not defined
   printlevel  4
#  file_write matlab
#  use_config_file = NONE
#  use_input_pdfs
#  constraint_op_and
END
END
```

PSUADE analysis

An example of the PSUADE analysis is presented in the following extract. METIS size 800 sampling is loaded into memory and the tests are performed with the Delta Test (DT) sensitivity metric, which obtains the prioritization of the parameters according to the sensitivity of the simulation outputs.

```
**********************************************************************
*      Welcome to PSUADE (version 1.7.5)
**********************************************************************
PSUADE - A Problem Solving environment for
         Uncertainty Analysis and Design Exploration (1.7.5)
(for help, enter <help>)
======================================================================
psuade> load psDataMETIS800
readApplication WARNING: app driver ./MASE-Driver.py not found.
```

```
load complete : nSamples = 800
                nInputs  = 4
                nOutputs = 11
psuade> delta_test
Enter output number (1 - 11) = 4
No transformation (e.g. log) on sample inputs nor outputs.
**********************************************************************
DeltaTest for variable selection
This test has the characteristics that the more important
a parameter is relative to the others, the smaller the
subset is at the end of the test (sharp zoom into the most
important subset).
Thus, the purpose of this test is to identify a subset of
important parameters.
Note: If both nInputs and nSamples are large, this test
      may take a long time to run. So, be patient.)
======================================================================
Current best solution for output 4:
To stop the search, create a psuade_stop file in local directory.
----------------------------------------------------------------------
1 1 1 0  = 1.531177e+02
1 0 0 1  = 5.777625e+00 (1 of 100)
1 0 0 1  = 5.777625e+00 (2 of 100)
1 0 0 1  = 5.777625e+00 (3 of 100)
\vdots
1 0 0 1  = 5.777625e+00 (100 of 100)
**********************************************************************
Final Selections (based on 3 neighbors) =
Rank  1 => 1 0 0 1 : delta =  5.7776e+00
Rank  2 => 1 1 0 1 : delta =  3.0144e+01
Rank  3 => 1 0 1 1 : delta =  3.2172e+01
Rank  4 => 1 1 1 1 : delta =  5.3933e+01
Rank  5 => 1 0 0 0 : delta =  1.3737e+02
Rank  6 => 1 0 1 0 : delta =  1.4299e+02
Rank  7 => 1 1 0 0 : delta =  1.4644e+02
Rank  8 => 1 1 1 0 : delta =  1.5312e+02
Rank  9 => 0 0 1 1 : delta =  1.7724e+02
```

```
Rank 10 => 0 0 0 1 : delta =  1.8211e+02
----------------------------------------------------------------------
Delta test ranking is now in matlabdelta.m.
Order of importance (based on 20 best configurations):
(D)Rank    1 : input    1 (score = 89 )
(D)Rank    2 : input    4 (score = 87 )
(D)Rank    3 : input    2 (score = 26 )
(D)Rank    4 : input    3 (score = 25 )
**********************************************************************
Final test using the most important parameters incrementally:
----------------------------------------------------------------------
0 0 0 0  = 3.137032e+02
1 0 0 0  = 1.373724e+02
1 0 0 1  = 5.777625e+00
1 1 0 1  = 3.014355e+01
1 1 1 1  = 5.393320e+01
**********************************************************************
AnalysisManager: analysis metric = 5.78e+00
```

Appendix F

MASE-Driver

Configuration file

The following description is an excerpt of the configuration file that MASE-Driver uses to run the multiple simulations at MASE-BDI.

```
{
  "sampling_methods": [
    "MC",
    "FACT",
    "LH",
    "OA",
    "OALH",
    "MOAT",
    "SOBOL",
    "LPTAU",
    "METIS",
    "FAST",
    "BBD",
    "PBD",
    "FF4",
    "FF5",
    "CCI4",
    "CCI5",
    "CCIF",
    "CCF4",
    "CCF5",
```

```
    "CCFF",
    "CCC4",
    "CCC5",
    "CCCF",
    "SFAST",
    "UMETIS",
    "GMOAT",
    "GMETIS",
    "SPARSEGRID",
    "LSA",
    "RFF4",
    "RFF5"
  ],
  "refinement_types": [
    "adaptive"
  ],
  "lastConfiguration": {
    "output_variables": [
      {
        "name": "time"
      },
      {
        "name": "qtyAgents"
      },
      {
        "name": "percentageAgents"
      },
      {
        "name": "figureOfMerit"
      },
      {
        "name": "producersaccuracy"
      },
      {
        "name": "usersaccuracy"
      },
      {
```

```
      "name": "wrongchange"
    },
    {
      "name": "rightchange"
    },
    {
      "name": "wrongpersistance"
    },
    {
      "name": "nullModel"
    },
    {
      "name": "simulatedNullModel"
    },
    {
      "name": "steps"
    }
  ],
  "input_variables": [
    {
      "lowerBound": "1",
      "name": "transformationAgentQty",
      "upperBound": "100"
    },
    {
      "lowerBound": "10",
      "name": "transformationAgentGroupPercentage",
      "upperBound": "100"
    },
    {
      "lowerBound": "1",
      "name": "individualExploration",
      "upperBound": "500"
    },
    {
      "lowerBound": "1",
      "name": "groupExploration",
```

```
        "upperBound": "1500"
      }
    ]
  },
  "psuadeLocation": "/pathToPsuadeInstallFolder/psuade",
  "maseLocation": "/pathToMaseFolder/MASE-murl.jar"
}
```

Output file

The results presented in the output file are generated by MASE-BDI framework as a result of a simulation. The predicted maps are also generated and stored in a directory. The MASE-Driver captures each of these results and sends it automatically to PSUADE. The results can be analyzed in PSUADE with multiple UA and SA metrics.

```
6490
8
23
6.921068396839571
7.09748464124261
73.57606344628695
0.6697305545149708
1.8648295522033496
24.409683834793704
0.26274513386997056
0.25079414389308674
365
```

Appendix G

Software and Data Availability

MASE-BDI Software:

https://gitlab.com/InfoKnow/MASE/MASE-BDI/SourceCode/tree/master/MASE-BDI

PSUADE-MASE Software:

https://gitlab.com/InfoKnow/MASE/MASE-BDI/SourceCode/tree/master/MASE-PSUADE

Data Availability

The primary data derived from the model analysis are available for review, and replicability. The data is organized first by sampling method and then by sensitivity measure. The data is available at: https://gitlab.com/InfoKnow/MASE/MASE-BDI/SourceCode/tree/master/PSUADE%20Raw%20Data or at following QR code:

Figure G.1: Link to the primary data used in the model analysis