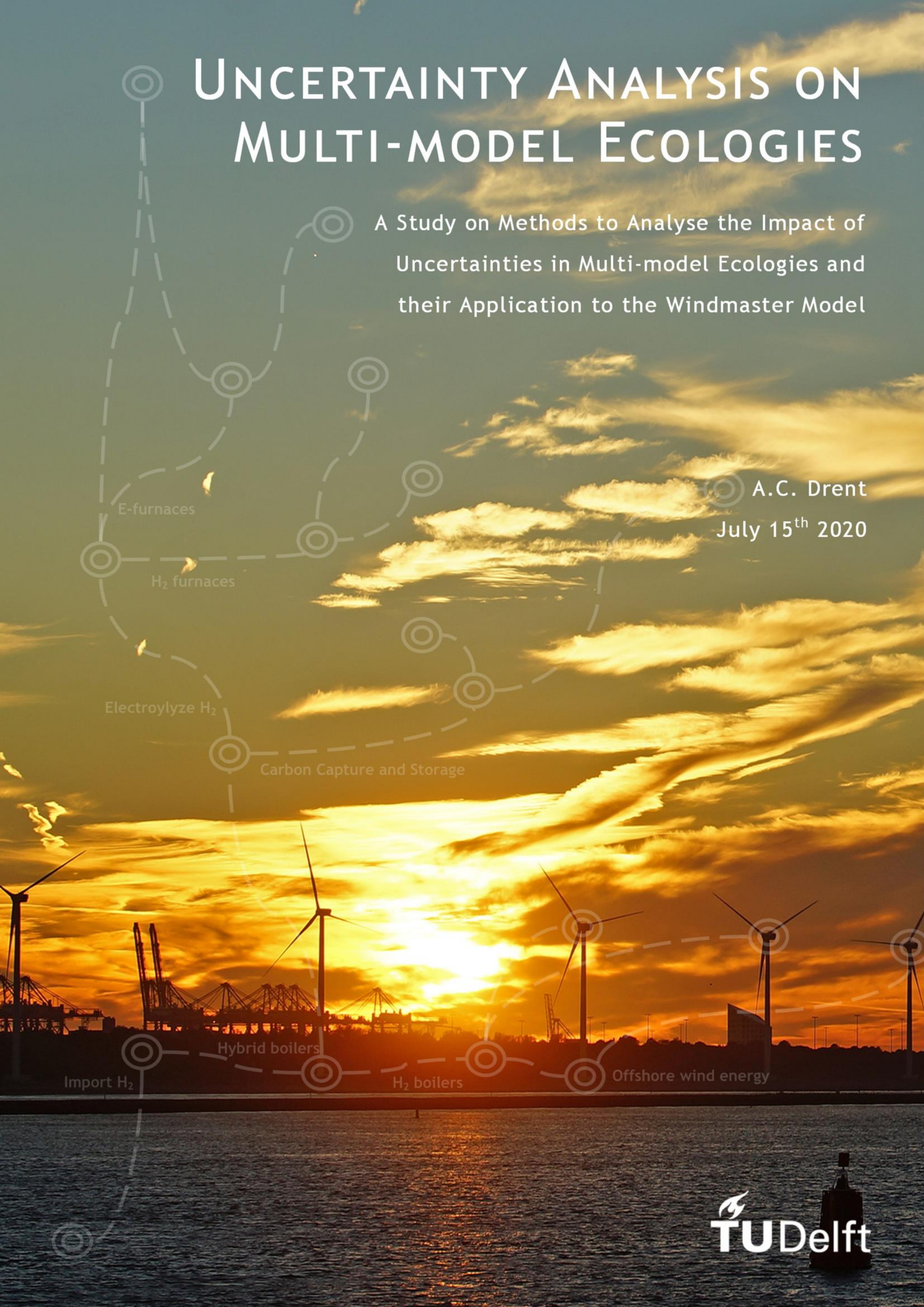
Uncertainty Analysis on Multi-model Ecologies

A Study on Methods to Analyse the Impact of Uncertainties in Multi-model Ecologies and their Application to the Windmaster Model

A.C. Drent

July 15th 2020



Cover photo: Torn, 2019

Uncertainty Analysis on Multi-model Ecologies

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **ENGINEERING AND POLICY ANALYSIS**

Faculty of Technology, Policy and Management

by

A.C. Drent

Student number: 4240367

To be defended in public on July 15th 2020

Graduation committee

First Supervisor : Dr.ir. J.H. Kwakkel, section Policy Analysis

Second Supervisor : Dr.ir. I. Nikolic, section Systems Engineering and Simulation

PREFACE

"There is nothing more certain and unchanging than uncertainty and change."

These words, already spoken some time ago by John F. Kennedy, reflect the uncertainties in our contemporary society that continuously affect our lives. Despite this ever-present uncertainty, how can we make better decisions? How to deal with these often deep uncertainties, is one of the most valuable insights I gained from (almost) completing my master's in Engineering and Policy Analysis. In my studies, I also developed an interest in simulation and modelling. It was inevitable that these two had to be combined in the thesis that is now in front of you. The importance of uncertainty analyses and the simulation and modelling of complex, socio-technical systems is now, I reckon, more relevant than ever.

At the beginning of my graduation, I imagined the end differently. Even though everyone had to withdraw a little, my supervisors, friends, and family had done a lot for me. This preface would not be complete without expressing my appreciation.

First of all, I would like to address my graduation committee. When I came to the two biggest geeks of the TPM faculty when putting together my thesis committee, I was afraid that I would feel ignorant. The opposite is true. In the beginning, we agreed that documenting "first" and "second" supervisors was only a distinction on paper. I am glad that this remained so until the end of my thesis. I find it extraordinary how enthusiastically you have guided and motivated me.

Jan Kwakkel, thank you for your guiding attitude in the world of uncertainty analysis. If I had read or come up with something, you would always be able to name a term or concept, and you would spoon down a number of articles relating to it. Despite your busy schedule, you made time to help me and to have extensive discussions.

Igor Nikolic, thank you very much for your help in writing my thesis. I really appreciated your relaxed attitude and sincere interest. You like to link the methods and uncertainty analyses to the real world: what real-life insight can we get out of this and how do we communicate this to people who understand it even less than I do? This has helped me to keep the link to the "real world".

I would like to thank my family for reading my thesis, hearing endless stories about uncertainty, multi-models, and how this combination can be applied to, for example, the port of Rotterdam. And of course for your unfailing support.

I would also like to thank my friends for the welcome distraction. With some of them, I went through the same graduation stages. That helped me a lot to keep going.

All that's left is for me to wish you a lot of fun reading my thesis.

Alexander Drent

Delft, July 15th 2020

Abstract

Our contemporary society consists of increasingly complex systems-of-systems in which technical and social subsystems influence each other. Multi-models reflect this complexity and are useful to analyse these complex socio-technical systems. They consist of connected simulation models which may each have their focus on a subsystem or modelling paradigm. However, these models contain different types of uncertainties influencing the fidelity of the results. Several methods are available to identify and analyse these uncertainties. It is yet unknown if and how existing uncertainty analysis methods can be applied to multi-model ecologies. In this thesis, we aim to provide for an answer to the following research question:

"To what extent can we apply existing uncertainty analysis methods to multi-models?"

To answer this question, it is important to first identify additional uncertainties in multi-model ecologies compared to single models. Next, we identify and apply methods to analyse these additional uncertainties. As proof of principle, a multi-model is used which focusses on the expansion and decarbonisation of the energy grid in the Port of Rotterdam.

There are three dimensions of uncertainty in simulation models: location, level, and nature. The different locations include the conceptual model, the computer model, input data, the technical model implementation, and the processed output data. In multi-models, we found an additional location: the interface. This is where the exchange of parameters between the models takes place. Within the multi-model locations, we identified some aspects that increase uncertainty. Epistemic opacity and computational expense are properties of multi-models that limit the analyst's knowledge of the multi-model and the feasibility of extensive uncertainty analyses.

The methods we identified for analysing these types of uncertainties are divided into sensitivity analysis and calibration. Sensitivity analysis quantifies the contribution of specific uncertainties to the overall uncertainty in the model outcome. Well-established methods are extra-trees feature scoring and Sobol. Calibration methods are based on the notion of equifinality. Using a specified likelihood function, they determine parameter values that lead to results with a high likelihood. Monte Carlo Markov Chain (MCMC) methods are often used for calibration purposes.

The applicability of these methods depends on the multi-model configuration. To describe these configurations, two limiting archetypes were used: directed graphs, and undirected graphs with feedback mechanisms over the model components. We found that uncertainty analysis of direct graphs can be carried out on both the model components and the whole. For undirected graphs, the research showed only added value in performing an uncertainty analysis on the whole multi-model. Otherwise, the changing context and emerging path dependencies cannot be included.

The application of extra-trees feature scoring, Sobol, and MCMC on the case study model showed that methods for uncertainty analysis are applicable on multi-models by including uncertainties on the interface. Furthermore, it is possible to reduce computational costs by factor fixing, distinguishing between deep and stochastic uncertainties, and assessing the convergence of sensitivity indices. Epistemic opacity can be dealt with by performing multiple replications and by including uncertainties related to the technical implementation of the multi-model. MCMC methods are suitable for scenario discovery, which provides insight into parameter values that lead to specific model results.

For future research on this topic, it is recommended to apply uncertainty analysis on multi-models with different network structures and a higher number of model components. This research offers a single case study multi-model, being an undirected graph with two model components. The role of uncertainties on the interface, epistemic opacity, and computational expense should then be further investigated in these different configurations. In addition, other uncertainty analysis methods in the context of multi-modelling could be investigated. Especially the application of moment-independent sensitivity analyses could be interesting to investigate further, since multimodal outcomes can arise from the interaction of heterogeneous models.

TABLE OF CONTENTS

“ *There is nothing more certain and unchanging than uncertainty and change.*

John F. Kennedy

1.

INTRODUCTION

1.1 Research Context

Large-scale, complex systems are an increasingly important component of our evolving modern-day society, consisting of both social and technical aspects, combined in a network of different actors. Examples of such large-scale socio-technical systems are the complex infrastructures used to provide energy, telecommunication, transports of goods and people, water management, and healthcare. Typical for these systems is that they show unpredictable, highly uncertain and emergent behaviour, which dynamically evolves. They often consist of subsystems, all influencing each other. An evolving field where these dynamics occur is, for example, the field of energy systems and its challenges relating to climate change, forcing the energy infrastructure to develop and become more sustainable (Herder, Bouwmans, Dijkema, Stikkelman, & Weijnen, 2008; Nikolic, Kwakkel, & Bletsis, 2019; van Dam, Nikolic, & Lukszo, 2013).

1.1.1 Simulation models

To analyse socio-technical systems, we use simulation models. These models are a theoretical, simplified representation of reality and help to assess the impact of policies and uncertainties to improve the real-world system. It is essential to consider that the outcomes of models only imitate some relevant features of real-world processes. Models use input data for initial and boundary conditions, to process in the technical implementation of the model. This implementation consists of all hardware and software that are involved in executing the simulation model. Simulation models, in turn, also generate data, which is generally processed using statistical methods and software. A vital element of this processing is the visualisation of this data to understand and interpret the outcomes. There are a lot of different methodological approaches to build these models and to prepare and process the input and output data (Petersen, 2006).

1.1.2 Multi-models

In the field of modelling and simulation, there is a growing need for tools capable of analysing large-scale socio-technical systems. A new approach is the use of multi-models, consisting of multiple models, connected in a multi-model ecology. These models may each have their own (modelling) paradigm and focus on a specific part of the (sub)system. In this way, a broader picture of the system can be taken into account since one formalism usually is not able to capture the complexity of the whole system. The individual models may focus on different technical or societal aspects of the system or both. The output of one model can be input for another model and vice versa. In this way, the models co-evolve and adapt to each other over time (Bollinger, Nikolic, Davis, & Dijkema, 2015; Carley, Morgan, Lanham, & Pfeffer, 2012; Mikulecky, 2001). Following Bollinger, Nikolic, Davis, & Dijkema (2015), we define a multi-model ecology as *"an interacting group of models and data sets co evolving with one another within the context of a dynamic socio-technical system"* (p.254).

Using and developing multi-models has many advantages. Multi-models enable us to investigate different uncertainty paths across the entire width of the system. This ensures that undesirable results can be prevented and allows us to make robust no-regret decisions that work regardless of how the future unfolds. Developing and testing simulation models is a time-intensive exercise, involving a high level of expertise. It is sensible to reuse these models in other configurations and couple them with other models. In decisions in which several parties are involved, the combination of multidisciplinary model formalisms increases the recognizability of the results. The principle of authority, combining different models developed by authorities on their specific domains, gives the combined multi-model an even more credible status.

1.1.3 Windmaster project

An example of a multi-model is the Windmaster project, conducted by the Delft University of Technology and Siemens, in collaboration with Gasunie, TenneT, Stedin, Port of Rotterdam, Deltalinqs, ISPT, and the province of South-Holland. This multi-model focusses on the energy transition in the industrial complex located at the Port of Rotterdam. It is developed to shape adaptive and robust investment policies to assure that the industrial complex in the Port of Rotterdam is decarbonised by 2050. The multi-model combines three models: an exploratory modelling scenario model, a technical-economical infrastructure model (which again consists of two sub-models for gas and electricity infrastructure) and a model that focusses on the investment behaviour of network operators (Wurth et al., 2019).

1.1.4 Challenges regarding multi-models

There are many challenges when developing multi-models (Nikolic, Kwakkel, et al., 2019; Nikolic, Warnier, et al., 2019). First, it needs to be technically possible to connect two or more models. This aspect is referred to as interoperability: the way the different models exchange data and how models are implemented and executed. Another aspect of connecting two or more models is composability. Composability deals with the question if it makes sense to connect these models on a conceptual level; are concepts across the different models defined in the same manner? Do assumptions in one model also stand when its outcome is transferred to another model? Composability consists of many aspects: consistency of syntax, semantics, and pragmatics, assumptions and validity. Validity concerns the question if the model is correct (Davis & Tolk, 2007). Validity closely relates to fidelity: the extent to which the model is capable of reproducing real-world behaviour (Feinstein & Cannon, 2001). Pace (2015) defines fidelity as a multi-faceted concept, consisting of attributes as accuracy, precision, timeliness, potential error sources and uncertainties, consistency, and repeatability.

1.1.5 Uncertainty

Uncertainties and potential error sources impact fidelity. In modelling and simulating socio-technical systems, there are a lot of different sources for uncertainty. These sources include stochastic variables and processes, a lack of accuracy and precision, and errors (Pace, 2015). A widely accepted definition of uncertainty is *"a lack of precise knowledge as to what the truth is, whether qualitative or quantitative"* (National Research Council, 1994, p. 161).

There are different kinds and levels of uncertainty identified in the available literature. Funtowicz & Ravetz (1990) differentiate between 'inexactness', 'unreliability' and 'border with ignorance'. Inexactness is expressed by using a range, and unreliability by a statistical confidence level. The 'border with ignorance' is impossible to express statistically. According to Walker et al. (2003), uncertainty is roughly divided into two types: uncertainty due to lack of knowledge and deep uncertainty. Deep uncertainty is defined as *"the condition in which analysts do not know or the parties to a decision cannot agree upon (1) the appropriate models to describe interactions among a system's variables, (2) the probability distributions to represent uncertainty about key parameters in the models, and/or (3) how to value the desirability of alternative outcomes"* (Lempert, Popper, & Bankes, 2003, p. 3). Pennock & Gaffney (2018) distinguish three kinds of uncertainty: aleatory uncertainty, epistemic uncertainty, and errors.

Aleatory uncertainty is any uncertainty or variation that is impossible to be reduced by new measurements, as opposed to epistemic uncertainty, where new measurements are able to reduce uncertainty. Epistemic uncertainty can be nestled in phase, structural, and ontological uncertainty.

1.1.6 Decision-making under deep uncertainty

Uncertainties impact our ability to make decisions. Often our reaction is to do more research, make more predictions, or use more models. Following the aphorism by Box (1976), all those models have one thing in common: they are all wrong. Due to the complexity of socio-technical systems and a high degree of uncertainty, we are unable to sketch an unambiguous picture of the future.

Decisions concerning large socio-technical systems are never made by one party. Different parties are involved who look at the same problem in a different way. They often disagree on assumptions, problem definitions, desired outcomes and how the system works. They interpret the same results or data differently and claim to know how the system works. We cannot predict the future, so we do not know who is right. That is why we have to look at problems and systems in a broader perspective using uncertainty analysis and multi-models. If we have to make decisions for the future now, we have to accept the associated uncertainty: given that we do not know, what is the best we can do?

1.2 Research Question

When creating multi-models, different problems arise around the composability of two or more models. Composability is affected by the validity and fidelity of the models. Factors influencing the fidelity include uncertainties and errors in the models. Since models are subject to different types of uncertainty, ranging from errors to epistemic and deep uncertainty, these uncertainties might propagate when coupling these models in a multi-model ecology (Nikolic, Kwakkel, et al., 2019; Nikolic, Warnier, et al., 2019). To analyse the impact of uncertainties and errors in (single) simulation models, a range of methods is available. These methods are capable of assessing the influence of varying input parameters on the outputs of the model. However, it is uncertain if these methods also provide sufficient insight into uncertainty in the domain of multi-models. Also, there are no known methods to assess the uncertainty propagation in multi-models. To be able to assess the validity of such models, methods are needed to provide insight into this uncertainty propagation.

The knowledge gap this research is focussing on is the ability to apply existing methodologies for assessing the impact of uncertainties on single simulation models on multi-model ecologies. This research will be conducted as part of graduating from the master Engineering & Policy Analysis (EPA). Typical EPA- problems involve multiple actors, faced with a wicked problem, where the problem is perceived differently by different actors, and no clear solution is available (Rittel & Webber, 1973). To be able to gain a better understanding of complex systems, we need to be able to develop models, reflecting the complexity of large-scale socio-technical systems. This means we need to gain an understanding of how uncertainty propagates in a multi-model ecology and how this impacts the model results.

Therefore, the research question is defined as follows:

“To what extent can we apply existing uncertainty analysis methods to multi-models?”

To be able to understand the uncertainty propagation in multi-models, we use a case study approach, using an existing multi-model to apply single-model methods for uncertainty analysis. To answer the main question, we distinguish the following sub-questions:

1. What additional sources of uncertainties exist in multi-model ecologies in comparison to single models?

This question focusses on sources of uncertainties. There are currently frameworks available for describing the uncertainty in single models. The aim is to give insight into what is different and what may be additional sources of uncertainty in multi-model ecologies. The question will be answered using a literature study.

2. What methods for uncertainty analysis exist to analyse uncertainty propagation in single-models?

This question focusses on methods to analyse uncertainties in single-models. The goal is to select methods and compare them to make a selection of methods that will be put to further use in the research. The question will be answered through a literature study.

3. How can these methods be applied to analyse a whole multi-model ecology?

Based on the provided additional uncertainties in multi-models and the available methods, this question aims to develop an understanding of how these methods can be applied on a multi-model ecology. The selected methods are going to be used on a case study multi-model, the Windmaster model, as proof of principle. Based on this uncertainty analysis, we aim to conclude on uncertainty analyses of multi-models in general.

1.3 Structure

The structure of this research is as follows. It starts with elaborating on uncertainty in simulation models and multi-models in particular. Next, the available methodologies are discussed and subsequently, a general approach to the use of these methods in the context of multi-modelling is given. This approach will be used to apply to the Windmaster model. Based on this specific case, a synthesis will be made regarding the use of uncertainty analysis on multi-models.

2.

Uncertainty in Multi-model Ecologies

This chapter aims to answer the first sub-question: "What extra sources of uncertainties exist in multi-model ecologies in comparison to single models?" First, a synthesis will be made of what is known about uncertainty regarding single simulation models. This will be based on definitions and frameworks found in the literature. This synthesis will lead to a framework to assess uncertainty in single models, which will subsequently be supplemented with sources of uncertainties that are expected to play a role in multi-model ecologies.

2.1 Dimensions of Uncertainty

Uncertainty consists of many aspects, and there are different types of uncertainty. Uncertainty is typified using three dimensions: location, level, and nature. Location concerns the source of uncertainty, the level of uncertainty deals with the severity of uncertainty, and the nature of uncertainty describes whether this uncertainty is part of the system, or caused by lack of knowledge or consensus amongst experts (Kwakkel, Walker, & Marchau, 2010; Walker et al., 2003). In the next three paragraphs, we discuss these dimensions in more detail.

2.1.1 Location of uncertainty

The location of uncertainty refers to which part of the model the uncertainty occurs. Based on Kwakkel et al. (2010), Petersen (2006), and Walker et al. (2003), we distinguish five locations of uncertainty:

- Conceptual model
- Computer model
    - Model structure
    - Model parameters
- Input data
- Technical model implementation
- Processed output data

The conceptual model is at the essence of the simulation model. It determines the context and boundaries of the system and what general theories, principles, assumptions, and relationships underlie the model. Due to these demarcations -what is included and what is not- and assumptions, uncertainties may arise in this location. There may also be alternative or even contradicting conceptual models, based on other theories, concepts, or relationships, which could be a source of uncertainty as well.

The computer model comprises how the conceptual model is translated into computer code using model parameters, relations, and mathematical functions. In Petersen (2006), this location is referred to as the 'mathematical' model. We subdivide the computer model location into the structure of the model and the used parameters. These parameters are inside the model or used as a representation of an external effect or policy.

The input data determine the initial values and boundary conditions of model parameters. As input data may, for example, come from other simulation models, experiments, observations, or 'guesstimations', there might be different sources of uncertainty in this location.

The technical model implementation consists of all hardware and software used for implementing and running the simulation model. Uncertainties in this location could be caused by bugs and errors in the model code or hardware.

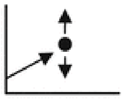
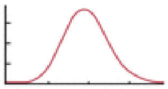
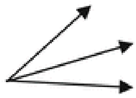
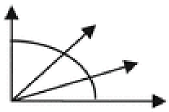
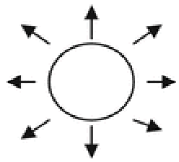
Simulation models, in turn, also generate data which is generally processed and visualised using statistical methods and software to understand better and interpret the outcomes of the model. The simulation model's output data is the location where the uncertainties of the other locations stack after having propagated during a simulation run. Additional uncertainties may arise during the processing of the output data and in the interpretation of this data (Kwakkel et al., 2010; Petersen, 2006; Walker et al., 2003).

2.1.2 Level of uncertainty

The level of uncertainty indicates the degree, or severity, of uncertainty. On the two extreme sides of this dimension are deterministic knowledge, when there is no uncertainty at all, and total ignorance, where we can say nothing knowledgeable because we simply do not know that we do not know. Between these two extreme levels of uncertainty, we define five levels of increasing uncertainty: from marginal uncertainty, shallow uncertainty, medium uncertainty, deep uncertainty, to recognised ignorance. These levels, included in Table 1, are defined by Pruyt & Kwakkel (2014):

- Deterministic knowledge applies to systems in which there is no uncertainty involved. We consider this type of uncertainty the 'zero-point' on the uncertainty scale.
- Level 1: Marginal uncertainty addresses systems in which there might be some recognised uncertainty, but the influence of this uncertainty on the outcomes of the model is deemed negligible.
- Level 2: Shallow uncertainty applies to systems in which some uncertainties are present that can be described sufficiently in statistical terms. This kind of uncertainty is captured through a forecast within a confidence interval or multiple forecasts (scenarios) with certain probabilities.
- Level 3: Medium uncertainty refers to uncertainties in which scenarios are developed, but the probability of these scenarios cannot be expressed in statistical terms. However, they can be arranged in order from probable to unlikely.
- Level 4: Deep uncertainty applies to situations in which we may develop future scenarios, but we do not know or cannot agree upon the probability of occurring, and we cannot determine if one scenario is more or less likely to occur than another scenario. Deep uncertainty can be defined as *"the condition in which analysts do not know or the parties to a decision cannot agree upon (1) the appropriate models to describe interactions among a system's variables, (2) the probability distributions to represent uncertainty about key parameters in the models, and/or (3) how to value the desirability of alternative outcomes"* (Lempert et al., 2003, p. 3).
- Level 5: Recognised ignorance is the most severe level of recognised uncertainty. It refers to situations in which we have no clue about the underlying mechanisms in the system, and we are unable to estimate the possibility of certain events happening. The only thing that is known is that we do not know. In the literature, this kind of uncertainty may also be referred to as 'black swans' (Taleb, 2015) or 'coconut' uncertainty (Makridakis, Hogarth, & Gaba, 2009), indicating that we know certain phenomena or events exist or may occur, but we cannot predict them nor explain the underlying causes.
- Total ignorance is the severest level of uncertainty. In these situations, we can say nothing about the uncertainty, because we do not know that we do not know. This type of uncertainty acts as the other limiting case on the uncertainty scale. Since this type of uncertainty cannot be specified, it is not covered by recognised uncertainty anymore.

Table 1 Five levels of uncertainty as found in Walker et al. (2013, p. 4): marginal uncertainty (level 1), shallow uncertainty (level 2), medium uncertainty (level 3), deep uncertainty (level 4), and recognised ignorance (level 5). Two extreme levels of uncertainty demarcate these levels: deterministic knowledge on the left side and total ignorance on the right side.

			LEVEL					
			Level 1	Level 2	Level 3	Level 4	Level 5	
LOCATION	Context	Complete Certainty	A clear enough future 	Alternate futures (with probabilities) 	Alternate futures with ranking 	A multiplicity of plausible futures 	An unknown future 	Total ignorance
	System model		A single (deterministic) system model	A single (stochastic) system model	Several system models, one of which is most likely	Several system models, with different structures	Unknown system model; know we don't know	
	System outcomes		A point estimate for each outcome	A confidence interval for each outcome	Several sets of point estimates, ranked according to their perceived likelihood	A known range of outcomes	Unknown outcomes; know we don't know	
	Weights on outcomes		A single set of weights	Several sets of weights, with a probability attached to each set	Several sets of weights, ranked according to their perceived likelihood	A known range of weights	Unknown weights; know we don't know	

2.1.3 Nature of uncertainty

Nature concerns whether uncertainty is caused by variability in the real-world system, or by lack of knowledge. Two classifications are used: ontic and epistemic uncertainty. Ontic uncertainty applies to systems that are chaotic by nature. Therefore, this kind of uncertainty cannot be reduced by increasing knowledge since it is unpredictable. This type of uncertainty, dealing with situations we cannot know, may also be referred to as 'aleatoric' or 'statistical' uncertainty (Kwakkel et al., 2010; Petersen, 2006; Van den Hoek, Brugnach, Mulder, & Hoekstra, 2014; Walker et al., 2003).

A lack of knowledge causes epistemic uncertainty. It refers to situations we do not know (yet), but could potentially be reduced. To deal with epistemic uncertainty, one could perform more research or involve experts (Kwakkel et al., 2010; Pennock & Gaffney, 2018; Petersen, 2006; Van den Hoek et al., 2014). However, epistemic uncertainty could also be caused by the fact that experts or stakeholders do not agree on the underlying mechanics of the system, or they interpret data differently (Kwakkel et al., 2010). This is caused by differences in perception, opinion, and values. Kwakkel et al. (2010) and Van den Hoek et al. (2014) call this ambiguity. This kind of uncertainty is fundamentally different as it does not refer to things we do not or cannot know, but things that actors know differently. We follow Van den Hoek et al. (2014): *"Ambiguity refers to the situation in which there are different and sometimes conflicting views on how to understand the system to be managed"* (p. 375). We divide epistemic uncertainty into two subtypes: uncertainties caused by not knowing enough (incomplete knowledge) and uncertainties caused by knowing differently (ambiguity).

2.1.4 Framework for assessing uncertainty

We present a framework to assess uncertainty in simulation models. This framework is based on the uncertainty matrices as proposed by Kwakkel et al. (2010, p. 311), Petersen (2006, p. 50), and Walker et al. (2003, p. 11). The uncertainty matrix is included in Table 2.

Table 2 Uncertainty matrix. A framework to assess uncertainty in simulation models by using three dimensions: location, level, and nature. Location concerns the part of the model where the uncertainty manifests itself, the level of uncertainty deals with the severity of uncertainty and the nature of uncertainty describes whether this uncertainty is ontic or epistemic. Since epistemic uncertainty can be caused by not knowing enough or knowing differently, this type of uncertainty is divided into incomplete knowledge and ambiguity.

Uncertainty matrix			**Level**					**Nature**		
								Epistemic		Ontic
			Level 1 Marginal uncertainty	Level 2 Shallow uncertainty	Level 3 Medium uncertainty	Level 4 Deep uncertainty	Level 5 Recognised ignorance	Incomplete knowledge	Ambiguity	
Location	Conceptual model									
	Computer model	Model structure								
		Model parameters inside the model								
		Input parameters to the model								
	Input data									
	Technical model implementation									
	Processed output data									

2.2 Additional Uncertainties in Multi-model Ecologies

In a single model, dealing with uncertainty, variability, and noise is hard. In multi-models, dealing with uncertainty might be even harder. In multi-model ecologies, the outcomes of one model, which are subject to uncertainties, are fed into another model, which are then permutated. This relates to uncertainty propagation, which is the effect that uncertainty in the input variables spreads to an even more substantial degree in the outcomes of a model.

In this paragraph, we look into additional sources of uncertainty that may play a role in these multi-model ecologies. These sources include a new location of uncertainty, extra sources of uncertainty within the defined locations, and aspects of multi-models that make it harder to analyse uncertainty propagation. The uncertainty matrix will be supplemented accordingly. First, we will go into aspects of multi-models that may play a role in all locations of the multi-model. Second, we will discuss what extra sources of uncertainty there are within existing and new locations.

2.2.1 Aspects

Epistemic opacity

In multi-model ecologies, models may often come from different modellers. This makes it difficult or even impossible to comprehend their internal processes, assumptions, and the exact meaning of the output variables and input parameters. This aspect is referred to as epistemic opacity. Not being able to understand the internal processes of a model entirely is at the same time the main reason for using simulation models at all as otherwise, we would not have to use them (Humphreys, 2009).

Computational expense

Simulation models may take a long time to run. As multi-models are dependent on the combined execution time of different models, the runtime increases considerably. Uncertainty analysis methods rely on running a high number of model runs. Computing power is often limited, so there is a trade-off between computational expense and extensiveness of the analysis. Even when we consider clusters of high-performance computing environments, the runtime might still be too long to perform, for instance, a full factorial design.

2.2.2 Locations

Conceptual model

Conceptual models of the multi-model level describe how concepts from different simulation models influence each other. As each simulation model in the multi-model ecology has its conceptual model as well, the uncertainties quickly become more extensive, and the conceptual models might contradict each other or make assumptions that will not hold in the theoretical framework of the other model. Uncertainties may arise because the paradigms of each model are different, and relations may be subject to discussion.

Computer model

Dealing with uncertainty in multi-modelling depends on how the multi-model ecology is structured. There are a lot of different possible configurations with each their own specific additional uncertainties. These may situate in the timing and process of exchanging information between the different models.

Connecting models from different modelling paradigms may lead to unexpected behaviour, for instance, if a deterministic simulation model receives chaotic input from agent-based models (ABM) or discrete event simulation (DES) models. These models rely on stochastic processes, which cause variance in the outcomes. Stochasticity may lead to different outcomes, based on the same chosen input variables.

Input data

In a single model, we can largely control the input that is given to a model through data cleaning and determining the initial value and upper and lower boundaries of parameters. In a multi-model ecology, the exchange of in- and output from the different models is largely autonomic. In combination with the property of epistemic opacity of multi-models, one or more models might be pushed into a parameter space for which it was not designed. The model might still function and give output, but these results might be highly unreliable.

Simulation models in a multi-model ecology may use scattered input data coming from, for instance, other models, external data sources, or educated guesses, which makes it hard to keep track of all the inputs that are fed to the models. Therefore, it might be the case that parameter values used in the different models are contradictory or should not be used in combination with each other.

The curse of dimensionality plays a role as well, which refers to various problems that may occur when applying analysis methods that work well with relatively low dimensionality of the input space to high-dimensional input space. Dimensionality increases through the growing number of parameters in multi-models, used for the individual models, or to switch between different sub-models. Due to the rapidly increasing complexity of the input space, methods may not work or may take an extremely long time to calculate, especially in combination with multi-models that have a long execution time.

Technical model implementation

The multi-model infrastructure enables models to interact. This infrastructure may be distributed amongst different computer systems and even different organisations. This results in the proliferation of all involved hardware and software in multi-model ecologies. Especially when there are a lot of heterogeneous computer systems involved, the simulation models running on these systems may respond differently or may not respond at all.

Processed output data

The processing and interpretation of output data of multi-models are bound to be subject to more sources of uncertainty. Although the processing of the output data may not be as complex on its own, the interpretation of the output is considerably more difficult, because one has to deal with all additional sources and locations of uncertainty.

Model interface

Multi-model ecologies depend on interactions between models. These interactions enable the different models to use each other's output as input. This process takes place between model interfaces. A model interface is key in a multi-model ecology, as it enables simulation models to exchange information. In the process of exchanging information, additional uncertainty may occur. Following the broad notion of a model interface as provided by Nikolic, Warnier et al. (2019), we focus on the social interaction part of the model interface. The backend connectors and other components which are needed to support this social interaction are allotted to the technical model implementation location.

Let us consider two simulation models in a multi-model ecology, models A and B, which are connected via their model interfaces and use each other's output as input. When the output of model A is transferred to the interface of model B, this output is subject to the accumulation of uncertainties of model A, propagated during a simulation run. Model B performs its calculations and feeds its output back to model A. This input for model A is now subject to uncertainty from the first step of model A and the additional uncertainty of model B, which was also subject to uncertainty from model A. Depending on the structure of the multi-model ecology, this kind of interaction may take place frequently per time step, between multiple models. Wilby & Dessai (2010) describe this process, causing a cascade of uncertainty, as the envelope of uncertainties is expanded every step of the model.

Another source of uncertainty is in the exchange of (numerical) information. Depending on the modelling paradigm of a model, the output of a model could be highly chaotic. This is the case in, for instance, ABM or DES models. The variability in the output of these models is high, caused by a lot of stochastic parameters and processes. Each time the model runs, the output might be different. Dealing with this kind of intrinsic uncertainty might be hard when we need to transfer this uncertainty as input for other models. It might not be trivial to determine which descriptive to use to communicate the uncertainty associated with the outcomes. Are we interested in the mean value, where the bulk is, or are we mainly interested in that one outlier?

2.2.3 Framework for assessing uncertainties in multi-model ecologies

With the identification of a new location for uncertainty in multi-model ecologies, we extend the framework with the interface location, included in Table 3. One might argue that the interface could also be part of the “model structure” location, and that is right. Therefore, it might be beneficial further to define the existing locations within the uncertainty framework.

The existing locations conceptual model, input data, and processed output data can both concern the multi-model ecology and the individual models. The technical model implementation can be applied to the individual models as well, while in the context of the multi-model ecology, it focusses on the required infrastructure to enable model interaction. The computer model location includes uncertainties within individual models and their parameters and structure. With the introduction of the new interface location, we focus on the interaction between the different involved computer models within the multi-model ecology. A conceptual overview of locations for a general multi-model, including two individual simulation models is included in Figure 1.

The level and nature of uncertainties also apply to a multi-model ecology. Within these dimensions, we see no aspects that could engage other levels or different kind of natures.

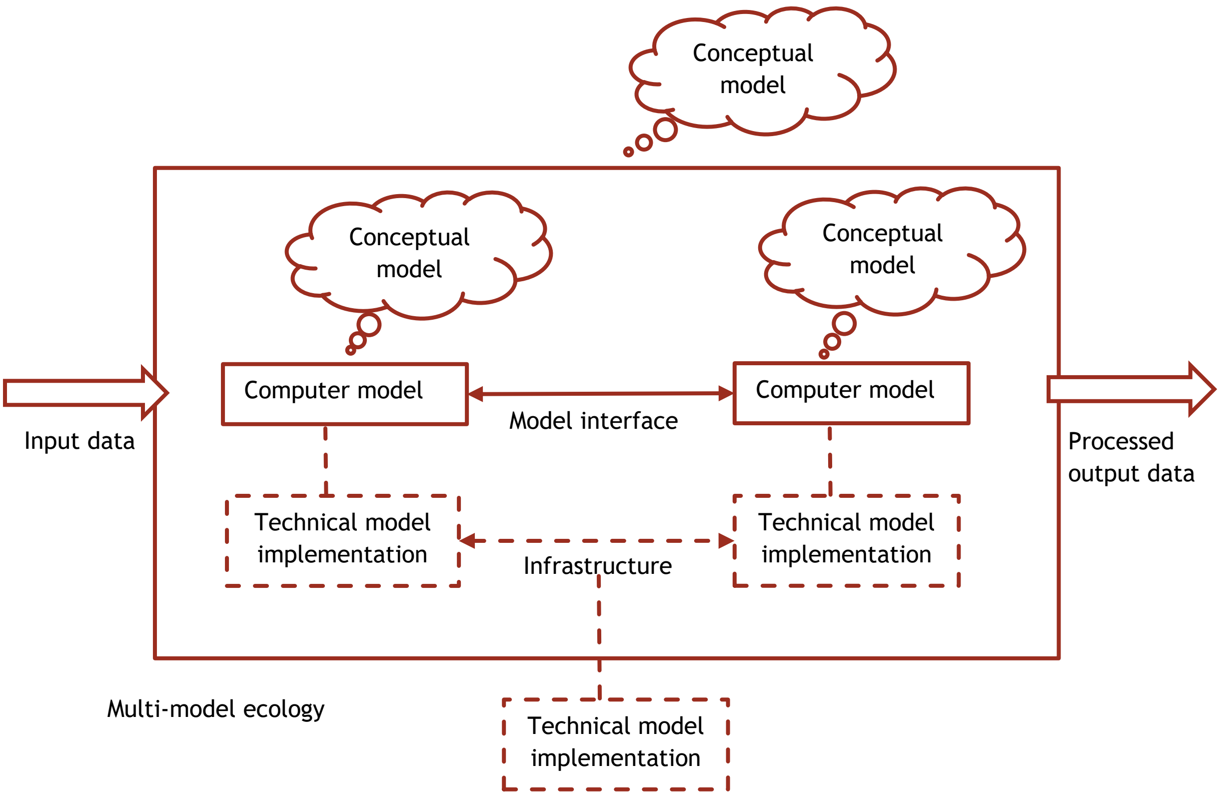


Figure 1 Conceptual overview of the uncertainty locations within the uncertainty matrix for a general multi-model, including two simulation models. The new uncertainty location ‘Model interface’ lies within the different sub-models and concerns the social part of the interaction between the models. Actual communication in the technical sense is covered by the technical model implementation and ‘infrastructure’ of the multi-model.

Table 3 Multi-model uncertainty matrix. A framework to assess uncertainty in multi-models by using three dimensions: location, level, and nature. Location concerns the part of the multi-model or individual simulation model where the uncertainty manifests itself, the level of uncertainty deals with the severity of uncertainty and the nature of uncertainty describes whether this uncertainty is ontic or epistemic. Since epistemic uncertainty can be caused by not knowing enough or knowing differently, this type of uncertainty is divided into incomplete knowledge and ambiguity. For illustrative purposes, the computer model location is divided into two parts: one for a hypothetical computer model A and one for a hypothetical computer model B. It will depend on the exact configuration and on the focus of the uncertainty assessment of a multi-model ecology how the different locations are divided over the individual simulation models and the multi-model ecology as a whole.

Multi-model uncertainty matrix			**Level**					**Nature**		
								Epistemic		Ontic
			Level 1 Marginal uncertainty	Level 2 Shallow uncertainty	Level 3 Medium uncertainty	Level 4 Deep uncertainty	Level 5 Recognised ignorance	Incomplete knowledge	Ambiguity	
Location	Conceptual model									
	Computer model A	Model structure								
		Model parameters inside the model								
		Input parameters to the model								
	Model interface									
	Computer model B	Model structure								
		Model parameters inside the model								
		Input parameters to the model								
	Input data									
	Technical model implementation									
	Processed output data									

3.

METHODOLOGIES FOR UNCERTAINTY ANALYSIS

To deal with uncertainty in simulation models, a variety of methods is available. These methods will be discussed in this chapter, which aims to answer the second sub-question: "What methods for uncertainty analysis exist to analyse uncertainty propagation in single-models?" We start with a quick overview of relevant methods, describing the general idea of these existing methods. From this overview, we select the methods that will be used in the uncertainty analysis of the Windmaster multi-model ecology.

From a policy-making perspective, the policy system, captured in a model, can be seen as a function, which output is influenced by uncertainty and policies. To capture this, we use the XPIROV framework included in Figure 2. XPIROV is a policy problem structuring framework, introduced by Agusdinata & DeLaurentis (2008).

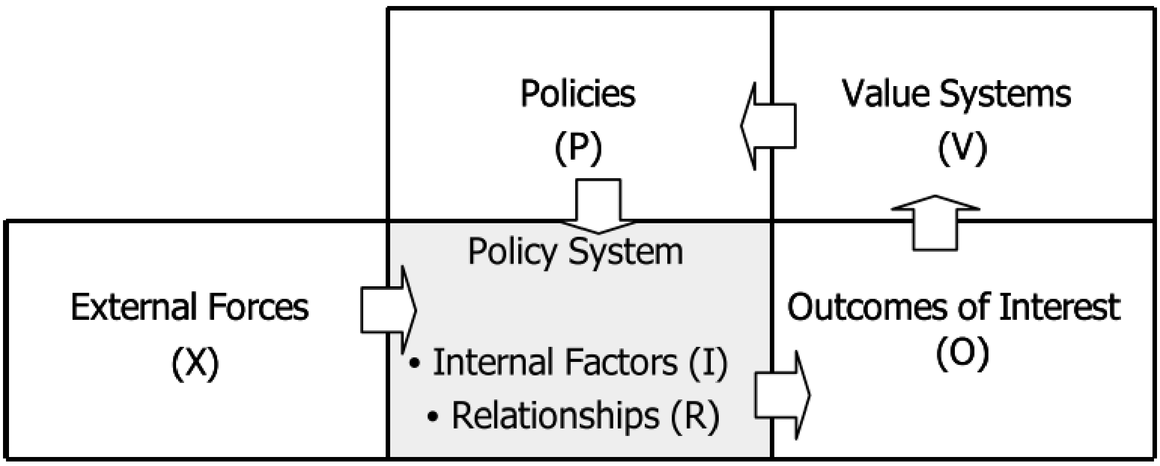


Figure 2 XPIROV policy analysis framework, as obtained from Agusdinata & DeLaurentis (2008, p. 5).

XPIROV consists of six elements: external forces (X), policies (P), the policy system under study consisting of internal factors (I) and relationships (R), outcomes of interest (O), and the value systems of the stakeholders and policymakers (V). The framework is used by policy analysts to present and structure the system under investigation (Agusdinata, 2008; Agusdinata & DeLaurentis, 2008; Castrejon-Campos, Aye, & Hui, 2020). The framework is an extension of the XLRM framework as introduced by Lempert et al. (2003).

External forces, or exogenous uncertainties, are factors that are out of control by the stakeholders and may influence the success of their strategy. Policies, or policy levers, are instruments that the decision-maker can deploy to influence the behaviour of the policy system to strive for better outcomes. Internal (endogenous) factors are part of the policy system and are influenced by the external forces and policies. Relationships link and influence the external forces, policies, and internal policies and produce the outcomes of interest. Outcomes of interest measure the performance of the policy system and relate to the objectives of stakeholders. The value systems reflect the goals, objectives, and preferences of the stakeholders. Based on these value systems, the desirable outcomes of interest and policies can be determined (Agusdinata, 2008; Agusdinata & DeLaurentis, 2008; Castrejon-Campos et al., 2020; Lempert et al., 2003).

To gain insight into the different possible outputs of the model, we make use of uncertainty analysis. Uncertainty analysis is the practice of quantifying the uncertainty in model outputs. It relies on the propagation of uncertainty in model inputs to the model outputs. A closely related method is sensitivity analysis (SA), which focusses on *"how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input"* (Saltelli, Tarantola, Campolongo, & Ratto, 2004, p. 45). Another purpose for uncertainty analysis is the calibration of model parameters to fit observed data or, more generally, a statistical model. This method is known as Bayesian inference and gives insight into the distribution of parameter values for which the resulting outcomes of a simulation model are in line with a target model or observed data.

3.1 Sensitivity Analysis

While uncertainty analysis focusses only on the uncertainty of the output, sensitivity analysis (SA) attempts to quantify what the sources of this uncertainty are. We distinguish two main approaches in sensitivity analysis: local and global sensitivity analysis (GSA). Local sensitivity analysis is executed by varying one input parameter around a defined base value while keeping the values of the other parameters at a constant level. The method studies how small variations around a base value affect the output and is particularly suitable for linear models. Global sensitivity analysis is an approach where the entire range of values of input parameters is considered, without defining some base value (Saltelli et al., 2008). We focus on the latter form of sensitivity analysis, as local methods do not adequately explore uncertainty in models with non-linearities (Saltelli et al., 2019).

3.1.1 Independent sampling methods

Sensitivity analysis methods use model parameter value sets which are specified a priori. We refer to this as independent sampling. We distinguish One-At-a-Time (OAT) and All-At-a-Time (AAT) sampling. OAT samples the value of one certain input parameter while keeping the others constant. AAT varies the value of all the input parameters simultaneously. The benefit of All-At-a-Time sampling is that interaction effects between input parameters can be evaluated, which is not possible using OAT sampling. The main drawback of AAT is that it is more computationally expensive, as it requires extensive sampling. Mainly, there are three commonly used AAT sample methods: Monte Carlo (MC), Latin Hypercube Sampling (LHS), and Sobol sequences (Pianosi et al., 2016).

MC sampling draws random values within the defined boundaries for each uncertainty. One combination of these different random parameter values makes for one sample, or experiment. Draws of Monte Carlo samples are purely random. This implies that all individual experiments do not correlate with other experiments. On its turn, this implies that we get no guarantees about the quality of the sample in terms of clustering or drawing samples across the whole distribution range. We might end up with a sample that focusses mainly on certain parts of the total input space. This may be solved by adding more samples.

To get more guarantees on the sampling quality, we might choose for Latin Hypercube Sampling. LHS subdivides the total range of one parameter into equal probable intervals and draws the same number of random samples from each interval. The draws are shuffled to combine different draws from one parameter with draws from the other parameters. Since the number of intervals for each parameter is dependent on the number of samples, we know that the whole input space is covered. The disadvantage of using LHS is that we need to know on beforehand how much samples we need to draw since it is not possible to simply add more samples if we might conclude that we need to generate more samples. The advantage of LHS is that generally, fewer samples are needed to reach the convergence of the estimated input space.

The Sobol sequence sampling method, like LHS, subdivides the input space into subspaces. However, Sobol subdivides the whole multi-dimensional input space at once. This ensures a better spread of the samples in the input space, at the cost of being more computationally expensive than both MC and LHS. An advantage of Sobol sequence sampling is that it is possible to add more samples, which will increase the number of subspaces from where to draw samples.

3.1.2 Objectives

Through the use of sensitivity analysis methods, one can classify the input variables into factors with (1) a high main effect, (2) a small main effect, but high total effect, and (3) variables with both a small main and total effect. The first category contains variables that mainly drive the output without interaction effects. Variables in the second class influence the model chiefly through interaction effects with other variables, while the third category contains variables that have a negligible influence on the output of the model. Therefore, SA is used for different applications, depending on the purpose of the analysis (Saltelli et al., 2008):

- Factor prioritisation (FP)
- Factor fixing (FF)
- Variance cutting (VC)
- Factor mapping (FM)

Factor prioritisation (also referred to as 'ranking') is used for identifying those parameters that have a great influence on the output of the model. These variables would fall into the first or second category. If these parameters were fixed at the base value, the variation of the output would decrease. These parameters might be interesting to focus on, as reducing the uncertainty of these parameters could potentially reduce the uncertainty of the model output. Factor fixing (also referred to as 'screening') can be understood as the opposite of factor prioritisation: it identifies the parameters which have a minimal effect on the variance of the output. These parameters, which are covered by the third class, could be fixed on an arbitrary value within its bounds, leading to fewer parameters to consider in further analysis. Variance cutting aims to bring the output uncertainty below a determined threshold by fixing the lowest amount of input values as possible. Factor mapping determines which region of the output space is associated with which region in the input space (Saltelli et al., 2008).

3.1.3 Sobol

One of the most used SA methods is variance-based sensitivity analysis, often referred to as the Sobol method. This method is a form of GSA with AAT sampling and determines which proportions of the output variance contribute to the variance of the model inputs while taking interaction effects between input parameters into account. These proportions are quantified in the Sobol sensitivity indices (SI's). SI's are divided in first, second, and total order indices. The first-order sensitivity indices are measures for the direct effect of one individual parameter's variance on the output variance. Second-order indices are the combined effect of two parameters' variances on the variance of the output. The total order indices indicate the total effect of one variable's variance on the output variance, including the main effect and all its interactions with other parameters (Ghanem, Owhadi, & Higdon, 2017; Saltelli et al., 2008; Sobol, 2001).

Although Sobol is computationally expensive, the method is quite popular, as the method can deal with interaction effects between parameters and the Sobol indices are easy to interpret. Another benefit of Sobol is that the method is model-independent, so the model itself does not have to be moderated to work. The method uses the model as a so-called 'black box': it feeds different input values to the model and record its output. Despite this 'black box' approach, the method allows more insight into the internal working of the model and the recognition of strongly related processes, through the second-order Sobol indices (Ghanem et al., 2017; Saltelli et al., 2008; Sobol, 2001).

A disadvantage of variance-based sensitivity analysis methods, including Sobol, is the assumption of a unimodal distribution of the output. After all, by definition, the variance is always considered around some mean value. Therefore, it is essential to keep in mind that we have to check the distribution of the output and determine whether the output is unimodal or multi-modal distributed (Ghanem et al., 2017; Saltelli et al., 2008; Sobol, 2001).

3.1.4 Morris

The Morris method, proposed by Morris (1991) and later improved by Campolongo, Cariboni, & Saltelli (2007) is a form of variance-based sensitivity analysis as well but differs from Sobol as it relies on OAT sampling. The method is used for factor screening, so it can be used to identify important factors in the high-dimensional input space and determine which factors are negligible, linear and additive, or are non-linear or have strong interactions with other input variables. Morris can deal with computationally expensive models as well, since the amount of needed samples is relatively low in comparison with other variance-based GSA methods (Campolongo et al., 2007; Morris, 1991; Saltelli et al., 2008, 2004).

The method determines for each input variable several local sensitivity measures, which are referred to as Elementary Effects (EE). From this EE, the sensitivity measures can be calculated, which consist of the mean (μ), the estimate of the mean of the distribution of the absolute values (μ^*), and the variance (σ). The mean is the overall influence of an input variable on the output. The variance is an estimation of the non-linear effect or interaction effects with other input variables. The estimate of the mean of the distribution of the absolute values is later added by Campolongo et al. (2007) and gives extra information on the sign of the relation between the input variable and the output. If this sign switches depending on the sampling value, then the μ^* would indicate that this variable has a strong influence, while at the calculation of μ the different signs would cancel each other out and thus indicates a low influence (Campolongo et al., 2007; Morris, 1991; Saltelli et al., 2008, 2004).

Since the Morris method can deal with interaction effects between parameters, needs a considerably lower amount of samples, is model-independent, and results can be interpreted easily, the method is quite popular. The drawback of the method is that, although interaction effects and non-linearities can be recognised, there is no information on the nature or source of these interactions. These properties render the method especially suitable for factor fixing (Cariboni, Gatelli, Liska, & Saltelli, 2007).

3.1.5 RBD-FAST

The RBD-FAST method is a Random Balanced Design (RBD) method based on the Fourier amplitude sensitivity test (FAST).

FAST is a variance-based GSA method proposed by Cukier, Fortuin, Shuler, Petschek, & Schaibly (1973) and later improved by Saltelli, Tarantola, & Chan (1999). It indicates the total effect of one input variable on the output and includes both its primary and interaction effects. The classic FAST method uses a space-filling curve, based on a periodic function, with a different frequency for each input variable. Random points on this curve are selected, and the model is executed repeatedly for these input values. Subsequently, the Fourier spectrum is calculated to estimate the sensitivity index of an input variable (Cukier et al., 1973; Saltelli et al., 2008, 1999).

RBD-FAST is a method introduced by Tarantola, Gatelli, & Mara (2006). A certain amount of points is selected over a curve with frequency 1, leading to a coverage of only a limited subspace of the total input space. Subsequently, independent random permutations are applied to the selected points to generate random input values and get a space-filling design. Then the Fourier spectrum is calculated on the model output at frequency 1, from which the sensitivity indices are calculated (Saltelli et al., 2008; Tarantola et al., 2006).

The benefits of the combined method are that it is model-independent, and a relatively low number of samples is needed. It is specifically a suitable method for computationally expensive models. With RBD-FAST, only the first-order effect can be calculated, so the interaction effects cannot be derived. If the explained variance is far below 1, another method considering the interaction effects may be more appropriate (Saltelli et al., 2008; Tarantola et al., 2006).

3.1.6 PAWN

The PAWN method is a form of density-based sensitivity analysis, introduced by Pianosi & Wagener (2015) and later revised in Pianosi & Wagener (2018). Density-based means that the method investigates the whole output distribution, rather than just the variance. This kind of method is also referred to as "moment-independent", as the analysis does not only focus on one moment (the mean) but considers the whole shape of the Probability Density Function (PDF). This is especially useful if the output distribution is not centred around just one mean (if the outcome distribution is multi-modal) or the output is highly skewed. The sensitivity measure derived from density-based methods, in general, is based on the distance between the unconditional probability distribution and the conditional probability distribution. The unconditional probability distribution is derived by sampling the whole input space, while varying all the input variables at the same time. The conditional probability distribution is derived by sampling the whole input space while keeping only one variable at a constant level (Khorashadi Zadeh et al., 2017; Pianosi et al., 2016; Pianosi & Wagener, 2015, 2018).

The PAWN-method is based on the Cumulative Distribution Function (CDF) rather than the PDF. According to Pianosi & Wagener (2015), this is because the CDF is much easier to approximate from sampled data. The distance between the unconditional and conditional CDF is expressed by using the Kolmogorov-Smirnov statistic, which takes the maximum absolute distance between two CDF's. This statistic gives a value between 0 and 1, with 0 meaning that this input value does not influence the output and 1 meaning that the parameter has a strong influence on the output. Although the first introduction of PAWN uses a tailored sampling method, the method can be used at datasets generated by generic sampling methods as well. The sample data is then divided in a subset for calculating the unconditional CDF, while using the remaining sample data to fix values of the various input parameters to a certain value or value interval (Khorashadi Zadeh et al., 2017; Pianosi et al., 2016; Pianosi & Wagener, 2015, 2018).

PAWN can be used for factor fixing, factor prioritisation, and factor mapping. The main benefit of the PAWN method is that the method is moment-independent, implying it can deal with a highly skewed or multi-modal output. The algorithm is further model-independent and works with a relatively low number of samples. With the method being able to work with generic sampling methods, the method is computationally cheap as well (Khorashadi Zadeh et al., 2017; Pianosi et al., 2016; Pianosi & Wagener, 2015, 2018).

3.1.7 Random Decision Forests

The random decision forest method is a statistical machine-learning approach, based on a large number of decision trees. First, we dive deeper into the concept of decision trees. Afterwards, we describe the random decision forest method.

A decision tree is a graphical representation, in the form of a sort of flowchart, which goes from observations on attributes of an item to a prediction of the target value of this item. The tree consists of a root, branches, decision points (nodes), and leaves (terminal nodes). In the starting node (the root), a first binary splitting rule is made to split the observations in two and from this root, two branches emanate. In these branches, the succeeding node will again split the observations. This process is repeated until the stopping condition is fulfilled. This stopping condition can be either the depth of the tree (the maximum number of successive nodes) or the maximum number of samples to be found in a leave. These stopping conditions can be varied, to find the right balance between overfitting and variance (James, Witten, Hastie, & Tibshirani, 2013; Jaxa-Rozen & Kwakkel, 2018).

A decision tree is used for either classification, where an item is classified as a discrete value or category, or regression, where the target value of an item is estimated by the mean of the samples in this node. These trees are respectively called classification trees or regression trees. The splitting value of a node is determined by an optimisation procedure, where the value of a variable is chosen that will lead to the greatest reduction in the impurity of the underlying ('child') nodes. Impurity describes how well the node can estimate the observed value. If 100% of the data points are classified correctly, we consider the nodes 'pure'. As long as it doesn't, we consider the nodes 'impure'. The impurity can be calculated using, for instance, the 'Gini' measure for a classification tree, or the mean square error for a regression tree (James et al., 2013; Jaxa-Rozen & Kwakkel, 2018).

The main benefits of decision trees are that the method can deal with heterogeneous input variables and interactions between them, the trees are quite easy to interpret, and can deal with non-linear models. The main downsides are that decision trees have a high level of variance, causing the trees to be very sensitive for small changes in the input data and have a high chance of overfitting the data. Therefore, decision trees are often used in ensemble methods, where the different trees are combined in so-called forests (Jaxa-Rozen & Kwakkel, 2018).

There are multiple ways to combine decision trees in a forest, one of which is the random forest. This is a type of bagged (or bootstrap aggregated) decision trees. Bagging is used to reduce variance and means that multiple training sets are constructed by bootstrapping the sampled data. The different trees are built by resampling the data with replacement. With a random forest, the individual trees are randomised by selecting a subset of the input variables on which the tree will be trained (Breiman, 2001). The predictions of the regression trees are then averaged, and the predictions of classification trees are taken as a majority vote (Breiman, 2001; James et al., 2013; Jaxa-Rozen & Kwakkel, 2018; Tin Kam Ho, 1995).

Another ensemble method is the use of extremely randomised trees (Extra-Trees), as introduced by Geurts, Ernst, & Wehenkel (2006). In this method, an extra form of randomisation is used by randomising cutting points for each node as well as the randomisation of the subset of input variables.

Apart from classification and prediction, random forests and Extra-Trees are widely used for analysing the importance of variables, for variable selection, and detection of outliers (Verikas, Gelzinis, & Bacauskiene, 2011). Most of the advantages of a single decision tree apply to the ensemble methods as well: the forests can deal with heterogeneous input variables and interactions between them and can deal with non-linearities. However, the interpretability of the forests is less than the interpretability of a single tree, since the result of the algorithm cannot be graphically displayed in one intuitive tree anymore. The additional benefits of forest methods are that the growing of forests and the analysis are performed quite fast, they can deal with common sampling methods, and perform well on relatively small datasets with high dimensionality.

3.1.8 Patient Rule Induction Method

The Patient Rule Induction Method (PRIM) was proposed by Friedman & Fisher (1999). The method aims to associate certain uncertainties with a subset of the output space, which is called 'factor mapping'. This subset can, for instance, be outcomes that are of particular interest, because they fall under or above a determined threshold value or fall into a certain quantile of the data. If we are interested in these specific points of the outcome, we might be interested in what uncertainties cause these kinds of outcomes as well, and especially under what input values. That is where scenario discovery comes in, where PRIM is an example of (Kwakkel, Auping, & Pruyt, 2013).

The method starts by repeatedly executing the model, sampling the input variables over the whole uncertainty space. The resulting dataset is then classified based on the value of the outcome of interest. The PRIM algorithm will create hyperrectangular "boxes" to describe the samples in the output space that violate the determined threshold value. In general, these boxes contain both samples that are of interest as samples that are not of interest. The goal is to select a box that contains as many samples of interest as possible, while also containing as least as possible samples that are not of interest. This trade-off is expressed in the use of two descriptive values: coverage and density. A high coverage indicates a high proportion of the total number of outcomes of interest within the subspace. High density indicates a high percentage of outcomes in the subspace that are of interest (Abu-Hanna, Nannings, Dongelmans, & Hasman, 2010; Bryant & Lempert, 2010; Friedman & Fisher, 1999; Kwakkel & Jaxa-Rozen, 2016).

The algorithm starts with a first box that contains all the data points of the experiments. The next step is to peel off a bit of this box, by discarding a small slice of the data. The amount of data points that will be peeled off in each iteration is dependent on the chosen 'peeling alpha'. Since the dimensions can be continuous, discrete or categorical, this step depends on the data type of the dimension. For a continuous dimension, the algorithm will consider peeling off a small slice from the top or the bottom. In the case of a discrete dimension, the method will consider the highest or lowest values. For categorical data types, the algorithm will just try to remove one of the categories. After considering all the slices of the different dimensions, the algorithm will choose the best one to remove from the dataset. Which one is 'the best', is determined by an objective function and depends on the input variables. One way to determine which is the best slice to remove is to consider the increase in the mean of the output value divided by the number of data points that are removed. The last step of the method is up to the analyst. Based on the 'peeling-trajectory', the analyst assesses the earlier mentioned trade-off between coverage and density. Subsequently, based on the chosen box, we deduce which uncertainties and what values ensure the outcomes in the selected box (Abu-Hanna et al., 2010; Bryant & Lempert, 2010; Friedman & Fisher, 1999; Kwakkel & Jaxa-Rozen, 2016).

3.2 Calibration

Calibration methods come from an understanding that multiple parameter sets may result in comparable outcomes and are able to match calibration data or a statistical model. This understanding is referred to as equifinality (Beven & Freer, 2001). These methods are mainly used to calibrate models on observed data. It is also possible to use the methods to fit models to specific outcomes, so they can also be used for scenario discovery purposes. Within calibration methods, we distinguish two main approaches, regarding the sampling of parameter values: forward modelling based on independent sampling, and inverse modelling based on dependent sampling. Forward modelling uses model parameter value sets which are specified a priori. These values may be adjusted manually to better match observations or calibration data. Inverse modelling attempts to start with the observations or calibration data and infer the suitable parameter values. It can be seen as a form of guided search through the model space for a range of plausible values for each unknown parameter (Vrugt, 2016; Vrugt, ter Braak, Gupta, & Robinson, 2009).

3.2.1 Generalized Likelihood Uncertainty Estimation

The Generalized Likelihood Uncertainty Estimation (GLUE) method, introduced by Beven & Binley (1992) makes use of forward modelling. It relies on a high amount of sampled model runs with varying input values. Based on the comparison between observed and sampled data, an estimation is given for how likely each sample resembles the observed data. This estimation is reflected in the likelihood measure. There exist multiple ways to determine this likelihood, one of which is the inverse error variance, which calculates the closeness between the model outcomes and the calibration data. For all the sampled data points with their corresponding likelihood, the samples are divided into 'behavioural' and non-behavioural parameters sets. Behavioural parameters sets indicate that these sets are a likely combination of parameter values. This division is based on a determined cut-off value, which is defined as an acceptable deviation or a percentage of the total number of model runs. Subsequently, these behavioural parameters sets are normalised to construct the cumulative distribution function of the model output. Based on this distribution, we calculate the median and associated uncertainty (Beven & Binley, 1992; Ratto, Tarantola, & Saltelli, 2001; Vrugt, 2016; Vrugt et al., 2009).

GLUE is especially useful to investigate to what extent the model is able to replicate observed behaviour. The benefits of using GLUE are that the method is model-independent and quite intuitive. It can deal with different classes of parameters, allowing for analysis of categorical or structural uncertainties as well. If the method is performed on distributed computer systems, the method can be performed quite fast. However, the drawback is that quite a large number of model runs might be needed to find parameter values that resemble the calibration data. Even if sample methods like Latin Hypercube Sampling are used, this might result in only a few behavioural runs (Beven & Binley, 1992; Ratto et al., 2001; Vrugt, 2016; Vrugt et al., 2009).

3.2.2 Markov Chain Monte Carlo methods

Markov Chain Monte Carlo (MCMC or MC^2) is an umbrella term for lots of different algorithms and approaches. MC^2 approaches make use of inverse modelling and are based on Bayesian statistics, estimating the posterior distribution of input variables and output. They make use of an efficient algorithm to sample in the uncertainty space, using a combination of Monte Carlo simulations and Markov Chains.

Monte Carlo simulations are simulations which are executed using random input values within a determined domain for each input parameter, following some sort of probability distribution. By sampling using a Monte Carlo method, the values with a higher probability are sampled more often than values with a lower probability. Markov Chains are connected, sequential points, which are probabilistically related. The next point in the chain is dependent on the current point and the probabilistic distribution of the possible next points. This implies that the generation of a Markov Chain is not dependent on the history of the chain: a Markov Chain is memoryless.

MC^2 methods use the randomness of Monte Carlo simulations and the generation of samples following the Markov Chains logic. Generally, the method will choose a random value for the parameter as an initial sample. The next sample is chosen randomly, based on the current sample and a random sized step in an arbitrary direction. We consider this next sample a proposed successive value. This proposed point is compared to the current point in the chain. If this new point is more likely to explain the data, this new point is accepted. Otherwise, the proposed point is either rejected or accepted, based on the generation of a random value and a predetermined threshold. If this value is lower than the determined threshold, the proposed point is accepted. Otherwise, this point is rejected. MCMC methods assure that the regions of high probability are sampled more than other regions. Because the first points of the chain are heavily dependent on the arbitrary chosen initial point, these are often thrown away. This approach is referred to as the 'burn-in' time. By removing these samples, the chain is no longer dependent on this starting point.

MCMC aims to approximate parameter values which explain observed data. Based on the accepted samples, the posterior output distribution can be approximated as well. Therefore, Markov Chain Monte Carlo methods can be used to calibrate the input parameters to fit some observed data. Based on the posterior distribution of the parameters, one can derive the relative influence of this parameter on the output of the model.

A variety of methods is developed to improve the efficiency of MC^2 methods further. The approaches can roughly be divided into single- and multi-chain methods. Single-chain methods consist of one Markov Chain, which searches through model space. Although this approach works quite well for Gaussian-shaped output distributions, it might not converge well for target distributions with long tails or multi-modal outcome distributions. Single-chain methods have difficulties in dealing with different regions of attraction and many local optima, as they might get stuck in a single region without exploring the rest of the input space. Multi-chain methods are MCMC approaches where multiple Markov Chains are deployed simultaneously to search the model space. This approach ensures that it is able to converge more efficiently, even if the target distribution is long-tailed, multi-modal, and has many local optima (Vrugt, 2016).

One of the best known MCMC algorithms is the Metropolis algorithm. This algorithm starts with a random initial value in the uncertainty space, for which the model is evaluated. The next jump size is taken from a jump distribution, centred on the current value. Most commonly, this is a normal distribution with a variance specified by the analyst. The model is evaluated again at this point, and the jump is accepted or refused based on the likelihood of the next point, divided by the current point. If this ratio is bigger or equal to one, the point is accepted. Otherwise, the point is accepted based on the ratio and the draw of a random value between zero and one. If the ratio is smaller or equal to this draw, then the next step is accepted. Otherwise, the current value is used again to evaluate the model. Subsequently, the algorithm is repeated with the proposal of a next point.

Another well-known algorithm, similar to the Metropolis algorithm, is the Metropolis-Hastings algorithm. The two algorithms are very much alike, although they differ in the distribution of the jump function. Where the Metropolis algorithm requires a symmetric distribution, the Metropolis-Hastings algorithm allows the use of an asymmetric jump distribution. This is particularly useful when we are interested in higher sample values (for which we would use a skewed distribution) or when there are particular constraints on the parameter values.

An example of a multi-chain algorithm is the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm, introduced by Vrugt, ter Braak, Clark, Hyman, & Robinson (2008). It has been developed to search for parameter sets that fit observed data efficiently. The chains use multiple starting points, so the method can better deal with multiple regions which may be of high interest and can better assess whether the algorithm has been converged. Since the different chains exchange information, the method removes chains that are stuck in spaces of the parameter space which do not contribute to the discovery of high probability areas. For the proposal function, used to determine the next point to consider, the scale and orientation are dynamically being updated during the analysis. The subset of dimensions to consider is also varied each step, as with high-dimensional input space it is often not optimal to consider all dimensions (Vrugt, 2016; Vrugt et al., 2008, 2009).

3.3 Comparison of Methods

In this chapter, we discuss and compare the different methods for analysing uncertainty. Which method to use is mainly dependent on the model, the goal of the analyst, and the characteristics and number of uncertainties.

Global Sensitivity Analysis methods Sobol, Morris, FAST, PAWN, Random Decision Forests, and PRIM all use an uncertainty space-filling design to determine the influence of uncertainties on the model outcomes. Their differences lie in the used sampling method, purpose, and the required number of model evaluations. The assumed output distribution is different as well, as most methodologies are variance-based, assuming the mean and variance are adequate indicators for the characterization of uncertainty. PAWN and RDF are able to deal with multimodal outcomes. The calibration methods GLUE and DREAM are both designed to fit the output of a model to observed calibration data. GLUE does this by randomly sampling the uncertainty space using a Monte Carlo approach; DREAM uses a Markov Chain Monte Carlo approach, actively searching for high probability regions, deploying multiple chains.

We typify the different methods and approaches by using a comparative framework provided in Table 4. The first attribute we consider is the intended purpose of the method: whether it is meant for factor prioritization, factor fixing, or factor mapping, or calibration (CB). The next attribute is the assumed output distribution. Another aspect is whether the method can deal with integer (INT), categorical (CAT), continuous (CNT) uncertainties, and interaction between them. The sampling method is another property of the methods. If the methods are able to perform their analysis on the base of standard sampling methods as Monte Carlo or Latin Hypercube Sampling, the same dataset can be used to perform multiple analysis, which is beneficial for computationally expensive models. The sampling methods differ as well in the use of OAT or AAT sampling, although only the screening method of Morris uses an OAT-approach. Also, we indicate if a method is available as a Python package or as part of the EMA-workbench. The last attribute is the number of samples which is needed to run the analysis reliably. The sample size generally depends on the number of uncertainties (M) that is included in the analysis and the complexity of the model output.

Table 4 Framework for comparing methods for uncertainty analysis. The methods differ in intended purpose, assumed output distribution, type of uncertainties, and sampling methods and size.

	Purpose				Assumed output distribution	Uncertainty types			Uncertainty interaction	Sampling	Available in Python	Sample size
	SA			CB								
	FP	FF	FM			INT	CAT	CNT				
Sobol	x	x			Unimodal, symmetric	x	x	x	Can deal with interactions	AAT, Sobol	Yes, and available in EMA-workbench	1000 x (M + 1)
Sobol (incl. 2nd order)	x	x			Unimodal, symmetric	x	x	x	Can deal with interactions and calculates 2nd order interaction	AAT, Sobol	Yes, and available in EMA-workbench	2 x 1000 x (M + 1)
Morris		x			Unimodal, symmetric			x	Can deal with interactions	Morris, OAT	Yes, and available in EMA-workbench	10 - 100 x M
FAST	x	x			Unimodal, symmetric	x	x	x	Indices only indicate the total effect	AAT, no specific method	Yes, and available in EMA-workbench	1000 x M
PAWN	x	x	x		Multimodal, skewed	x	x	x	Can deal with interactions	AAT, no specific method	Not yet implemented in Python	1000 x M
RDF	x	x	x		Multimodal, skewed	x	x	x	Can deal with interactions	AAT, no specific method	Yes, and available in EMA-workbench	100 x M
PRIM			x		Unimodal, symmetric	x	x	x	Can deal with interactions	AAT, no specific method	Yes, and available in EMA-workbench	100 x M
GLUE				x	Multimodal, skewed	X	x	x	Can deal with interactions	AAT, no specific method	Yes	1000 x M
DREAM				x	Multimodal, skewed	x	x	x	Can deal with interactions	AAT, MCMC	Yes, e.g. through PyDREAM package	1000 x M

4.

METHODOLOGIES AND MULTI-MODEL ECOLOGIES

Based on the provided additional uncertainties in multi-models and the available methods, what can we say about uncertainty analysis on multi-model ecologies in general? How can these methods be applied to analyse a whole multi-model ecology?

This answer depends largely on the composition and structure of the multi-model ecology: what and how many different models are included, how and to what extent do the models interact, what different types of simulation models are involved, is there an overlap between the used parameters and input data? The variety of possible configurations of multi-model ecologies render it nearly impossible to give a conclusive answer on this question that can be applied to all of these configurations.

4.1 Multi-model Archetypes

Therefore, we propose two archetypes of two coupled simulation models. These archetypes can be considered as building blocks, whereof numerous different multi-model configurations can be composed. We consider models as nodes in a network structure or graph. The interfaces form the edges of the graph. The differentiation lies in the extent to which feedback mechanisms are present and whether or not the graph is directed between the nodes (models). We limit ourselves to the conceptual level of the multi-model ecology. Many multi-models will use a combination of these archetypes. The archetypes can be considered as two extreme forms of model interaction:

- Multi-model structure as a directed graph
- Multi-model structure as an undirected graph and with feedback mechanisms

4.1.1 Directed graph

In this kind of configurations, visualized in Figure 3, there is no run-time interaction between the two separate computer models. The multi-model is a kind of pipeline implementation. One computer model performs its calculations based on the input data, provides its output and in the next step, the next computer model uses this output, combined with other input data, and provides the output data. Substantive interaction between the models is limited.

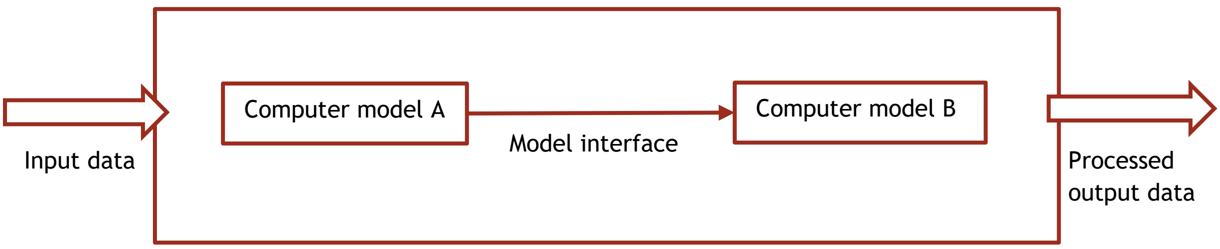


Figure 3 **Multi-model structure as a directed graph.**

4.1.2 Undirected graph

In this archetype, visualized in Figure 4, there is bidirectional run-time interaction between the models. Feedback loops run over the model components. These are related to the changing context and path dependencies that may emerge in the models.

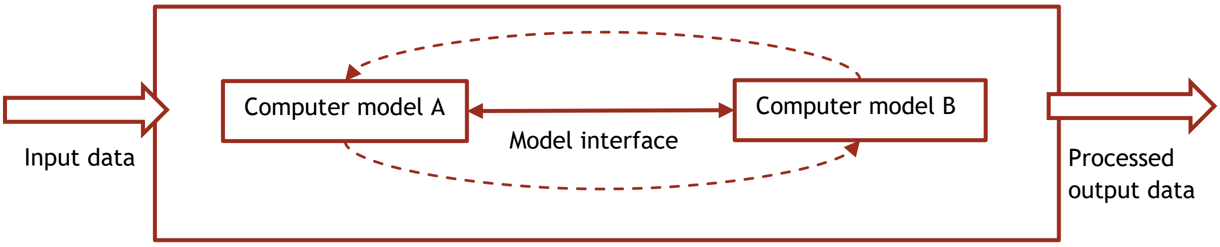


Figure 4 Multi-model structure as an undirected graph and with feedback mechanisms

4.2 Uncertainty Analysis on Different Multi-model Archetypes

4.2.1 Directed graph

There are no additional uncertainties compared to a single simulation model. It is sufficient to perform a sensitivity analysis on the first model in the chain and to capture the range of the different model outcomes in a probability density function. This PDF can then be used on the input parameter(s) to the second model, which are affected by the first model. A sensitivity analysis can be performed on the second model, taking into account the full posterior probability density function of the first model.

A similar approach is possible for calibration. In this case, it might be better to reverse the order of analysis: first, calibration could be applied to the second (or last) model in the pipeline. Next, it is clear within which outcome space the first model should come out. Based on this analysis, the first model can be calibrated to estimate the parameter values under which this sub-output space would be reached.

Depending on the multi-model, an alternative might be to perform both sensitivity analysis and calibration on the multi-model as a whole. Because there are no interaction effects between the models, it is expected that in terms of results, it does not matter whether we apply the methods to the individual models or the whole multi-model. In terms of computational costs, it should not matter in most cases, either. However, if the dimensionality of the uncertainty space in a multi-model becomes very large, some methods may require substantially more samples. When the first model affects several parameters of the second model, and there is a complex relationship between them, it might be better to apply the methods to the multi-model as a whole. Otherwise, to reduce the uncertainty dimensions, it would be better to perform the analysis onto the two separate models.

4.2.2 Undirected graph

Run-time interactions result in accumulation of uncertainties on the interface. The input of both models is needed to give sensible results. Because of the interwoven, two-sided dependency in the different model steps, only sensible parameter values can be provided using the logic of the underlying simulation models. It does not make much sense here to perform an uncertainty analysis on the individual models.

To provide insight into the sensitivity of the exchanged parameters on the multi-model outcomes, we propose to include the parameters that are exchanged between the different model in the uncertainty analysis. These are the values that are used to perform the next modelling step in the other model. One could think of setting a bandwidth around different (continuous and discrete) exchange parameters. For each data exchange, we can examine whether this exchange can be adjusted slightly by influencing the values. Other possibilities for manipulation could be, for example, the order of delivery of information, influencing the used descriptive statistics, or the (numerical) rounding of numbers. In this way, we can analyse to what extent the multi-model results depend on different ways of delivering information or small variations in the exchanged values.

Concerning calibration, a similar approach is possible. By performing calibration on the multi-model as a whole, we include all relevant parameters and processes. This does not mean that the parameter values found can also be used one-to-one for the individual models. However, we state that the sub-models cannot be used separately in a sensible way, or at least not on the defined multi-model outcomes.

4.3 Experimental Setup

In this section, we propose an experimental setup for multi-models with undirected graphs, which we then apply to the Windmaster model. We consider the Windmaster model a multi-model with undirected graphs and feedback loops over the different sub-models. First, we assess the uncertainties in the multi-model using the XPIROV framework and uncertainty matrix. Afterwards, we apply a selection of the methods introduced. To perform an uncertainty analysis on the model, we use two sensitivity analysis methods with independent sampling techniques and one calibration method with dependent sampling.

4.3.1 Factor fixing

The first step is the generation of plausible scenarios, using different combinations of values of the uncertainties. Based on these experiments, we get an overview of possible model outcomes. On these outcomes, we apply a factor fixing approach.

Sampling method

Since the uncertainty space is high-dimensional, a Sobol sequence-based sampling method becomes too computationally expensive. This leaves us with the choice for either Monte Carlo sampling or LHS. Although a better distribution may be reached using Latin Hypercube Sampling, it is not possible to add more samples afterwards and investigate the convergence of feature scores over the increasing number of used samples. To be able to gain insight into this convergence, we choose MC sampling.

Sensitivity analysis method

For the selection of the method, we must take into account some requirements:

- A limited number of required model runs; since we need to incorporate many uncertainties, we aim for a method that can calculate its sensitivity indices based on a relatively limited number of model evaluations.
- Python implementation; there is an implementation of the method available as Python package or, preferably, as part of the EMA-workbench.
- Uncertainties; the method can handle categorical, integer, and real parameters.
- Assumed output distribution; the method does not impose any specific requirements on the shape of the output values since we do not know what the output distribution for the KPI's will be.
- Sampling; the method can perform analysis on MC drawn experiments.
- Interaction effects; the method needs to cope with interaction effects between input parameters.

Based on the comparison framework in Table 4, we choose a Random Decision Forest feature scoring, based on the extremely randomized trees method.

Experimental design

We use the EMA workbench developed by Kwakkel (2017), as the Windmaster model already includes an implementation of the workbench. The EMA workbench feeds different combinations of uncertainties and policies to the Windmaster model and retrieves and structures the used model inputs and the model outcomes. In addition, the EMA workbench carries out the computation of the experiments in parallel. We make use of this by performing the experiments on a cloud-based High-Performance Cluster.

To limit the number of required model evaluations, we distinguish deep uncertainties we want to explore in a broad, global context and stochastic uncertainties which we want to explore in a more limited, computationally cheaper, context. Deep uncertainties will be included in the generation of experiments, where stochastic uncertainties will be included in the replications of the model. Replications are a set of specified uncertainty values, for which each experiment will be repeated. For each deep uncertainty, we may trace back its impact on the different outcomes of interest. We may analyse the impact of a specific replication of the model outcomes, but we cannot determine the impact of a specific parameter included in these replications, because we will not sample enough combinations of parameter values within this category and therefore would lead to unreliable estimates for the individual stochastic uncertainties. We include policy options as a deep uncertainty as well, to limit the required amount of model evaluations. If we were to include the policies in a partial factorial design, we would need more model evaluations.

Subsequently, we need to determine the number of experiments and replications, which together determine the required model evaluations. We include 11,000 experiments over ten different replications, what comes down to 110,000 model evaluations.

4.3.2 Global Sensitivity Analysis

We perform a Global Sensitivity Analysis to calculate first, second, and total order sensitivity indices for the remaining uncertainties. We use the results of the factor fixing method to determine what uncertainties to fix on an arbitrary value within its bounds. We also include the uncertainties relating to the interface, because it is expected that there will be interaction effects between the (intermediate) outcomes and the other uncertainties. Based on the comparison framework, we choose Sobol because this method is available as part of the EMA workbench, it can deal with interactions between uncertainties and provides insight into these interactions. A Sobol sequence sampling method is required.

Experimental design

The design is largely the same as the previous one. Since the required number of experiments is largely dependent on the amount of included deep uncertainties, we will require around 22,000 experiments. We will perform ten replications for each experiment. This amounts to 220,000 model evaluations for the Sobol sequence sampling.

4.3.3 MCMC method

For the dependent sampling calibration method, we choose an MCMC method. MCMC methods are not yet implemented in the EMA workbench, but there are several Python packages available that we can use on the model. Preference is given to a multi-chain approach, in which the chains can work in parallel to achieve convergence faster and reduce computational costs. We, therefore, choose the DREAM algorithm. Since there is no data available on which the model can be calibrated, we use the MCMC method for the purpose of factor mapping or scenario discovery. We deploy the DREAM method to search for scenarios in which an outcome of interest violates a specific threshold value. These outcomes and threshold values are based on the value systems of the stakeholders.

Experimental design

We use the Python package PyDREAM as developed by Shockley, Vrugt, & Lopez (2018) to sample experiments in the area of interest. We deploy three chains with different starting points and use a distance function for the likelihood. For outcomes under the threshold value, we define this function as the negative distance. For all outcomes higher than the threshold value, the function returns 1.

Since the PyDREAM package cannot cope with different types of parameters by itself, we sample all the parameters as if they were real numbers. In the model, we then convert the numbers to integers or categories by using a floor function. This mathematical function converts real numbers to the highest integer lower than the sampled number. Contrary to the common rounding function, the probability of any integer within the domain is equally high.

Because each chain runs on a single processor core, there is no added value in deploying more than three cores. This entails that we can make limited use of distributed computer systems. Consequently, the viable number of samples is somewhat limited. The advantage of dependent sampling is that the experiments can always continue later since the history of each chain is recorded. The chains then continue where they left off. We perform 500 iterations per chain. No replications will be executed in order to reduce the computational expense.

5.

Uncertainty Analysis of the Windmaster Model

5.1 Introduction to the Windmaster Model

Global warming, driven by increasingly higher CO_2 emissions, is one of the grand challenges of current times. Effects of global warming are already visible through melting glaciers, rising sea levels and long periods of drought. Since climate change is affecting everybody in all countries, the United Nations acknowledged global warming as part of their Sustainable Development Goals (SDGs) (United Nations, 2018). These SDGs also underline the need for affordable and clean energy. Over the last few decades, CO_2 emissions have increased significantly. Reduction of CO_2 emissions is paramount in order to reduce global warming and to preserve our planet for future generations. One hundred ninety-five countries signed the Paris agreement to limit global warming. As the Netherlands is also part of this agreement, CO_2 emissions need to be zero in 2050 (Nederlandse Emissieautoriteit, 2015).

Despite this agreement, the Netherlands is lagging on the reduction of CO_2 emission. The Port of Rotterdam emits approximately 15% of this emission. The Port aims to reduce CO_2 emissions by 50% in 2025 as compared to 1990, which entails a total reduction of 12 Megaton of CO_2. However, research shows that emissions are only increasing (Plomp, Wetzels, Seebregts & Kroon, 2013).

Electrification is seen as a solution with great potential to reduce CO_2 emissions in the Port of Rotterdam to decrease these emissions, especially within the industrial (petrochemical) complex. Electrification entails that these industries use electricity as energy source, instead of fossil fuels such as oil and gas. Although this solution has great potential, industries cannot adapt their processes and assets on an individual basis, because all industries are interconnected and therefore dependent on each other. Besides, the current energy grid in the Port of Rotterdam, consisting of infrastructure from network operators TenneT, Stedin, and Gasunie, cannot cope with an increased amount of demand and supply of energy. Therefore, investments in the energy infrastructure need to be made to support the energy transition in the Port of Rotterdam.

However, how this energy transition will take place is highly uncertain. Besides, the involved investment costs are high, and the lead times of investments tend to be long. Investments require collaboration and consistent investment policies between the different network operators to make sure the right circumstances are created for industries to switch to alternative energy technologies.

The Windmaster model is developed to discover robust policies regarding the investments in the infrastructure of the industrial cluster of the port of Rotterdam. The multi-model consists of three connected single-models with different modelling paradigms: an exploratory modelling scenario model, a technical-economical infrastructure model, and an investment behaviour model. Viewed on a conceptual level, these components of the Windmaster model generate different possible (equally likely) energy transition pathways, determine their impact on the electricity and gas networks, propose and consider investments, make a choice for an investment, update the energy network accordingly, and keep a track record regarding which events from the transition pathways could be supported and which could not (Wurth et al., 2019).

5.2 Uncertainties within the Windmaster Model

5.2.1 External factors

The Windmaster model is developed to explore uncertainties in the development of supply and demand of energy in the industrial cluster of the Port of Rotterdam. After all, the objective of network operators is to support the industrial cluster with their energy infrastructure. Since the development of new energy infrastructure and the expansion of the current infrastructure tend to take long and involve high investment costs, it is important to gain insight in the uncertainties involved in the energy transition in the Port of Rotterdam. Based on this exploration, insights can be obtained regarding robust adaptive investment strategies, which can support these different transition pathways.

Following the logic of the Windmaster model, we define transition pathways as a series of discrete events influencing the peak demand of energy of an asset, the required feedstock of an asset, or the realization of new energy production or conversion assets on a specific location.

To generate different plausible pathways, we need a variety of parameters. The Windmaster model has incorporated these parameters, and they will be introduced in this section. The industrial cluster in the Windmaster model consists of different assets, which are either supply, conversion, or final demand assets.

The supply side consists of the supply of feedstock for the assets in the cluster. They include the delivery of waste, coal, gas, oil, and electricity. Electricity comes from the Dutch electricity grid, the BridNet connection on the Maasvlakte, the offshore wind landing site on Maashaven, and the offshore wind landing site on Simonshaven. Uncertainties in the supply side of the cluster include the growth of wind energy landing on Simonshaven and Maashaven from 2030 and onwards. In addition to wind energy, the production and import of hydrogen gas are uncertain as well. Hydrogen gas could be either imported, produced at existing SMR locations, or produced at a new hydrogen gas production site at the Maasvlakte or the Botlek area. Production sites of hydrogen gas in the industrial cluster are included as conversion asset.

Conversion assets are assets where feedstocks are used and (partially) converted to one or more other feedstocks. They include oil refineries, energy plants, and water electrolysers. At these assets, different types of energy can be produced, such as steam, heat, electricity, and syngas. The assets can be divided into the following categories:

- Boilers; production of steam
- Furnaces; production of heat
- Cogen plants; combined production of heat and electricity
- Steam Methane Reformers (SMR's); production of hydrogen gas
- Coal plants; production of electricity
- Oil refineries; production of oil-based fuels

All these assets have different kind of (future) technologies or feedstocks to switch to, as part of the energy transition. These transition alternatives are laid out in Figure 5.

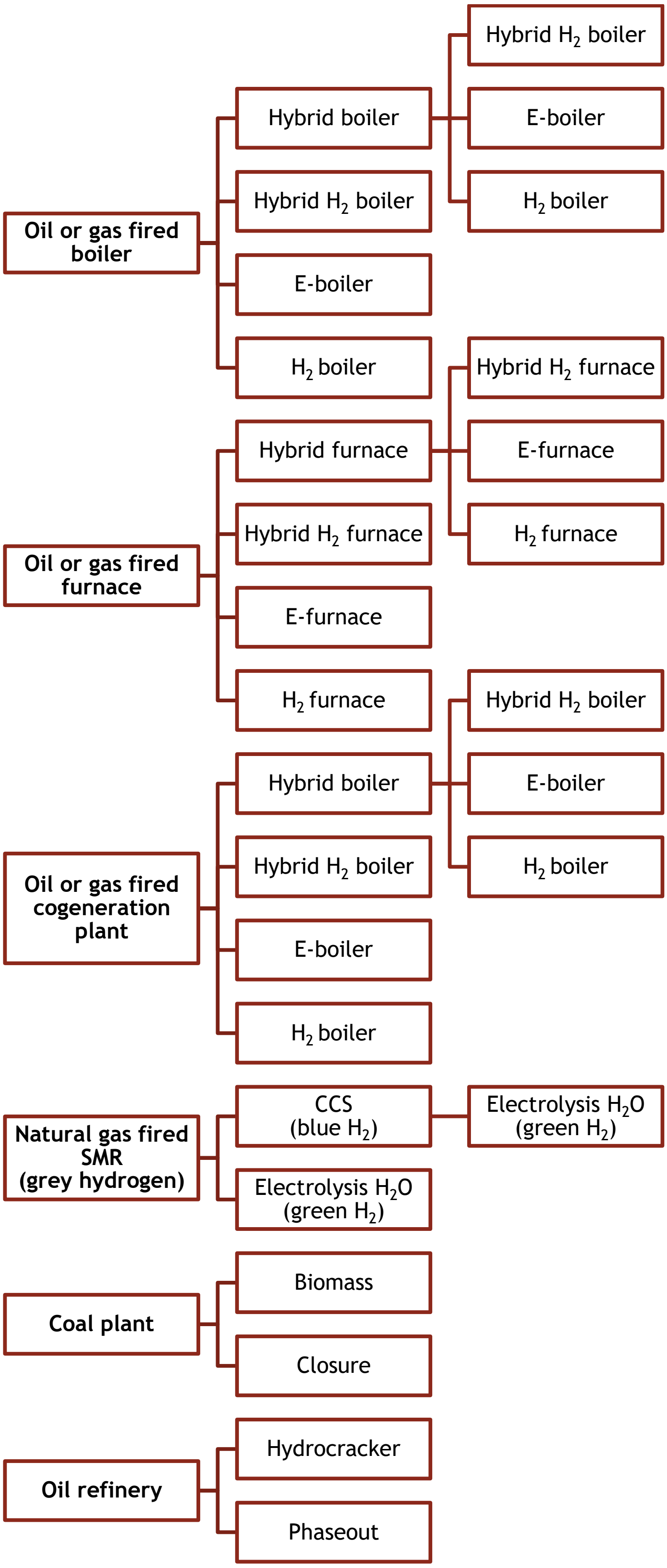


Figure 5 Transition alternatives per conversion asset category. Boilers, furnaces, and cogeneration plants switch to other feedstocks or a combination using a hybrid implementation. SMR's install a Carbon Capture and Storage (CCS) facility, switch to electrolysis of H_2O, or both. Coal plants either switch to biomass feedstock or close. The oil refineries either install a hydrocracker or phase out.

Each alternative has an impact on the amount and type of both required and produced feedstock and emission of CO_2. The choice for specific technologies in itself is uncertain, but also the timing of implementation and the time of availability of these technologies are uncertain. The uncertainties included in the generation of transition pathways are:

- Year of availability of technology
  - Availability of CCS
  - Availability of electrolysis H_2O
- Year of implementation
  - Switch to biomass or closure of Uniper's coal plant
  - Switch to biomass or closure of Engie's coal plant
  - Lead time for implementation per conversion asset

The lead time for implementation per conversion asset is the time that passes before this asset indeed implements the chosen alternative. For example, if each oil or gas-fired boiler is switching to electricity from 2020 onwards, the actual implementation per asset may take place between 2020 and 2025.

On the demand side of the Windmaster model, we identify uncertainties in demand for hydrogen gas in the Hinterland and the demand for natural gas in the built environment of the port. The demand for natural gas in the municipality of Rotterdam determines when backbones of Gasunie's infrastructure could be used for transporting hydrogen gas. The demand for hydrogen gas in the Hinterland determines whether the port of Rotterdam will function as a transit port for H_2 and what volumes are required per year.

An overview of all the parameters that together control the generation of transition pathways is included in Table 5.

Table 5 Overview of parameters used in the Windmaster multi-model ecology to generate transition pathways. The parameters may be used to either influence the peak demand of energy of an asset, the required feedstock of an asset, or the realization of new energy production or conversion assets on a specific location. The variable name refers to the used parameter name in the Windmaster mode; description gives a short explanation of the controlled aspects of the parameter; a type can be either Boolean (bool), categorical (Cat), real, or integer (Int). Limits or categories indicate the possible parameter values.

	#	Variable name	Description	Type	Limits or categories
Supply	1	offshore wind growth	Growth of wind capacity between 2030 and 2050. If True, the supply to both TenneT's 380 kV stations at Simonshaven and Maasvlakte grow each year following an S-curve from 3 (in 2030) to 13 (MV) or 12 (SH) in 2050	Bool	True; False
Conversion	2	BOILER paths	Steam production from current (gas, oil, or hybrid) to alternative technology. Assets can change to Hybrid (a combination of current and e-boiler), Hybrid H_2 (H_2 and e-boiler), e-boiler, or H_2 boiler	Cat	Hybrid; HybridH_2; Electricity; H_2; Hybrid_HybridH_2; Hybrid_Electricity
	3	COGEN paths	Cogeneration (combined heat and power) from current (gas, oil, or hybrid) to alternative technology. Assets can change to Hybrid (a combination of current and e-boiler), Hybrid H_2 (H_2 and e-boiler), e-boiler, or H_2 boiler	Cat	Hybrid; HybridH_2; Electricity; H_2; Hybrid_HybridH_2; Hybrid_Electricity
	4	FURNACE paths	Heat production from current (gas, oil, or hybrid) to alternative technology. Assets can change to Hybrid (a combination of current and e-boiler), Hybrid H_2 (H_2 and e-furnace), e-furnace, or H_2 furnace	Cat	Hybrid; HybridH_2; Electricity; H_2; Hybrid_HybridH_2; Hybrid_Electricity;
	5	SMR paths	Steam Methane Reformers from current (with or without CCS) to water electrolyser or current (without CCS) equipped with CCS. Both these options are dependent on their year of technology introduction	Cat	SMR_CCS; Electrolysis_H2O; SMR_CCS_Electrolysis_H2O
	6	timing CCS	Year of the introduction of Carbon Capture and Storage technology (possible extension for SMR sites)	Int	2022 - 2030
	7	year of introduction ELECTROLYSIS_H20	Year of introduction hydrogen gas	Int	2028 - 2050
	8	location h2 production	Location of possible new H_2 production site (SMR)	Cat	None; Maasvlakte; Botlek
	9	Delta per conversion asset	For each conversion asset, a 'delta' value is included, which determines the lead time (in years) for implementation of a new technology	Int	0 - 5

Conversion	10	SF factory 2032 location	Location for a new production plant for synthetic fuels based on CO_2 obtained from the air in 2032	Cat	None; Maasvlakte; Botlek
	11	SF factory 2036 location	Location for a new production plant for synthetic fuels based on CO_2 obtained from the air in 2036	Cat	None; Maasvlakte; Botlek
	12	SF factory 2040 location	Location for a new production plant for synthetic fuels based on CO_2 obtained from the air in 2040	Cat	None; Maasvlakte; Botlek
	13	SF factory 2044 location	Location for a new production plant for synthetic fuels based on CO_2 obtained from the air in 2044	Cat	None; Maasvlakte; Botlek
	14	SF factory 2048 location	Location for a new production plant for synthetic fuels based on CO_2 obtained from the air in 2048	Cat	None; Maasvlakte; Botlek
	15	yearHydrocrackerBP	Hydrocracker BP	Bool	True; False
	16	baseload biomass or closure C3	Switch to biomass (True) or closure (False) of Uniper's coal plant	Bool	True; False
	17	C3 year	Year in which Uniper's coal plant switches to biomass or closes	Int	2020 - 2030
	18	baseload biomass or closure C6	Switch to biomass (True) or closure (False) of Engie's coal plant	Bool	True; False
	19	C6 year	Year in which the Engie's coal plant switches to biomass or closes	Int	2020 - 2030
	20	yearBPOffline	Year BP offline if BP has not invested in hydrocracker capacity	Int	2035 - 2040
	21	yearGunvorPhaseout	Year of Gunvor's oil refinery phaseout	Int	2025 - 2030

	22	yearKochPhaseout	Year of Koch's oil refinery phaseout	Int	2025 – 2030
Demand	23	easterly demand H2	Possible H_2 market development, following an S-curve from 0 in the year of H_2 introduction, to parameter value in 2050	Int	30 – 172
	24	year end demand gas Rotterdam	End year of natural gas demand from the built environment of Rotterdam and surroundings	Int	2030 - 2045

5.2.2 Policies

The policies in the Windmaster model focus on the investment decision making strategy of the individual network operators. There are four defined investment strategies: reactive, current, proactive, and collaborative. The alternatives influence various characteristics of the decision-making process:

- Time horizon; the time frame that network operators may consider in their decisions.
- Investment goal; preference for an investment alternative based on price and expected overcapacity.
- Investment budget; the specified budget for each network operator.
- The propensity to save; whether the focus of a network operator is on saving for future investments.
- Lead time per investment; the lead time for the implementation of an investment is dependent on whether a network operator tries to collaborate to shorten the lead time.

Table 6 Controlled aspects of decision-making process per network operator and decision-making policy. Stedin, Gasunie, and TenneT can apply a reactive, current, or proactive decision-making policy. The 'collaborative' network operator is a fictive combined operator which is responsible for all the energy infrastructure in the Port of Rotterdam industrial cluster.

	Investment aspect	Reactive	Current	Proactive
TenneT	Time horizon (years)	0	10	20
	Investment goal	Lowest overcapacity	Satisfy as many clients as possible	Highest overcapacity
	Investment budget (M €)	10	75	150
	Propensity to save	No	Yes	Yes
Gasunie	Time horizon (years)	0	10	20
	Investment goal	Lowest overcapacity	Satisfy as many clients as possible	Highest overcapacity
	Investment budget (M €)	5	25	150
	Propensity to save	No	Yes	Yes
Stedin	Time horizon (years)	0	10	10
	Investment goal	Lowest overcapacity	Satisfy as many clients as possible	Highest overcapacity
	Investment budget (M €)	3	5	8
	Propensity to save	No	Yes	Yes
Collaborative	Time horizon (years)	20		
	Investment goal	Create transport capacity as fast as possible		
	Investment budget (M €)	411		
	Propensity to save	Yes		

5.2.3 Internal factors and relationships

The internal factors and relationships are defined in the agent-based investment behaviour model and the technical-economical infrastructure model. This section will elaborate on the global operation of the multi-model and the interactions between the two models. For a full description of the Windmaster model and its assumptions, we refer to the model documentation and report by Wurth et al. (2019).

The Windmaster model consists of the two 'main' models, an exploratory modelling part, and different datasets used by both models to make sure the input data and assumptions are the same for both models. An overview of the multi-model ecology is given in Figure 6.

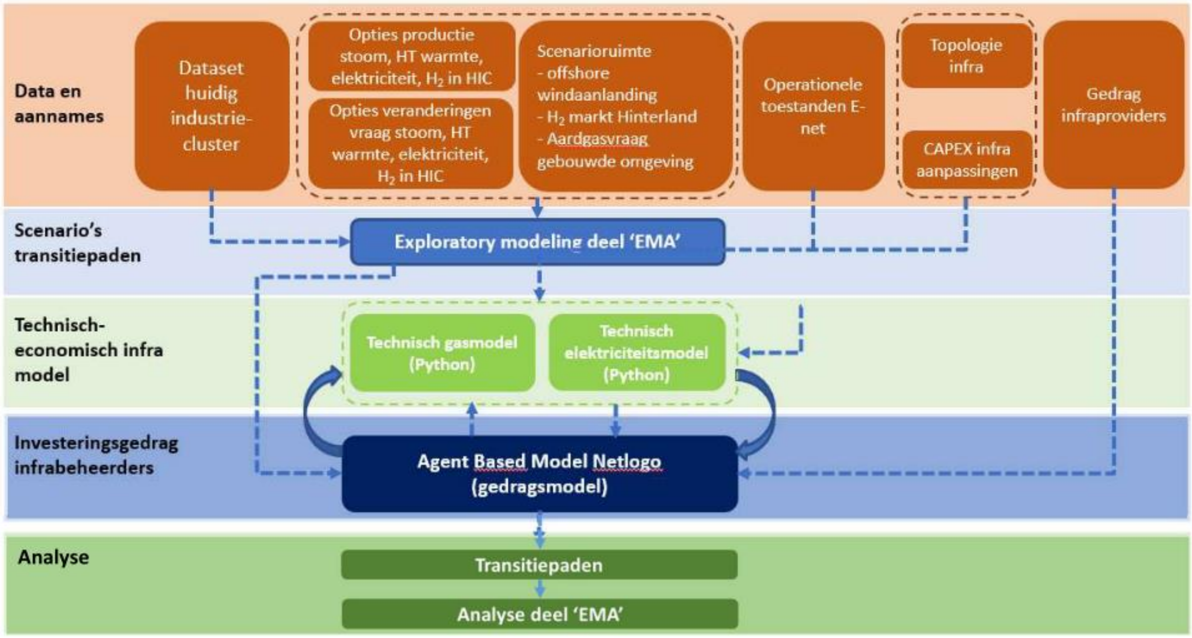
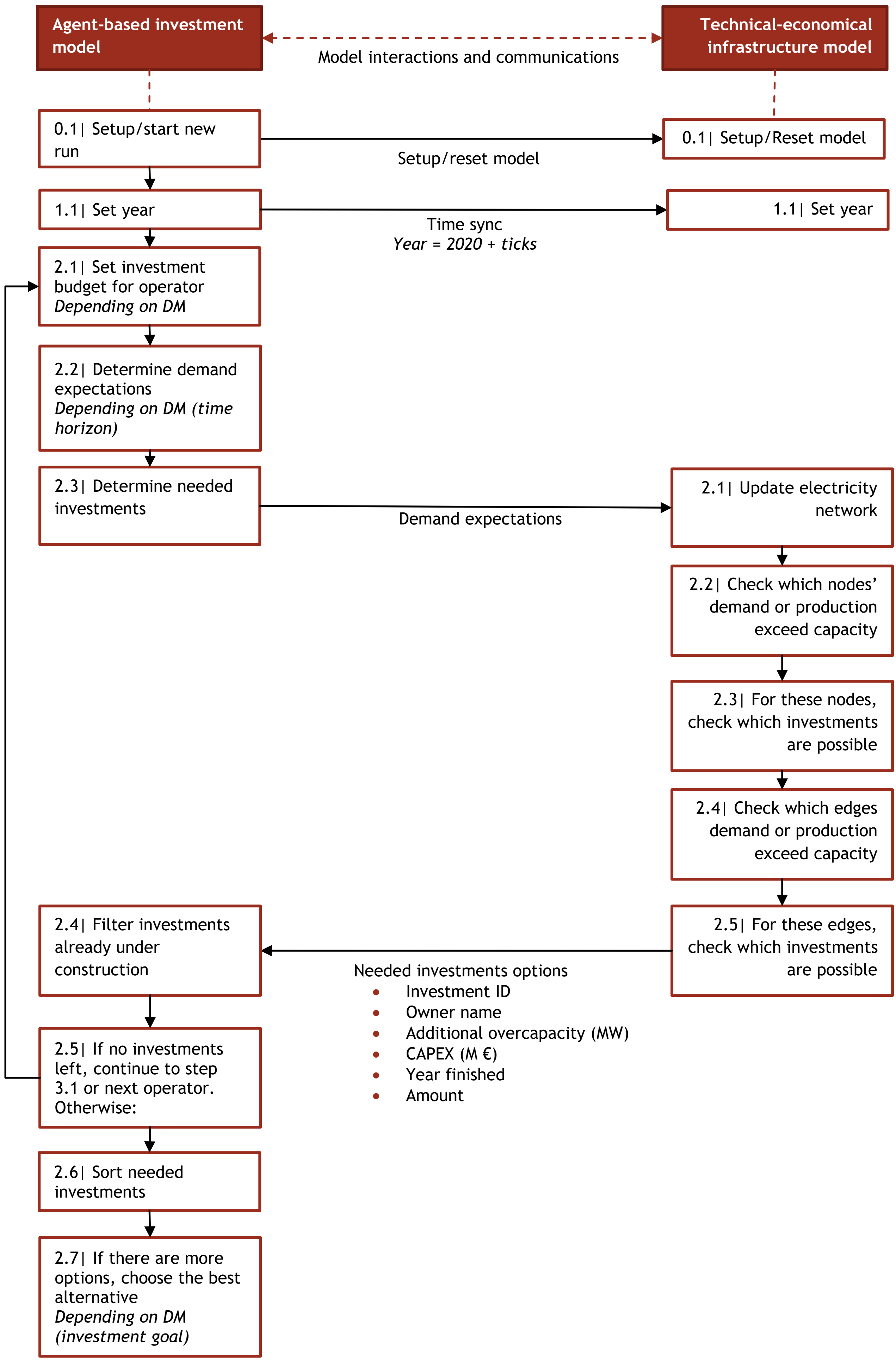


Figure 6 Schematic view of the Windmaster multi-model, as obtained from Wurth et al. (2019, p 14).

The technical-economical infrastructure model is developed in Python and is used to calculate the load flow over the energy infrastructure in the Port of Rotterdam, considering changes in one of the connected demand or supply assets in the Port. For these changes, the model determines if there are bottlenecks in the distribution networks, consisting of energy stations and distribution lines. If one of these components is expected to overload, the model generated a set of investment options to obviate these bottlenecks. The investment behaviour model is an agent-based model developed in Netlogo. Based on the information from the technical-economical model, it determines alternative investments.

Finally, the Exploratory Modelling and Analysis (EMA) part of the multi-model is used for generating different energy transition pathways in which the energy transition may take place. These pathways, consisting of certain 'events', may, for instance, be the switch of one industry to another form of energy technology, the introduction of energy technology, or the closure of an oil refinery. These pathways are fed to the two connected models, which are used to determine the impact of these events on the capacity of the energy infrastructure. After each model run, the output of the different models is recorded and transposed again by the EMA part to be able to perform analysis. To provide a conceptual overview of the model interaction between the two models, the global model steps of both models and interactions between the models are included in Figure 7.



Agent-based investment model
Technical-economical infrastructure model
Model interactions and communications
0.1| Setup/start new run
0.1| Setup/Reset model
Setup/reset model
1.1| Set year
1.1| Set year
Time sync
Year = 2020 + ticks
2.1| Set investment budget for operator *Depending on DM*
2.2| Determine demand expectations *Depending on DM (time horizon)*
2.3| Determine needed investments
Demand expectations
2.1| Update electricity network
2.2| Check which nodes' demand or production exceed capacity
2.3| For these nodes, check which investments are possible
2.4| Check which edges demand or production exceed capacity
2.5| For these edges, check which investments are possible
2.4| Filter investments already under construction
Needed investments options
Investment ID
Owner name
Additional overcapacity (MW)
CAPEX (M €)
Year finished
Amount
2.5| If no investments left, continue to step 3.1 or next operator. Otherwise:
2.6| Sort needed investments
2.7| If there are more options, choose the best alternative *Depending on DM (investment goal)*

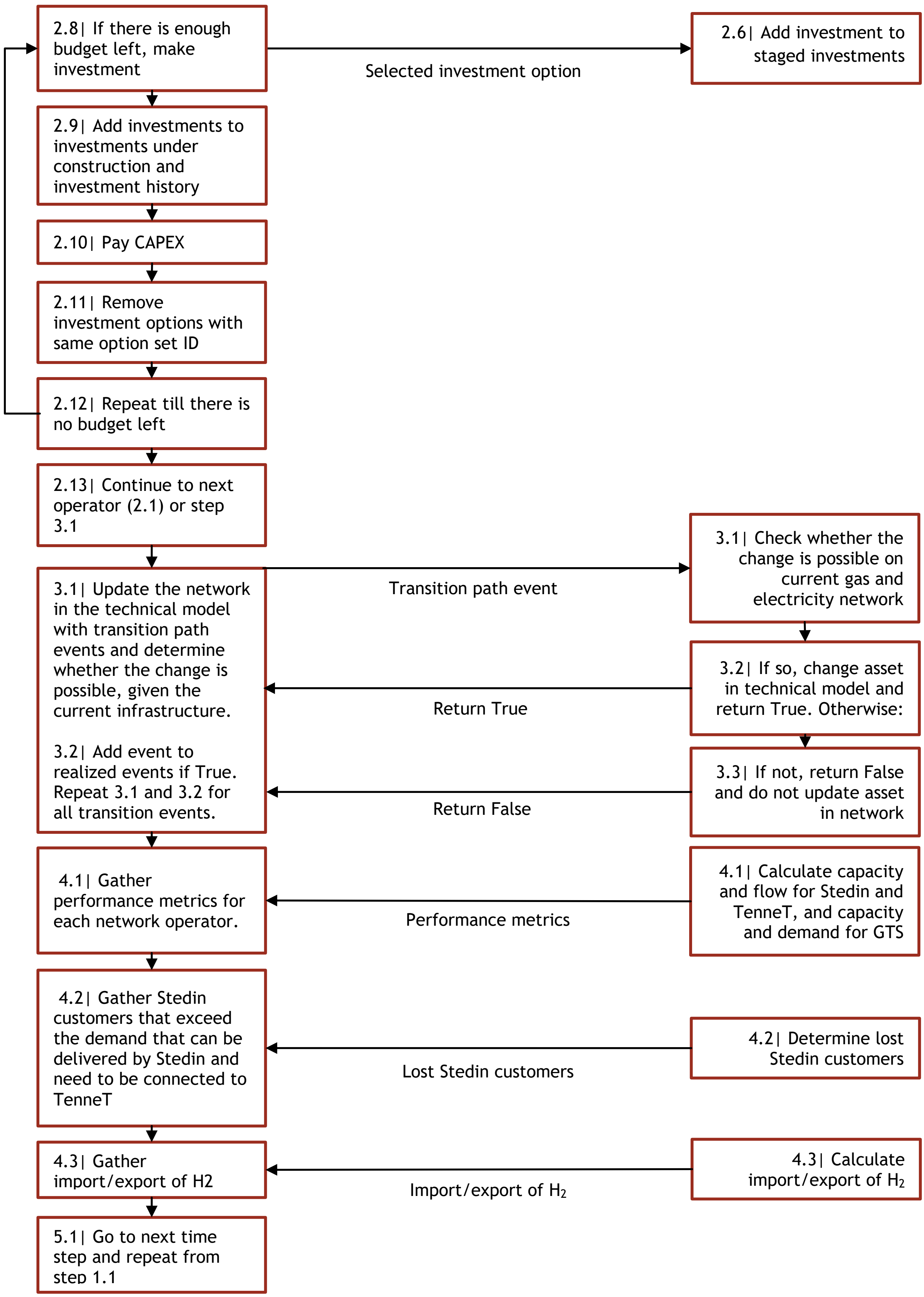


Figure 7 Simplified overview of model steps and interactions on the interface location between the agent-based decision-making model (left) and the economical-technical infrastructure model (right).

The interactions between the agent-based model and the technical model are mainly related to the transition path events and the communication of needed investments with assumptions on CAPEX, lead time, and realised overcapacity. Since we already cover the influence of uncertainty in the transition path events by the external factors, we focus on the communication of the investments. A couple of aspects of this communication may influence the decisions taken by the network operators: the sequence of the selected investment alternatives, the estimated lead time of the investment, and the estimated CAPEX of the investment. Also, the sequence in which network operators select the investments may influence the model outcomes.

To analyse these interactions on the interface level between the two models, we use four parameters to influence the interaction between the models:

- The sequence of the lists of investments

Every time step, the technical-economical infrastructure model generates a list of possible investments per network operator. The sequence in which these investment options is presented could influence the order in which the investments are chosen and thus, which investments are less chosen, because of budgetary constraints. We will use a Boolean parameter which determines if the sequence is shuffled or not.

- Lead time per investment option
- CAPEX per investment option

For each investment option, the infrastructure model presents the lead time and CAPEX per investment options as a fixed number. By influencing these values and varying them in an interval around these fixed numbers, we analyse the sensitivity for these values in the decisions taken by the investment behaviour model. We will use a stochastic continuous factor between 0.8 to 1.2 for both parameters.

- Random seed ABM

The random seed of the investment behaviour model determines in which sequence the network operators may make their investment decisions. By fixing the random seed, we analyse the influence of these decision orders. As an added benefit, we increase the reproducibility of the experiments, since the agent-based model will always give the same outcomes if the same random seed is used. We will use an integer parameter between -2147483648 and 2147483647, which are all the options for a random seed in Netlogo.

5.2.4 Outcomes of interest

According to Saltelli et al. (2019), a proper sensitivity analysis should always be based on the question that is aimed to be answered instead of the model in general. After all, a simulation model generally has many outputs, and the relation between the input variables and the output may differ substantially. The Windmaster model has a range of different outcomes, focussing on different aspects of the model. An overview of all defined model outcomes is presented in Table 7.

Table 7 Overview all of the defined outcomes of the Windmaster model with description and units.

#	Outcome	Description
1	Stedin load	The load on the Stedin energy infrastructure
2	Stedin capacity	The total capacity of the Stedin energy infrastructure
3	TenneT load	The load on the TenneT energy infrastructure
4	TenneT capacity	The total capacity of the TenneT energy infrastructure
5	Gasunie capacity	The total capacity of the Gasunie energy infrastructure
6	Investments	The type and amount of the implemented investment per investment type
7	Stedin CAPEX	The capital expenditure of the implemented investments by Stedin

8	TenneT CAPEX	The capital expenditure of the implemented investments by TenneT
9	Gasunie CAPEX	The capital expenditure of the implemented investments by Gasunie
10	Collaborative CAPEX	The capital expenditure of the implemented investments by the fictive collaborative energy network operator
11	H_2 import	Seaside hydrogen gas import
12	Stedin lost	Customers that switched from Stedin to TenneT electricity connection
13	First failure	The first year in which a transition pathway event could not be supported by the energy infrastructure
14	Missed over time	Amount of transition pathway events that could not be supported by the energy infrastructure

These outcomes cover the types and amount of realized investments, the financial aspects of these investments (CAPEX), the technical load distribution aspects of the infrastructure, and amount of achievable (and missed) transition path events. To retain a broad view, we want to choose Key Performance Indicators (KPI's) that cover these aspects. The KPI's should give insight into the performance, efficiency, types, and costs of the realized investments. To achieve this, KPI's included in Table 8 will be used in the uncertainty analysis.

Table 8 Selection of Key Performance Indicators covering the different aspects of the model outcomes and their descriptions.

#	KPI	Description
1	Investments (#)	The type and amount of the implemented investment per investment type. This KPI records the total amount of realized investments per investment type per model run (30 years).
2	Missed over time (#)	The amount of transition pathway events that could not be supported by the energy infrastructure. This KPI records the number of missed events per year.
3	Used capacity TenneT (%)	The portion of the realized infrastructure capacity that is used by the assets in the port region. This KPI is calculated by dividing the TenneT load by TenneT capacity and is presented as a percentage per year.
4	Total CAPEX (€)	The capital expenditure of the implemented investments by the different network operators (Stedin, TenneT, Gasunie, and Collaborative). This KPI records the CAPEX per year.

5.2.5 Value systems

Based on the value systems, a preference for certain ranges of outcomes may be specified, or there may be an emphasis on a particular outcome of interest. In the analysis of the Windmaster model, we can apply the value systems for two purposes. First, value systems are embedded in the specified decision strategies: whether a network operator chooses an investment alternative which creates as little overcapacity as possible (reactive decision making), which satisfies as many clients as possible (current decision making), which creates as much overcapacity as possible (proactive decision making), or creates capacity as fast as possible (collaborative decision making). Second, we can use value systems as part of a scenario discovery (or factor mapping) approach, where we divide the outcomes of interest into a part that is desirable and a part that may be less desirable. In this case, we may define threshold values for an outcome of interest and analyse what uncertainty ranges play a role in generating these model outcomes. Given this approach, network operators will probably tend to prefer lower CAPEX, as much used capacity as possible, and as little missed events as possible. Depending on the network operator, there might be a preference for specific extensions of the infrastructure grid, such as H_2 capacity or high voltage transformers.

5.2.6 XPIROV framework

Based on the identified external factors, policies, internal factors, relations, outcomes of interest, and value systems, we fill the XPIROV framework for the Windmaster model. The application of the XPIROV framework is included in Figure 8.

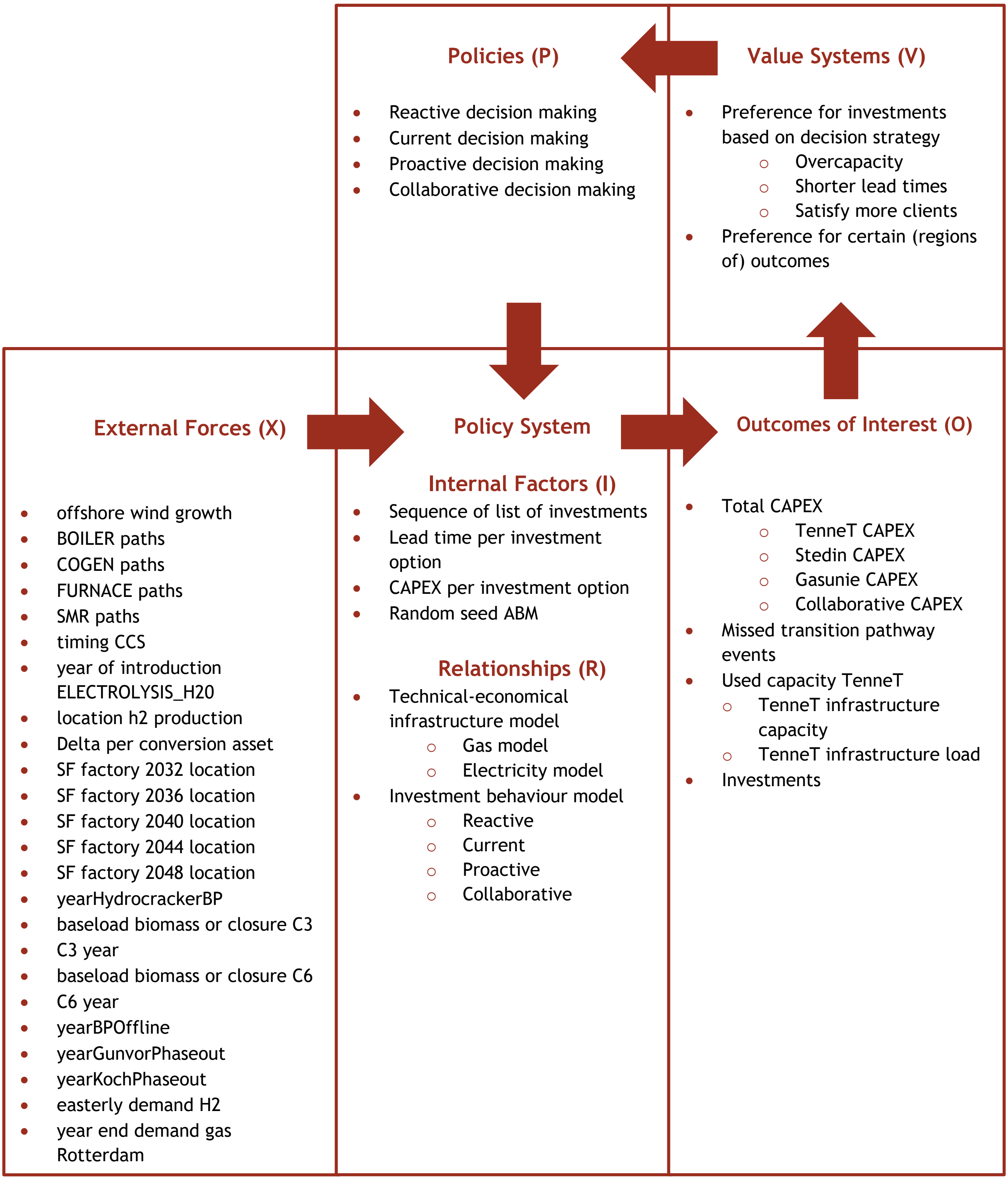


Figure 8 XPIROV framework applied to the Windmaster multi-model. The framework includes external factors (X), of which most are used to generate transition path events; policies (P) which influence different aspects of the decision making; internal factors (I) , which focus on the model interaction between the investment behavior model and the technical-economical infrastructure model; relationships (R) which include the different models of the multi-model; outcomes of interest (O) which contain the selected KPI's; and the value systems (V) which contain the investment goals of the network operators and the preference for (certain regions of) a KPI.

5.2.7 Uncertainty matrix

For a further structuring of the uncertainties in the Windmaster model, we apply the multi-model uncertainty matrix in Table 9. The focus is on the parameterized uncertainties and not on the corresponding technical or conceptual uncertainties. Therefore, we will distinguish uncertainties on both individual models, and for the remaining locations, we will combine the uncertainties on both the individual models and the multi-model ecology as a whole.

Table 9 Multi-model uncertainty matrix filled for the Windmaster model. The focus of analysis lies on uncertainties in the policy system and in the interface location. Due to the tightly coupled structure of the Windmaster model, external forces apply to both models and therefore, are combined.

Multi-model uncertainty matrix			**Level**				
			Level 1	Level 2 Shallow uncertainty	Level 3 Medium uncertainty	Level 4 Deep uncertainty	Level 5
Location	Conceptual model					• Included assets • Included future technologies	
	Agent-based investment model and the technical-economical infrastructure model	Model structures				Investment strategy	
		Model parameters inside the models			• Investment budget per infra provider • Time horizon per infra provider		
		Input parameters to the models		• Delta per conversion asset • Year BP offline • Year Gunvor phaseout • Year Koch phaseout	• Timing CCS	• Offshore wind growth • Boiler paths • Cogen paths • Furnace paths • SMR paths • year electrolysis H_2O • Location H_2 production • SF factory locations • Baseload biomass or closure C3 • Baseload biomass or closure C6 • Hydrocracker BP • Easterly demand H_2 • Year end demand gas Rotterdam • C3 year • C6 year	

	Model interface		• Lead time per investment option • CAPEX per investment option			
	Input data			• Investment alternatives • Current assets • Current infrastructure		
	Technical model implementation		Random seed	Sequence list of investments		
	Processed output data				• Selection of KPI's • Valuation of KPI's	

5.3 Extra-Trees Feature Scoring

In appendix A, we show the experiments used to generate the model outcomes. We note that there is no bias in the experiments that could lead to unreliable results of the analysis. In this section, we discuss the results of the factor fixing analysis. We look into the generated results. Next, we show the results of the sensitivity analysis. Finally, we will interpret these results.

5.3.1 Results

Overview of possible model outcomes per KPI

The line plots in Figure 9 visualize the KPI’s over the model run time. Each line represents a possible outcome of the model, based on one experiment, averaged over ten replications.

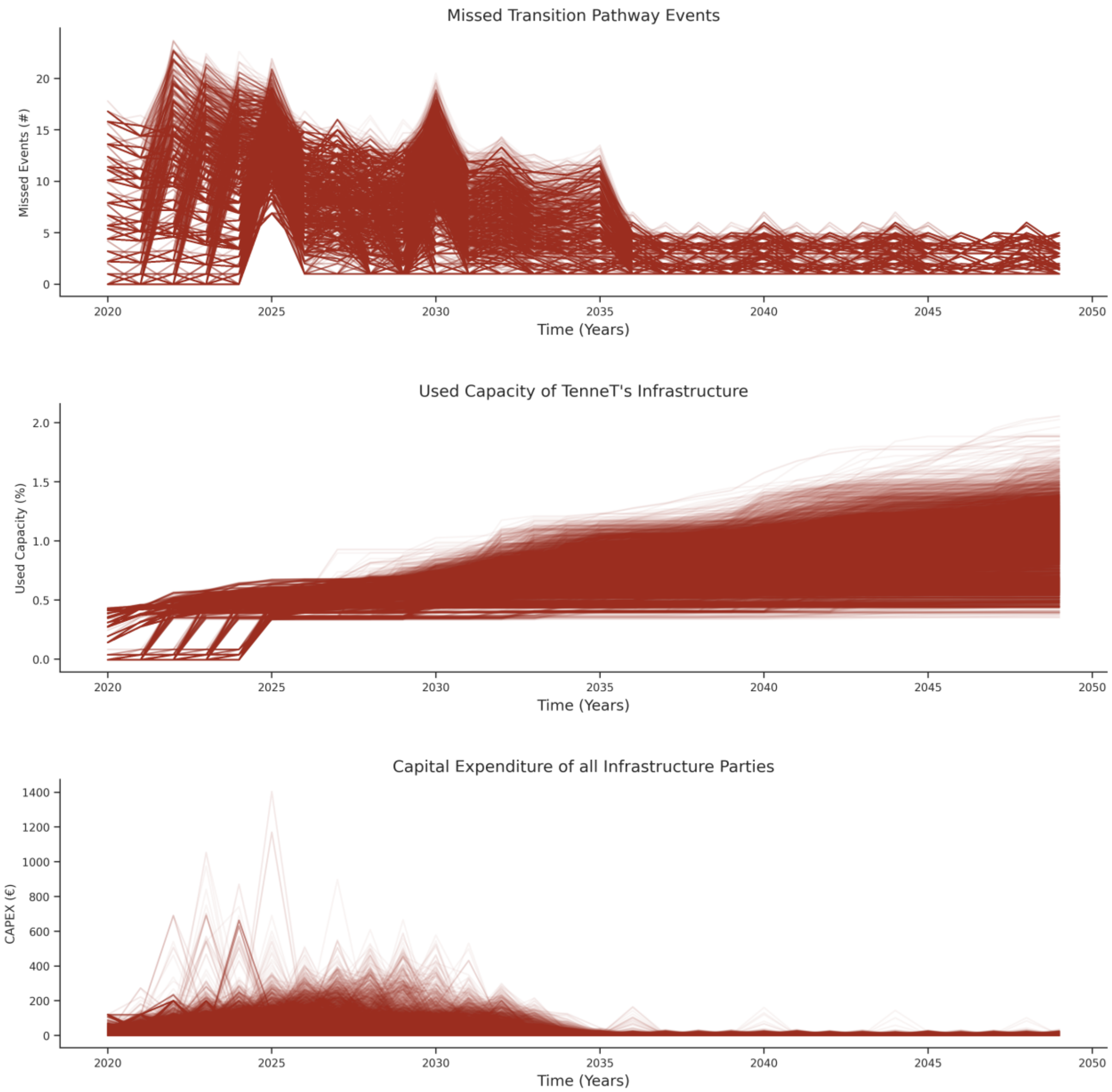


Figure 9 Line plots showing outcomes for each KPI over time. All lines represent the realisation of one experiment, averaged over ten replications. A high degree of opacity has been used to make clear where many realisations lie.

Feature scores

We first focus on the (cumulative) end state of each KPI. A heatmap of the calculated feature scores is included in Figure 10.

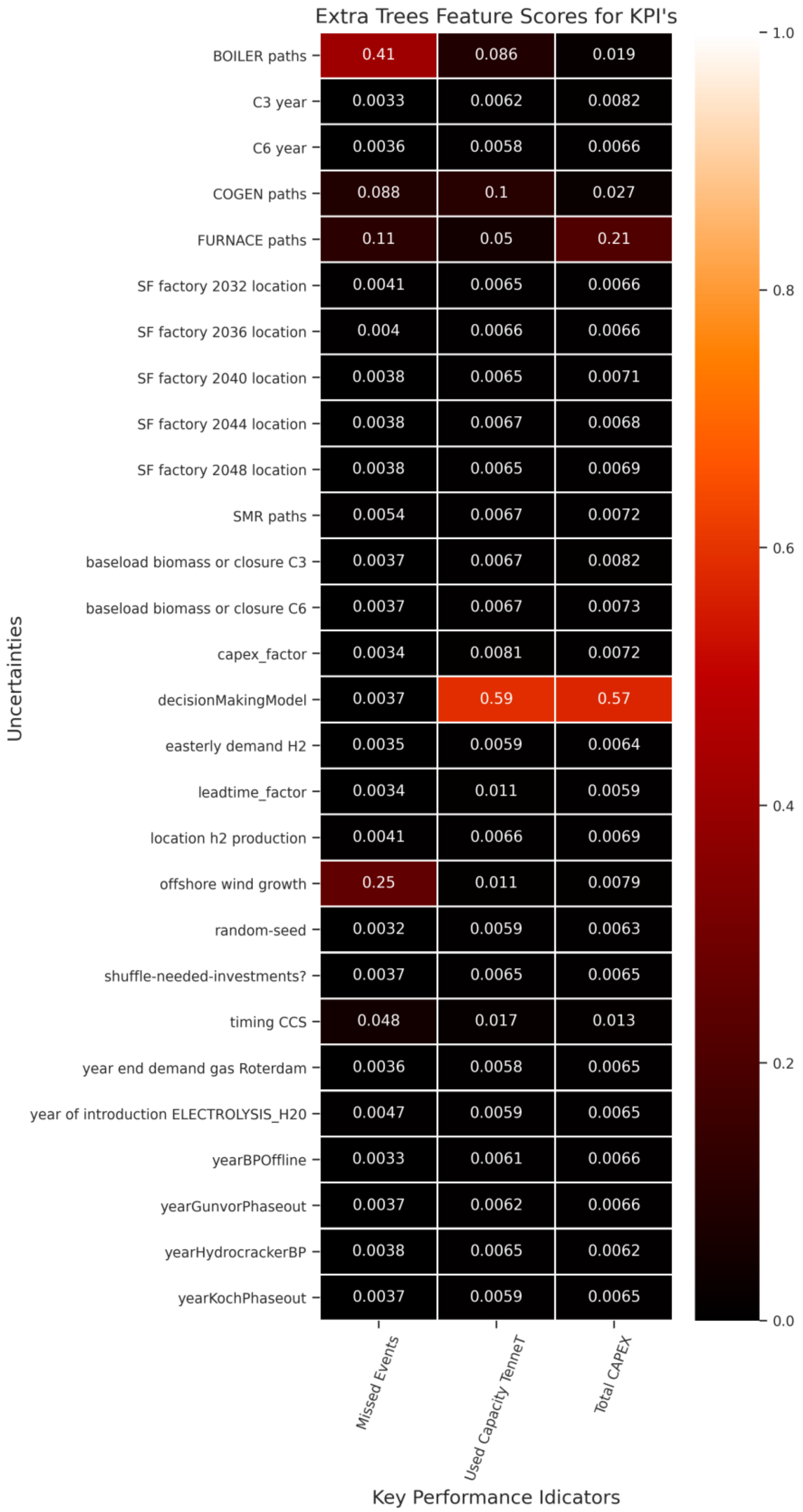


Figure 10 Heatmap of extra-trees feature scores on the (cumulative) end states of each KPI. The scores indicate the degree of influence each uncertainty has on the concerning KPI. No influence is indicated by 0, while high influence is indicated by 1.

It is striking that only a few indicators stand out, while most uncertainties do not seem to have a strong influence on any KPI. Besides, the influence of uncertainties differs per KPI. For example, the applied decision-making model seems to have a high impact on both the used capacity of the TenneT infrastructure and the total CAPEX, while it does not seem to influence the number of missed events. To explore the influence of the decision-making model on the KPIs, we group the outcomes by the decision-making models in Figure 11. Based on this grouping, we indeed distinguish different behaviour for the KPI's, except for the missed events. Moreover, for the used capacity of TenneT, the impact of the decision-making model seems to differ over time.

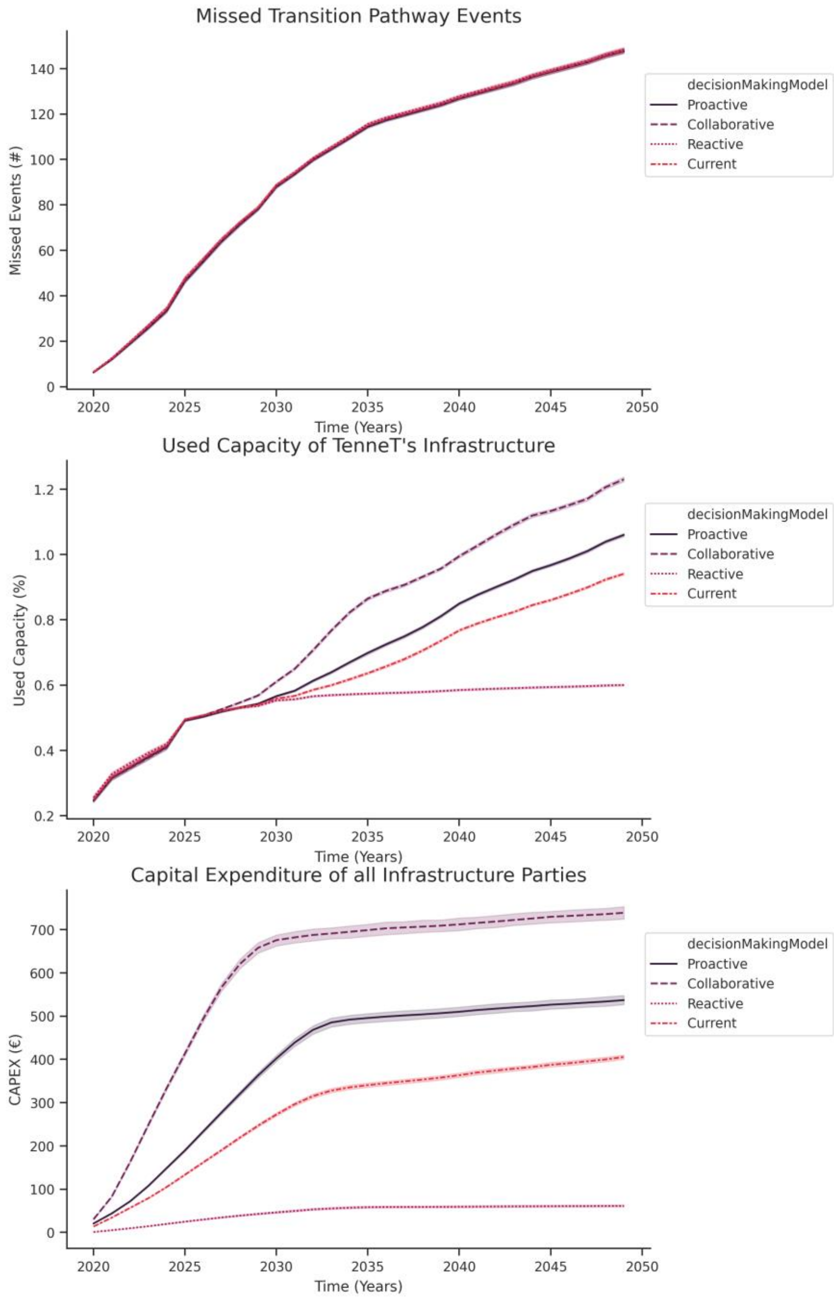
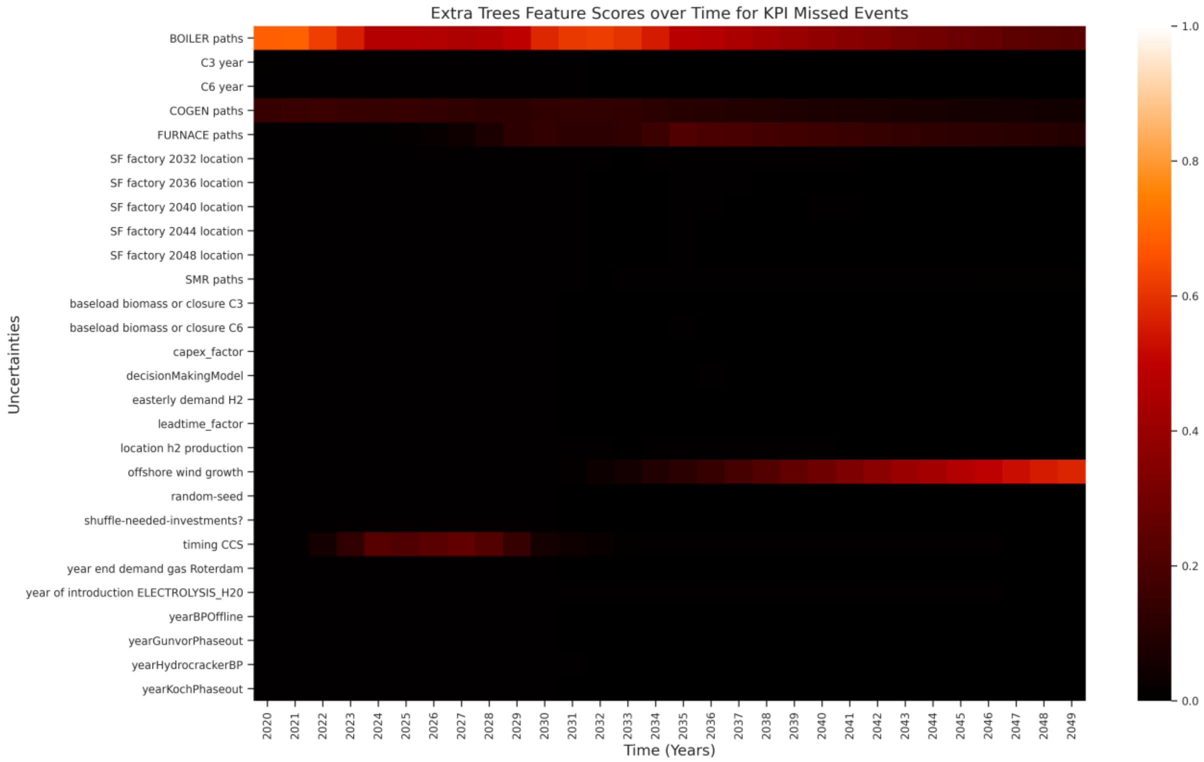
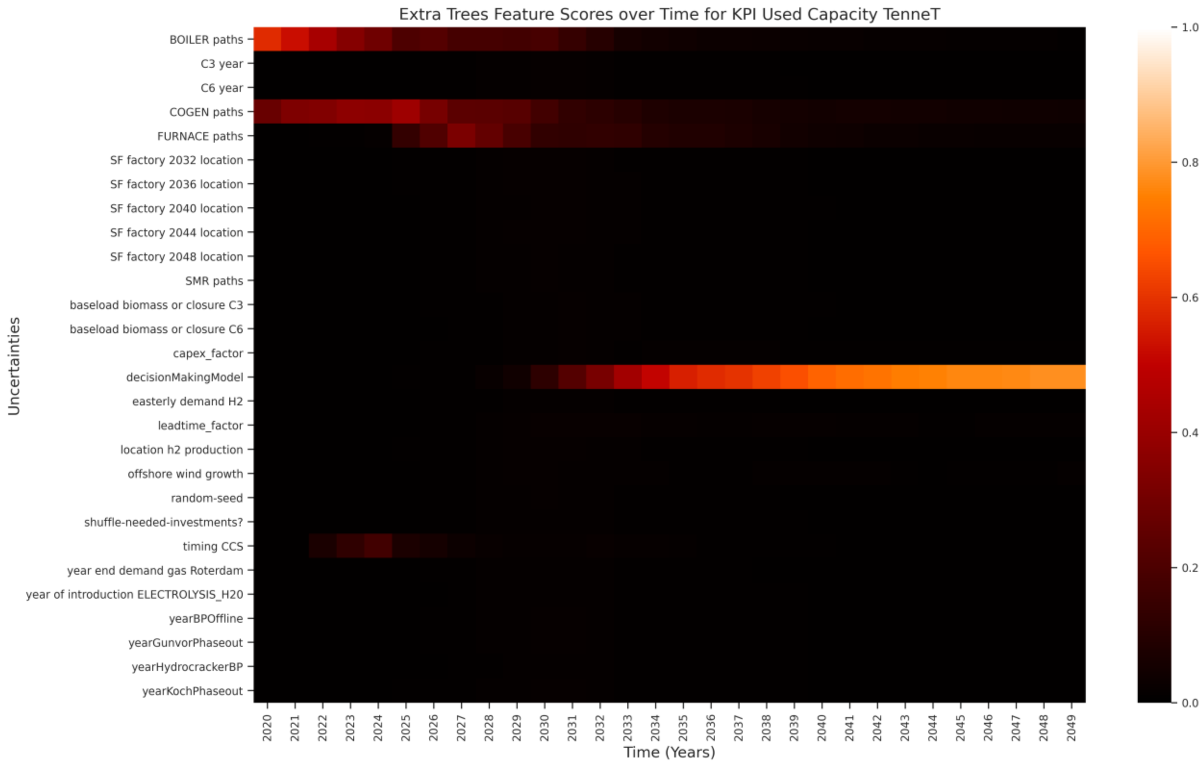


Figure 11 **The development of the Key Performance Indicators over time, disaggregated per applied decision-making strategy. The strategy does not seem to influence the amount of missed events, but it influences the used capacity of TenneT from 2025 on, while it influences the CAPEX from the beginning.**

We look further into the development of influence over time by performing extra-trees feature scoring per KPI for each additional time step in Figure 12.





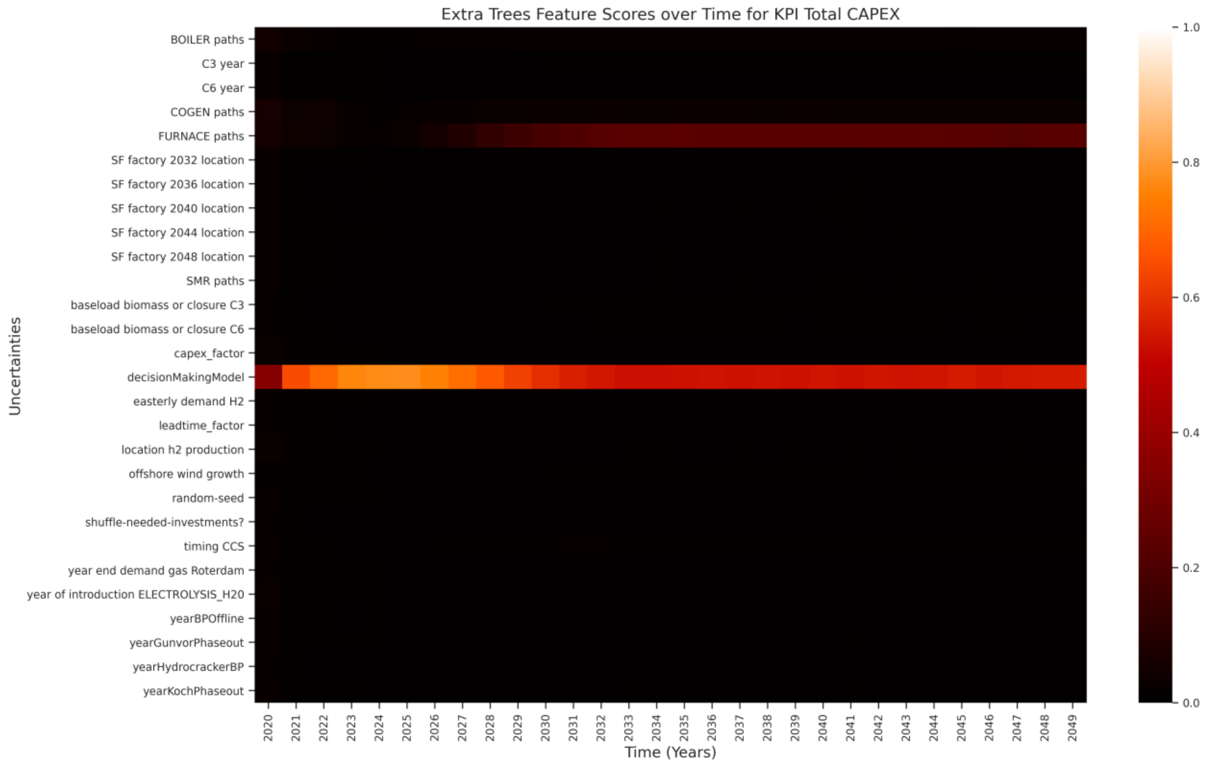


Figure 12 Extra-trees feature scoring over time per KPI. For each year, the influence of the uncertainty on the realization of the KPI has been recalculated. No influence is indicated by a black square, while a white square indicates a strong influence. Some uncertainties only influence the concerning KPI in specific moments in time, while others remain relatively constant.

For TenneT's used capacity, we indeed see that the influence of the decision-making model only starts impacting the outcomes from 2030 and onwards. Until that moment, the KPI is mainly driven by a combination of the boiler paths, cogeneration paths, furnace paths, and the timing of introduction of CCS. For the CAPEX, we see that the furnace paths start influencing the KPI from 2026 onwards. The influence of the decision-making model differs over time. For the missed events, the influence of the boiler paths remains relatively the same over the model run, while the offshore wind growth starts only influencing this KPI from 2035 onwards. The influence of the furnace paths seems to take over the influence of the cogeneration paths.

Investments

We now look into the remaining KPI: the investments. As this KPI is not recorded over time and consists of the different investment alternatives and their realized investments, the type of analysis we perform is different from the other KPI's. The first step is to gain insight into what type of investments are generally realized. To visualize this, we create a heatmap over the whole range of experiments in Figure 13.

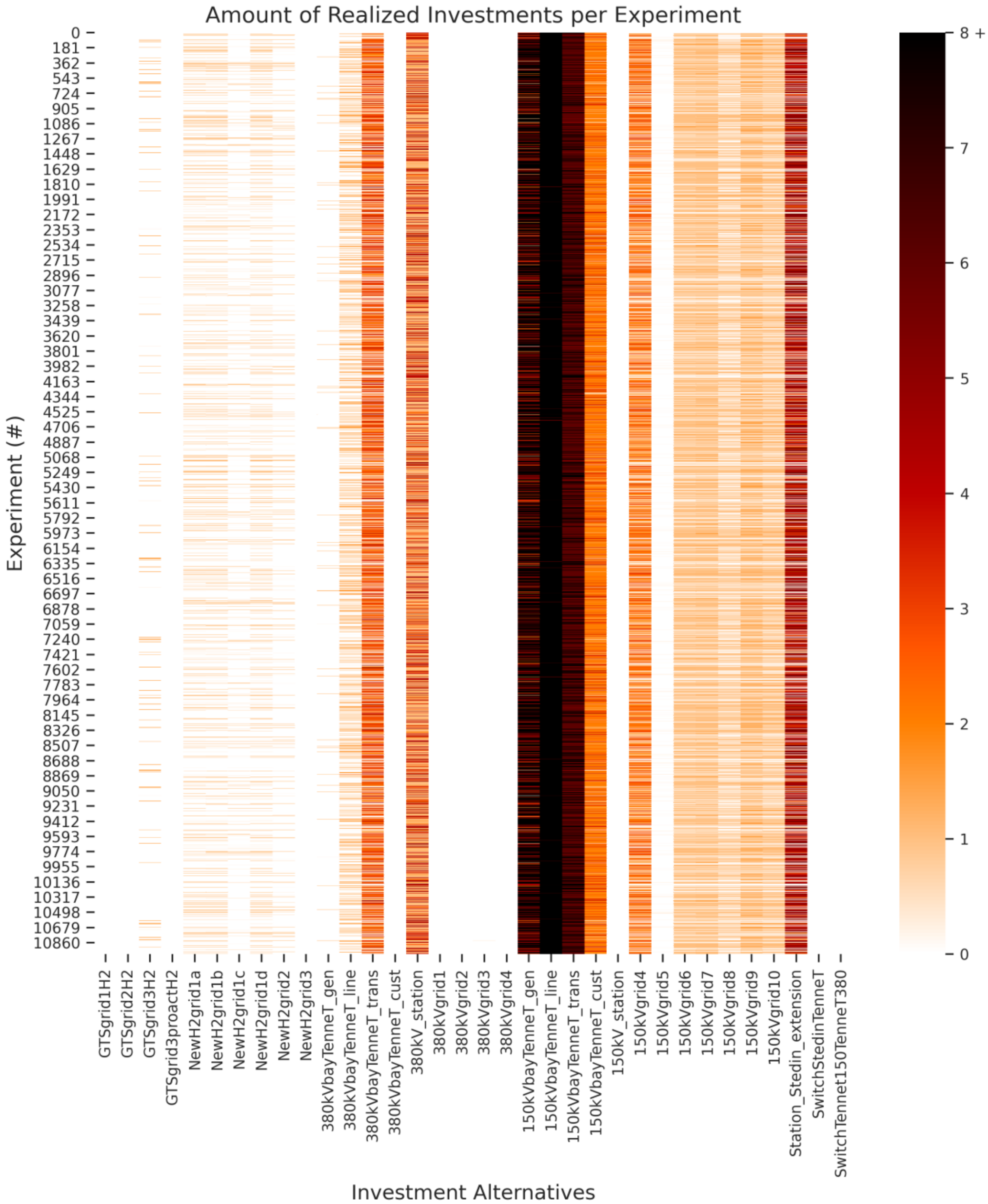


Figure 13 Heatmap of the realized investments in the various experiments. The colour indicates how many investments of a certain type have been made. The figure gives an idea of the realization of investments across the whole range of experiments.

As we can see, the type and amount of investments differ substantially for some investment types, while other investment types are either almost always or never realized. To gain insight into the driving uncertainties for the different type of investments, we again perform extra-trees feature scoring. The scores are visualized in a heatmap in Figure 14. Again, we see that only a small part of the uncertainties play a role in the decision for an investment type. The decision-making model seems to have a strong influence on the type and amount of investments like we also see on the CAPEX KPI. This is to be expected since the CAPEX is mostly related to the investments.

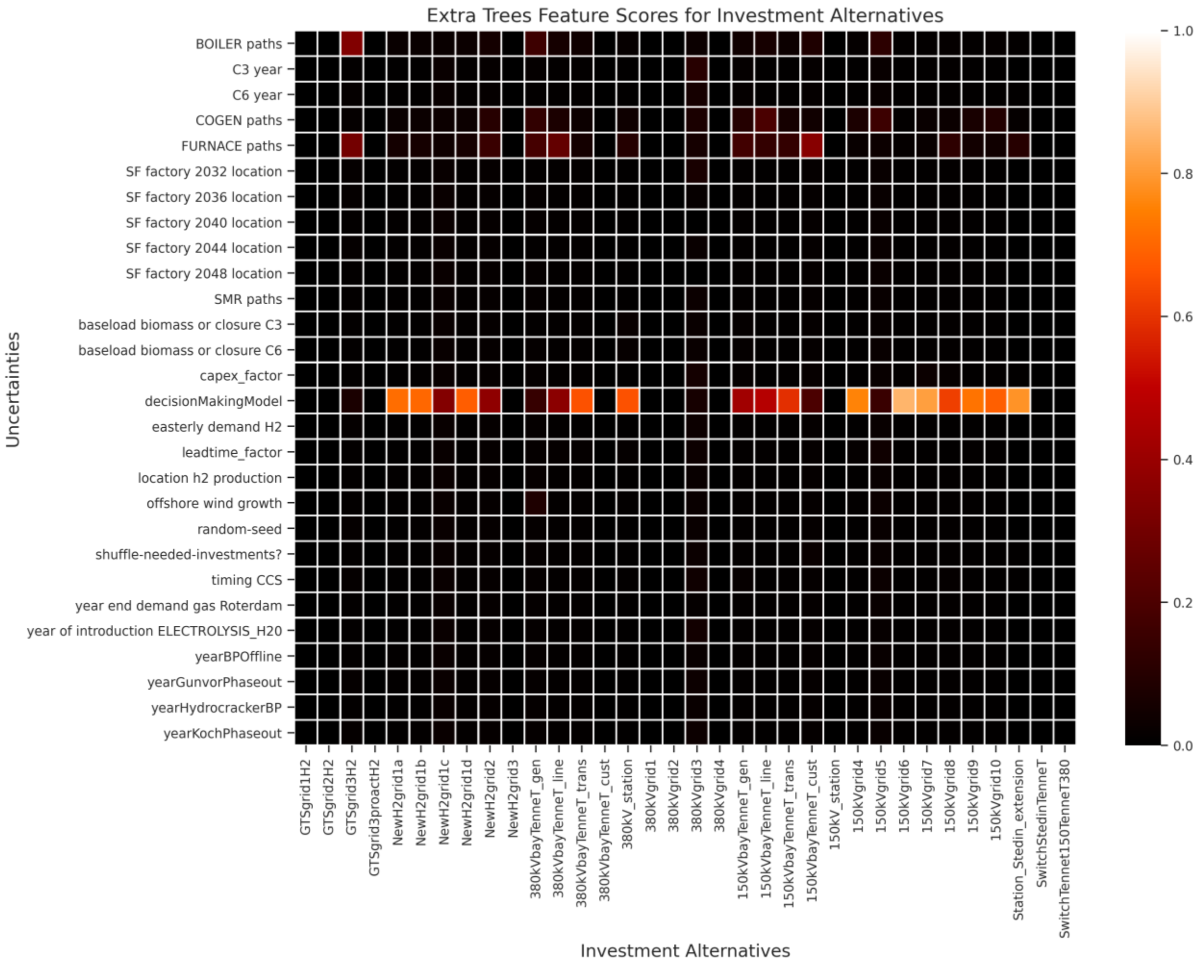


Figure 14 Extra-trees feature scores per investment alternative. The colour indicates the influence of uncertainties per investment alternative. No influence is indicated by a black square, while a white square indicates a strong influence.

To visualize the relationship between the decision-making model and the choice for investment types, we regroup the realized investment types on the underlying decision model in Figure 15. We indeed see that the choice for the type and amount of realized investments differs per decision making strategy, with the most striking aspect being the choice for new hydrogen gas grids, chosen almost only when applying the collaborative decision-making strategy. There is also a big difference between the reactive decision-making strategy and the other strategies, with a reactive strategy leading to less realized investments.

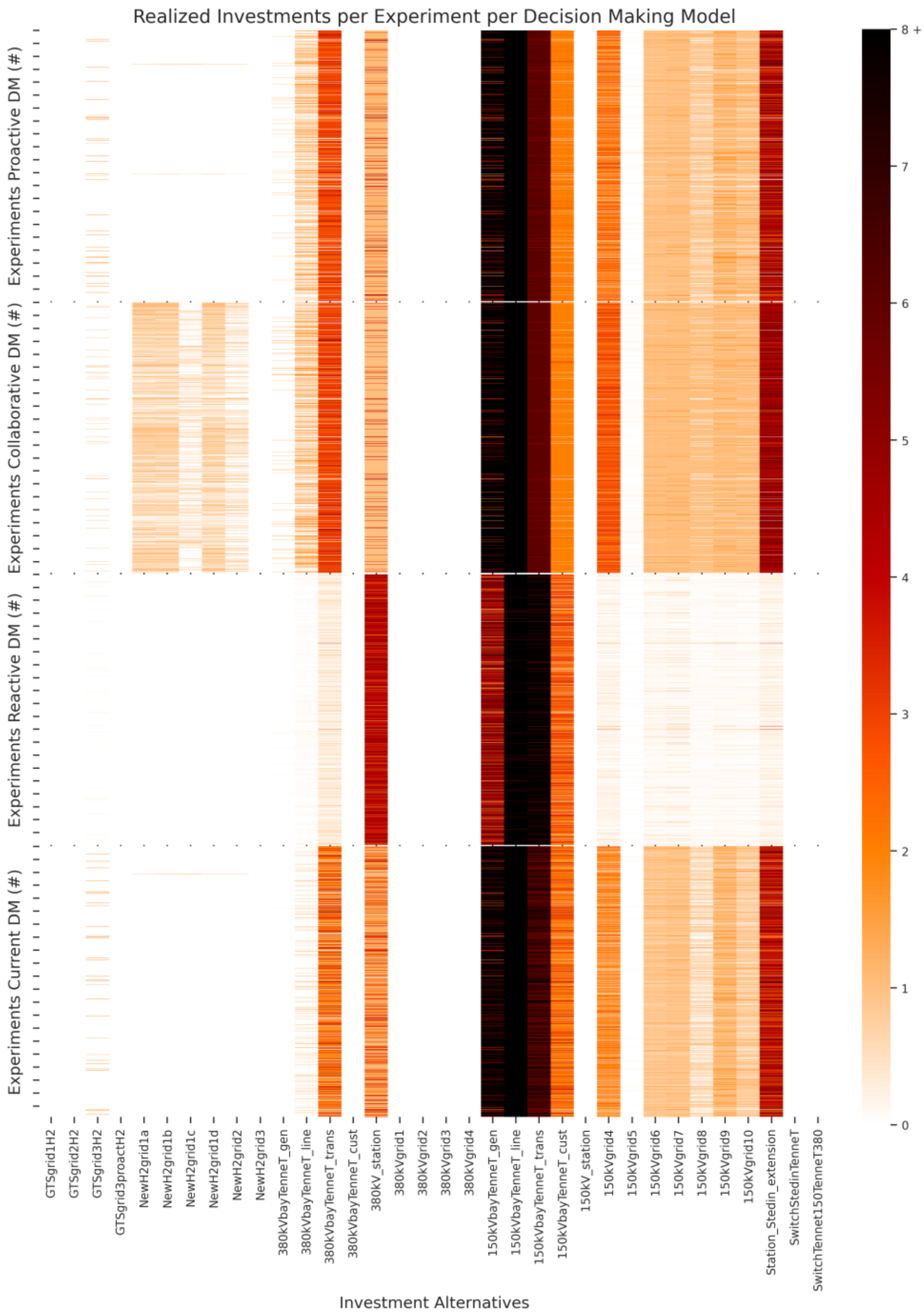
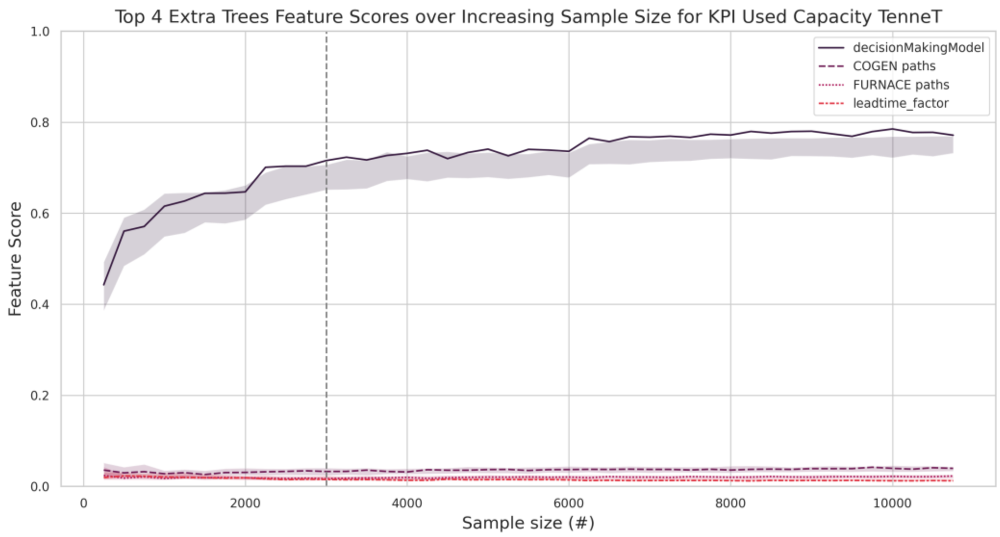
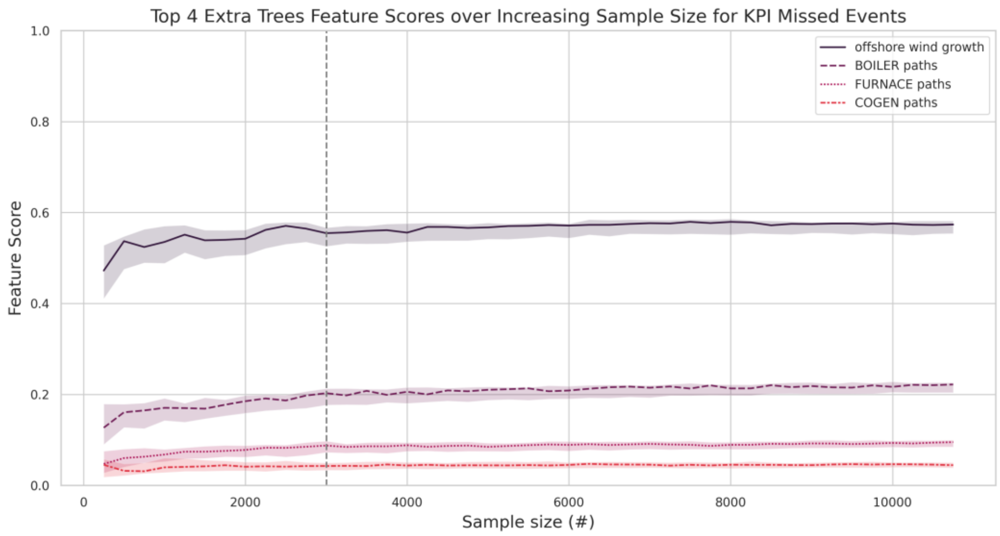


Figure 15 Types of investments implemented and their number, subdivided into experiments with the different decision-making strategies. The colour indicates how many investments of a specific type have been made. Because the experiments are broken down according to the investment strategy used, this indicates the difference in (types of) investments under the influence of different strategies. It is notable, for example, that hydrogen-related investments are made almost exclusively in collaborative decision making, and barely 150 kV investments are made under reactive decision making.

Convergence

The reliability of the method depends on the chosen amount of experiments and replications per experiment. In Figure 16, we analyse the convergence of the feature scores on the KPI's for the four uncertainties that have the most influence on these. This analysis is based on a repetitive extra-trees feature scoring while adding 500 samples each iteration. For each additional block of samples, we repeat the feature-scoring based on a bootstrapped repetition on the same samples and show the minimum and maximum value of the feature score, based on this bootstrapped repetition. We take into account both the order of importance of the uncertainties and the convergence of the scores to a relatively stable level. On that basis, we note that 3,000 experiments seem enough to get reliable feature scores.





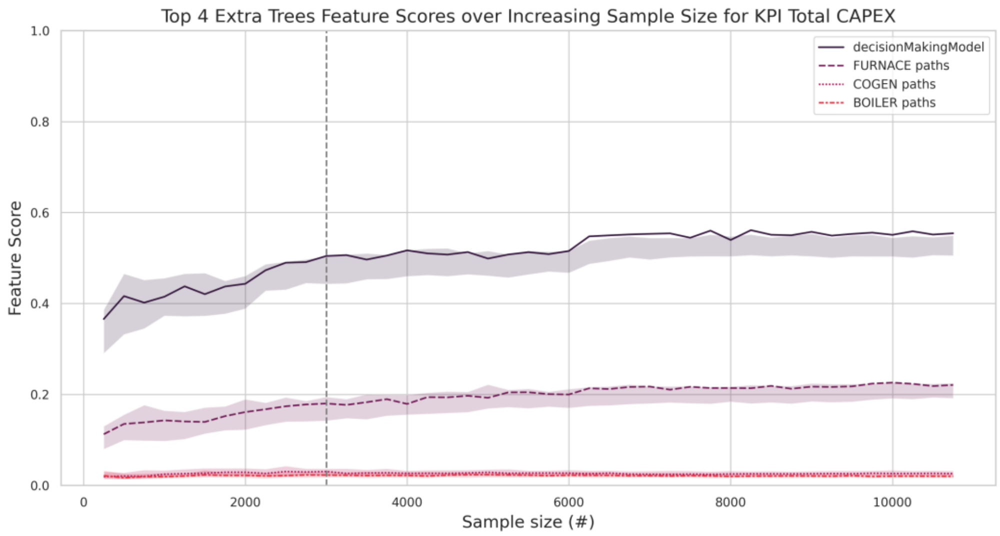


Figure 16 Feature scores based on increasing sample size, on the top 4 most influencing uncertainties. The (dashed) lines indicate the scores based on an Extra-Trees feature scoring over the whole sample set. The shaded areas indicate lower and upper limits based on a bootstrapped repetition of the feature scoring. Every 500 samples, the feature scores are recalculated and repeated for random sampling with replacement. The number of bootstrapped reruns is equal to the amount of included samples, divided by twenty. The shaded area, therefore, does not represent a confidence interval, and this explains why the (dashed) lines may exceed the shaded area. Although this might be solved by repeating the bootstrapping process more often, the goal of the figure is to get a sense of the decreasing uncertainty in the feature scoring by adding more samples, until both the feature scores and the bandwidth of the bootstrapped results converge.

Stochastic influence

To determine the influence of stochastic uncertainties on the model outcomes, we calculate the Extra-Trees feature scores for each replication in Figure 17. For the missed events, we see that there is little to no difference in the calculated scores. For the total CAPEX, we see some differences in the calculated scores on the decision-making model for the first replication (# 0) relative to the other replications, while the rest of the scores are pretty stable over the different replications. However, the impact of the different replications on TenneT's used capacity seems very high on the scores for all uncertainties. Based on this observation, we note that we may need to include more replications per experiment to get more reliable feature scores.

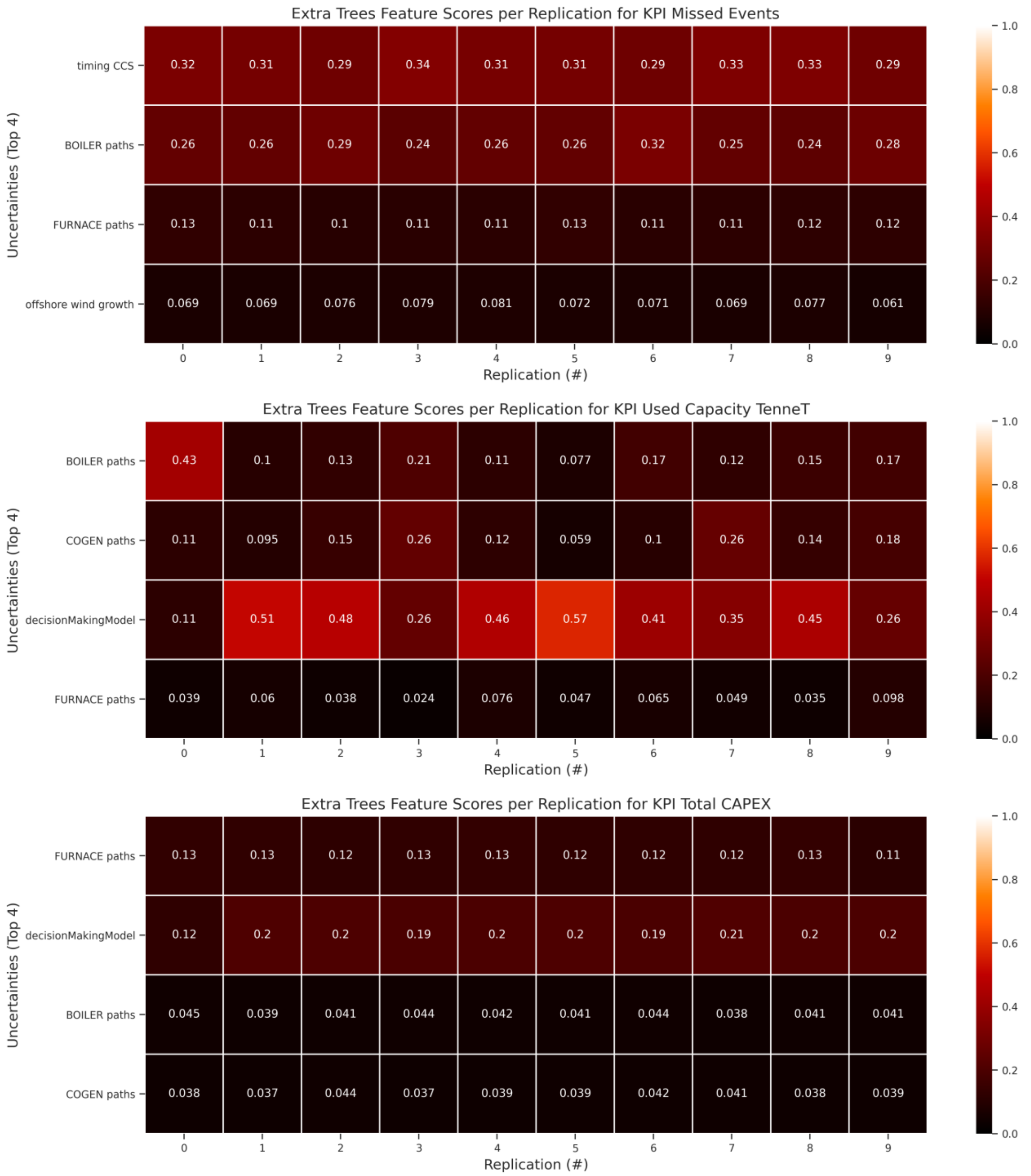


Figure 17 Extra-Trees feature scoring per replication. If the scores per replication do not differ that much, the KPI is not sensitive for a different set of deltas per conversion asset. For the missed events, there is little to no difference in the calculated scores. For the total CAPEX, the calculated scores on the decision-making model for the first replication (# 0) is lower relative to the other replications, while the rest of the scores are pretty stable over the different replications. The impact of the different replications on TenneT's used capacity seems very high on the scores for all uncertainties.

5.3.2 Main findings: factor fixing

As regards the distribution of the results, we see a (very) wide distribution in the number of missed transition events, while the used capacity has a narrower distribution. The total CAPEX is relatively narrow distributed as well. We attribute the uncertainty in the various outcomes to the uncertainties. The results show that the influence of uncertainties differs per KPI and per time period in the model. We distinguish the following uncertainties that overall have a limited impact on the outcomes:

- Baseload biomass or closure C3
- Baseload biomass or closure C6
- C3 year
- C6 year
- Easterly demand H_2
- Location H_2 production
- SF factory 2032 location
- SF factory 2036 location
- SF factory 2040 location
- SF factory 2044 location
- SF factory 2048 location
- SMR paths
- Year BP offline
- Year end demand gas Rotterdam
- Year Gunvor phaseout
- Year hydrocracker BP
- Year Koch phaseout
- Year of introduction electrolysis H_2O

We fix these uncertainties in the following analyses on randomly selected values within their bounds. The following uncertainties appear to have a significant influence and therefore remain in the analysis:

- Boiler paths
- Cogeneration paths
- Decision-making model
- Furnace paths
- Offshore wind growth
- Timing CC

The choices for technologies to produce steam (captured in the 'boiler paths'), for the combined production of heat and electricity (captured in the 'cogeneration paths'), and for the production of heat (captured in the 'furnace paths') mainly have an impact on the amount of missed transition events. Also, the technology choices contribute to the used capacity in the period from 2020 to 2035. For the total investment costs, we only see the impact of the technology choice to produce heat from 2025 to 2050. The remaining technologies do not influence investment costs. The used decision-making strategy has a considerable impact on the TenneT capacity used and the total investment costs but has a negligible influence on the amount of missed transition events. Although the influence of the strategy is relatively constant for the total investment costs, we see that it only starts to play a role from 2030 onwards for the capacity used. The increase in offshore wind energy from 2035 onwards has a high impact on the amount of missed transition events but does not influence the TenneT capacity used and the total investment costs.

The timing of the introduction of Carbon Capture and Storage for the production of hydrogen gas has a relatively strong influence on the TenneT capacity used and the amount of missed transition events, around 2025. The year of introduction does not influence the total investment costs.

The combination and degree of influence of the uncertainties differ per investment alternative. To provide an overview, we included the degree of influence per uncertainty per investment alternative, for only the four impactful uncertainties and the investment alternatives that are most dependent on these uncertainties in Figure 18.

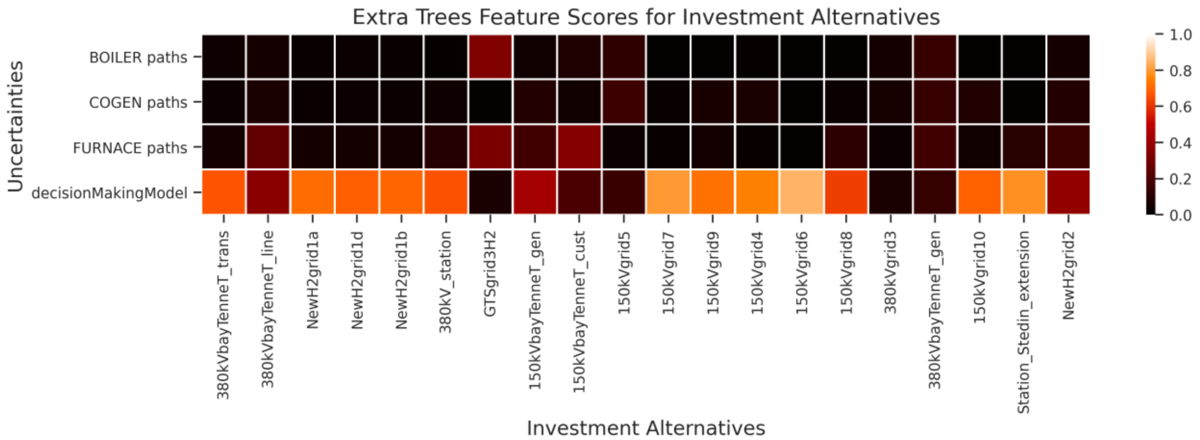


Figure 18 Degree of influence for the four uncertainties with the highest impact on the choice for investments. A black square indicates no influence of the uncertainty on the respective investment alternative, while a white square would indicate a powerful influence.

The uncertainties related to the interface have a limited impact. Still, we include them in the following analysis to analyse their effects on the model outcomes more deeply. Since the CAPEX factor and lead time factor are partial outcomes, we expect that there are interaction effects with other uncertainties influencing the CAPEX and lead time, such as the decision model. We increase the bandwidth of both the CAPEX and lead time factor from 20% to 30%, to investigate if under these parameter values the importance of these factors increases.

The stochastic influence on the model results is relatively limited. We, therefore, keep the number of replications the same for further analysis.

5.4 Global Sensitivity Analysis: Sobol

In appendix B, we provide insight into the distribution of the samples in the experiments. In this section, we will discuss the results of the Sobol analysis.

5.4.1 Results

Based on the Sobol sampling, we get slightly different results compared to the MC sampling with more included uncertainties. Although most results are comparable, the KPI 'missed events' shows a somewhat less continuous spread than the previous MC sampling. An overview of the different outcomes is included in Figure 19.

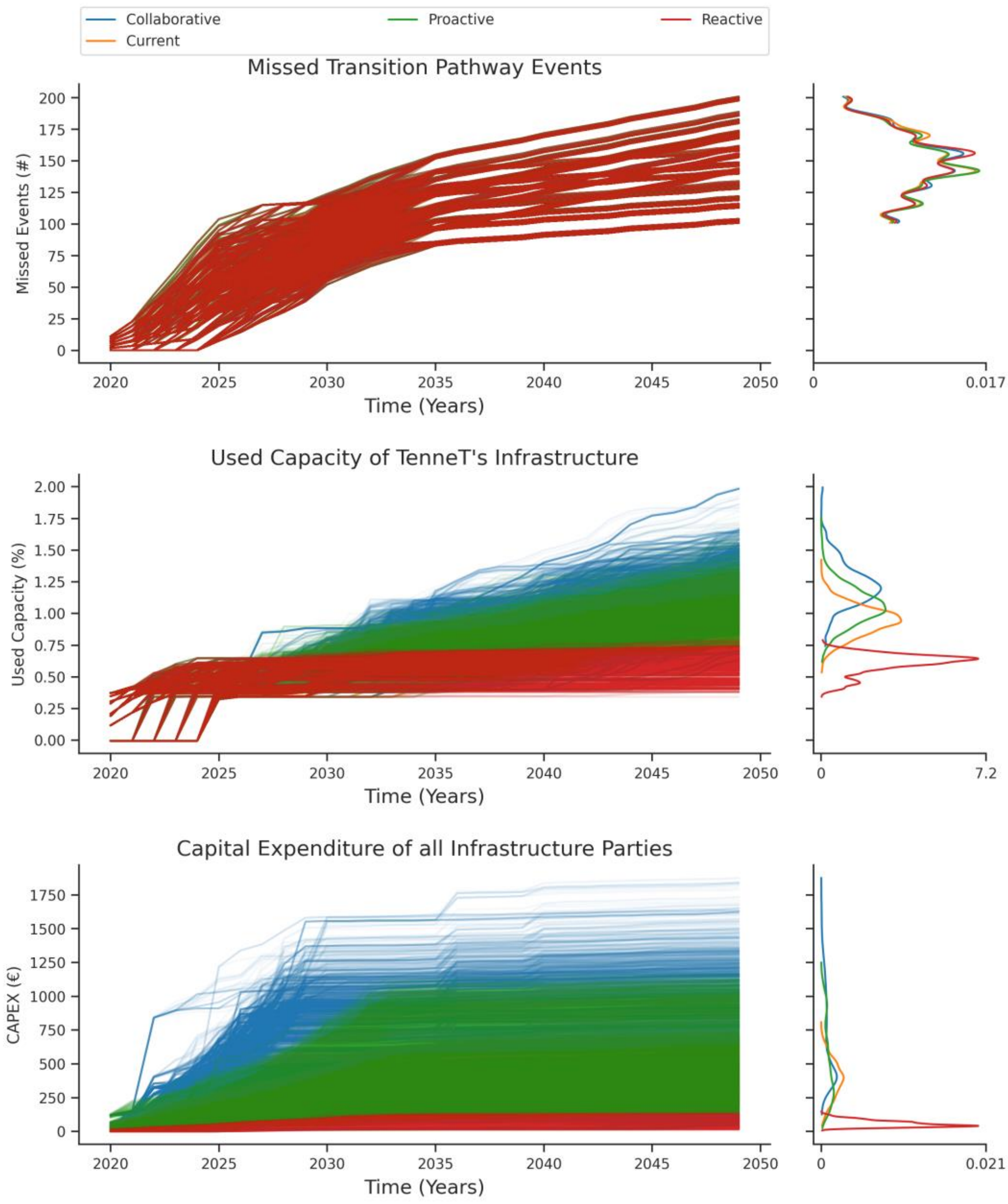


Figure 19 Line plots for the different KPI's based on the Sobol sampling. The colours indicate the underlying decision model. The plot on the right is a KDE plot and indicates the probability density function for each KPI, again divided based on the decision-making model.

First, we focus on the first- and total-order sensitivity indices for each outcome. These are included in Figure 20. It is striking that for the KPI missed events, there seem to be nearly no interaction effects between the uncertainties. This can be deduced from the minimal discrepancy between the first-and total order sensitivity index. For the other outcomes, there seem to be interaction effects between the uncertainties. As with the previous analysis, we see that the influence of uncertainties on the various outcomes is different. Uncertainties that have a strong influence on one outcome have little or no influence on the other outcome.

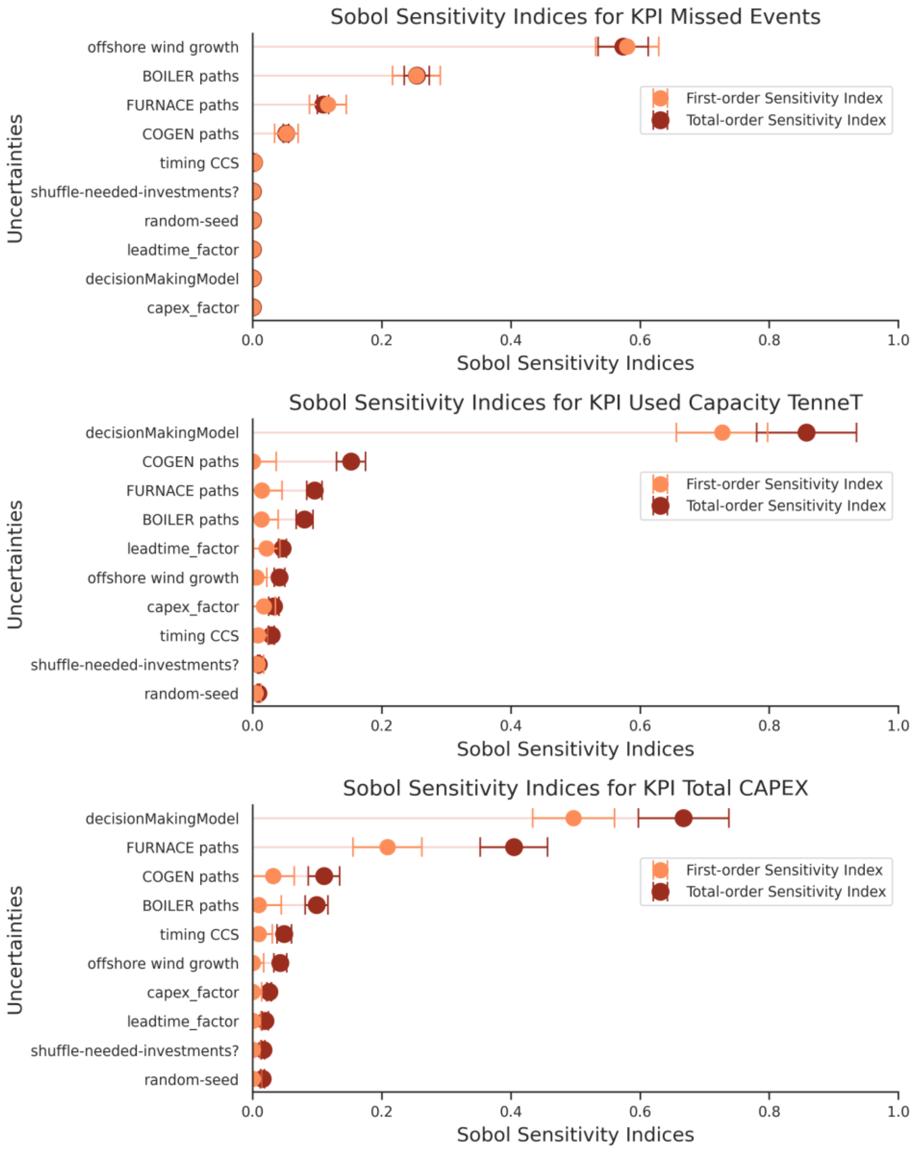


Figure 20 **Sobol Sensitivity Indices for each KPI. The total-order sensitivity index indicates the total contribution to explained variance, thus includes the first-, second-, third-, and higher-order sensitivity indices. The first-order sensitivity index indicates the main effect of the uncertainty on the variance of the KPI's. The further apart these two indices are, the higher the indirect (higher-order) interaction effects with other variables. For each index, the confidence interval is indicated by an error bar.**

The Sobol indices are initially calculated on the cumulative endpoints of the three KPI's. Here, too, it is interesting to look at the development of these scores over time. In Figure 21, we look at the total-order Sobol indices, which thus contain the main effect and all interactions with other uncertainties.

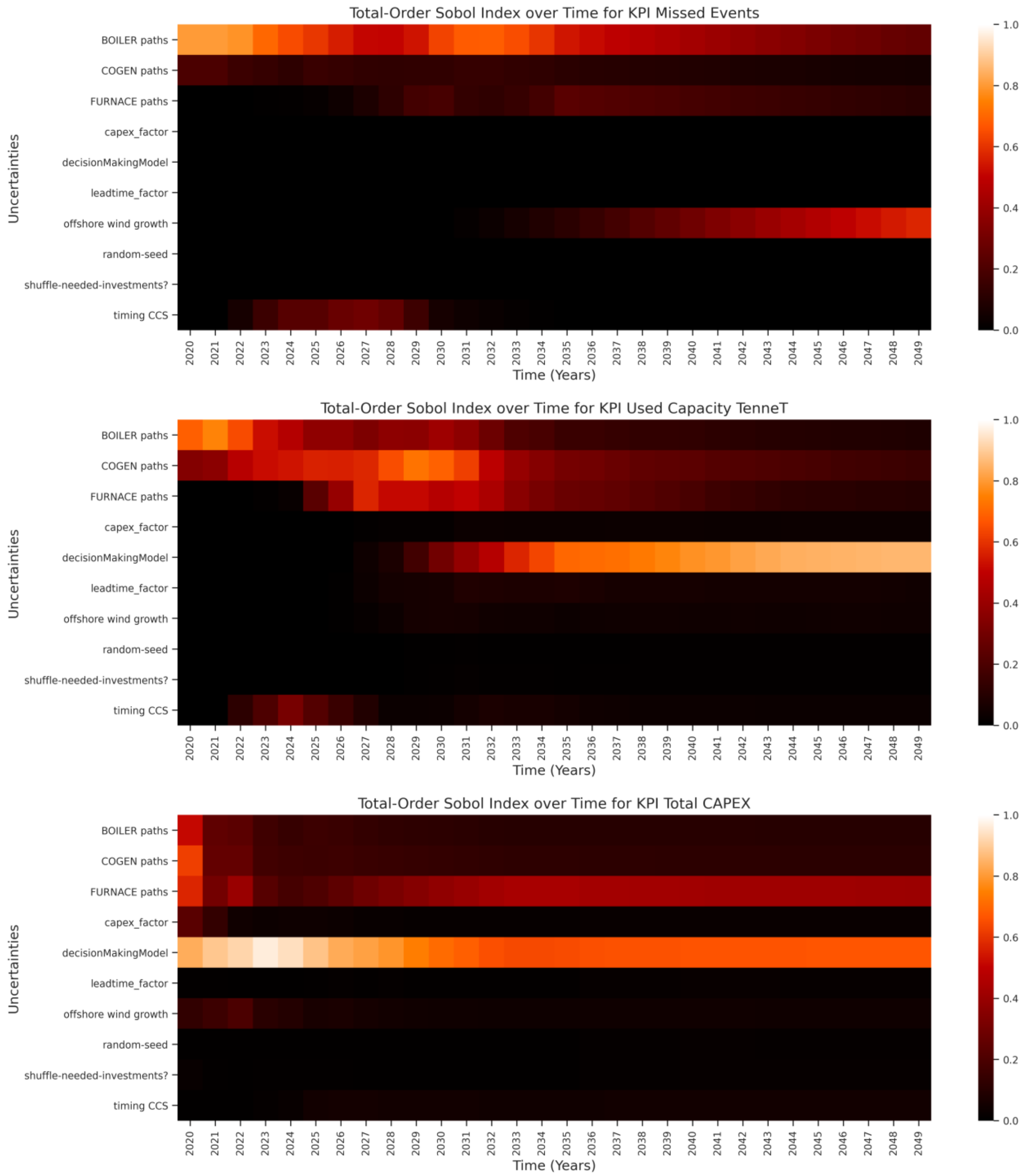
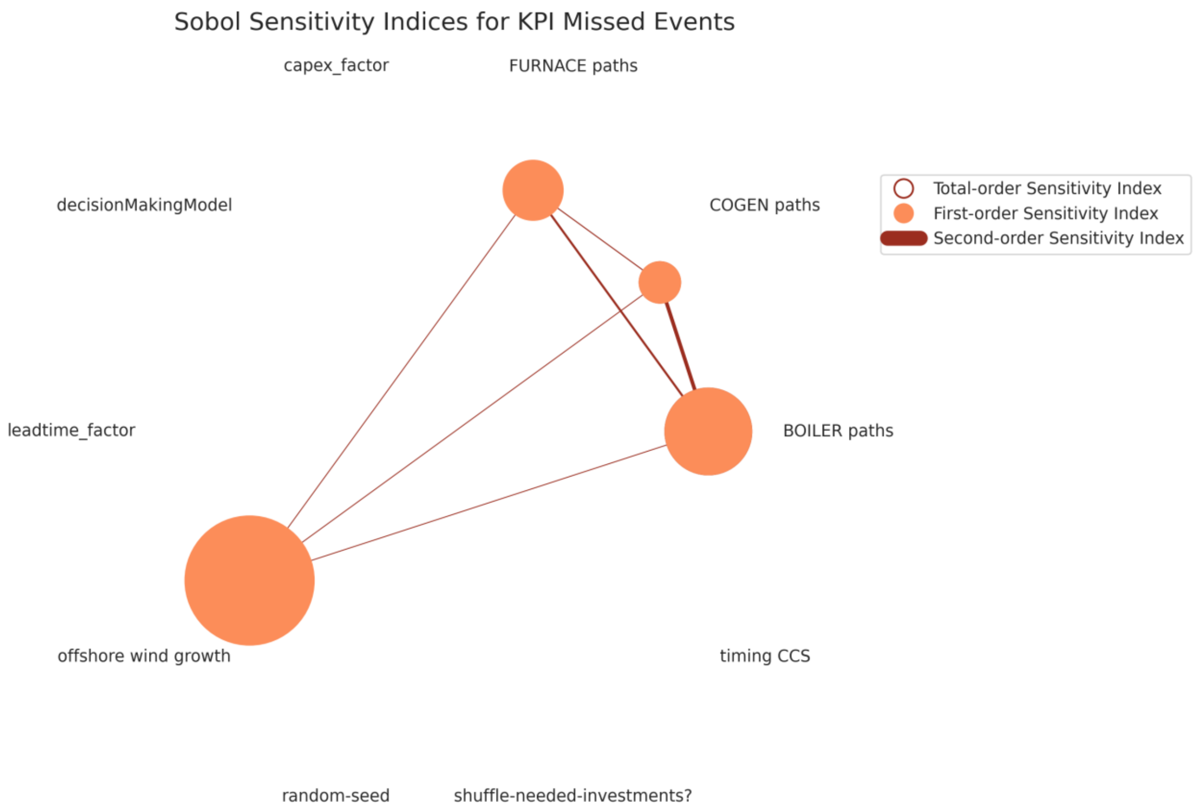


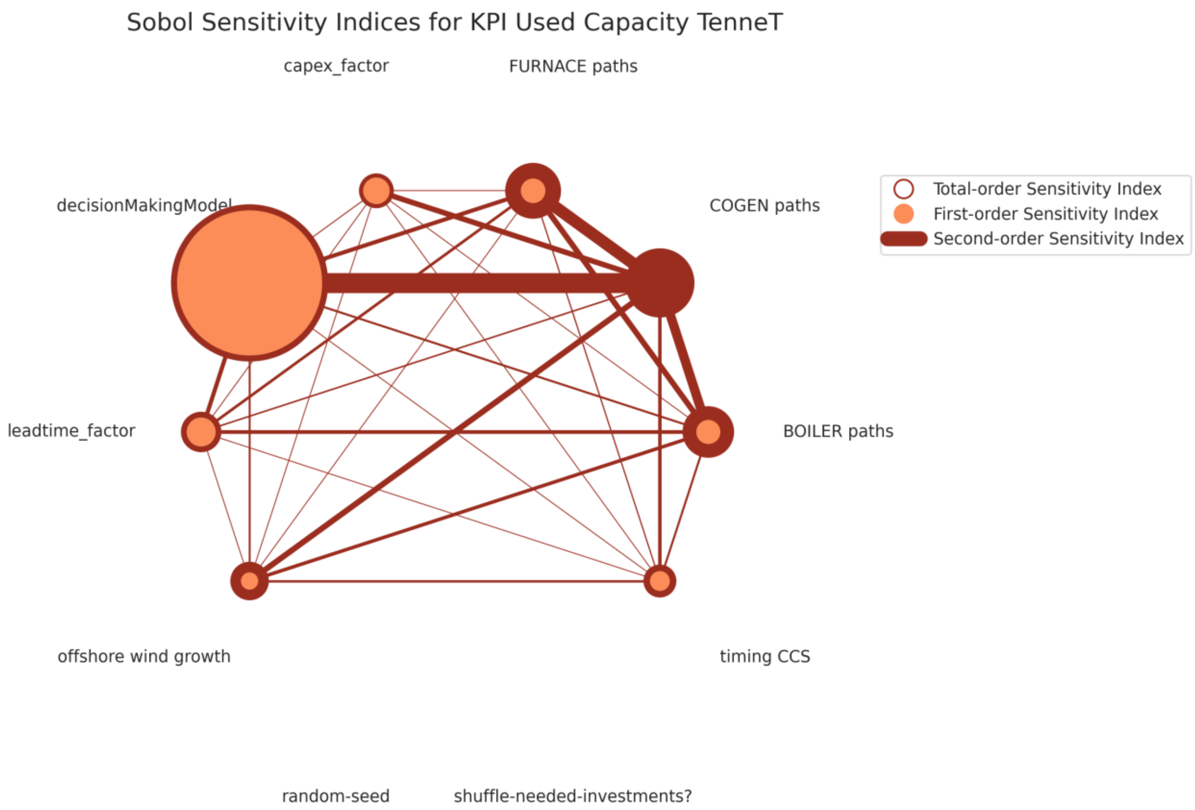
Figure 21 Heatmap for the total-order Sobol indices over time per KPI. The total-order Sobol index indicates the part of the variance in the outcome that is explained by the variance of the uncertainty, including the interaction effects with other uncertainties. For each year, the explained variance of the uncertainty on the realization of the KPI has been recalculated. No influence is indicated by a black square, while a white square indicates a strong influence. Some uncertainties only influence the variance of the concerning KPI in specific moments in time, while others remain relatively constant.

The difference between the total and first-order sensitivity index already indicates the degree of interaction between the uncertainties per KPI. To elaborate on this, we analyse the second-order sensitivity index for each pair of uncertainties. For the outcome missed events, the expectation is that these will be very low. For the two remaining KPI's, based on the difference between the total and first order sensitivity index, there is an expectation that there will be interaction effects.

To gain insight into the pairwise interactions, and the first- and total-order sensitivity indices, the relations are plotted in a diagram included in Figure 22. Indeed, we see that the second-order effects at the KPI missed events are limited. In the KPI used capacity, there is a strong interaction between the decision-making model and the paths for the cogeneration plants, which in turn interacts with the chosen paths for the furnaces and boilers.

Concerning the second-order effects of the total CAPEX, we see limited interaction, except between the decision model and the furnace paths.





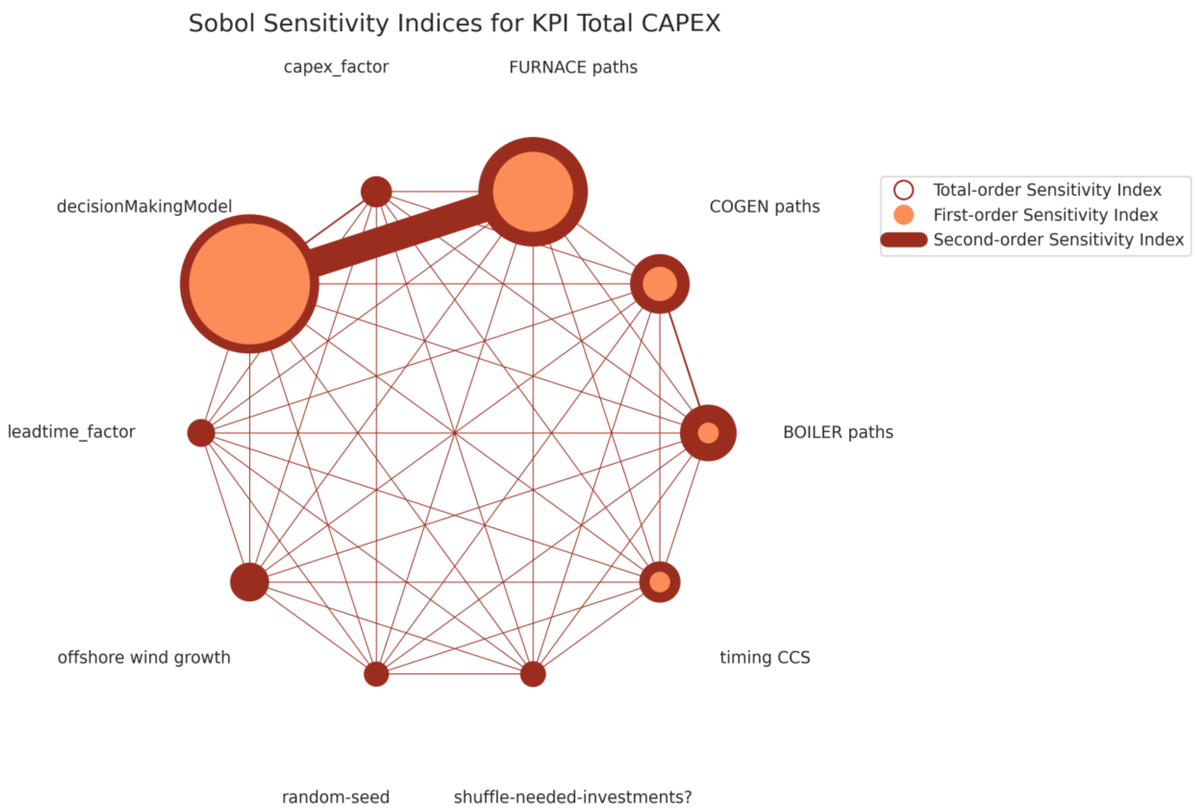
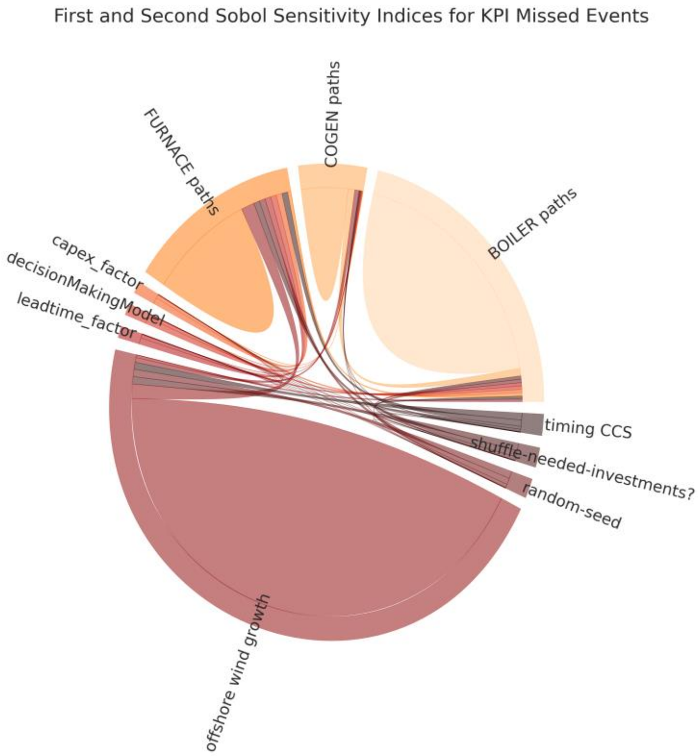
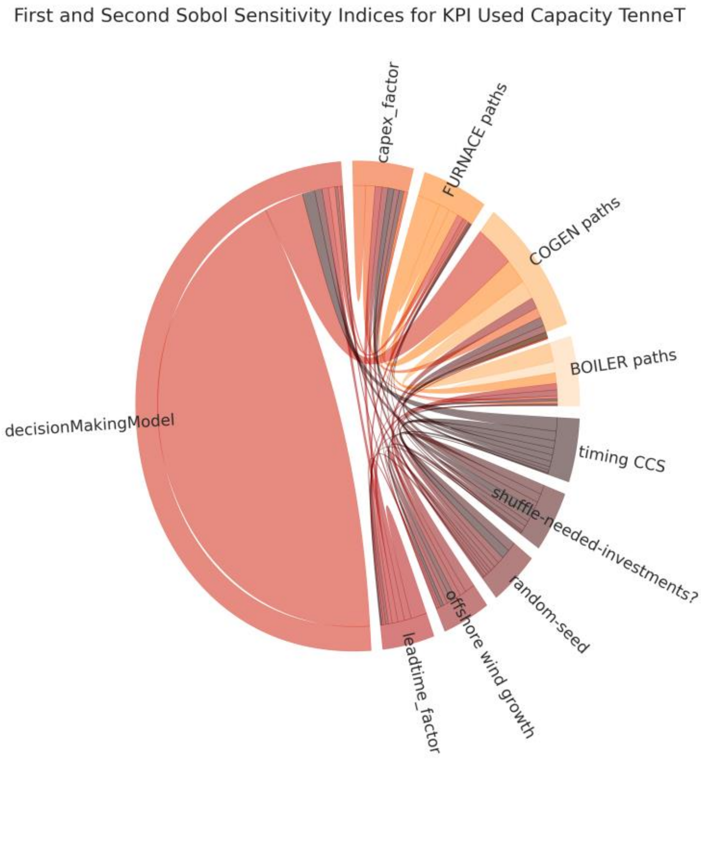


Figure 22 First, second, and total-order sensitivity indices for each KPI. The total-order SI is indicated by the outer (dark red) ring per uncertainty. The degree of (orange) padding indicates the proportion of the first-order sensitivity index. The connecting lines between the circles indicate the degree of interaction between the connected uncertainties: ticker means a higher degree of interaction. Only uncertainties with a minimum total-order sensitivity index of 0.01 are included.

To visualize the main effect and the composition of the interaction effects of the various uncertainties, we use a chord diagram in Figure 23. The connection with itself represents the main effect, while the connection with other uncertainties represents the second-order interaction. The size of the uncertainty indicates the total effect compared to other uncertainties.





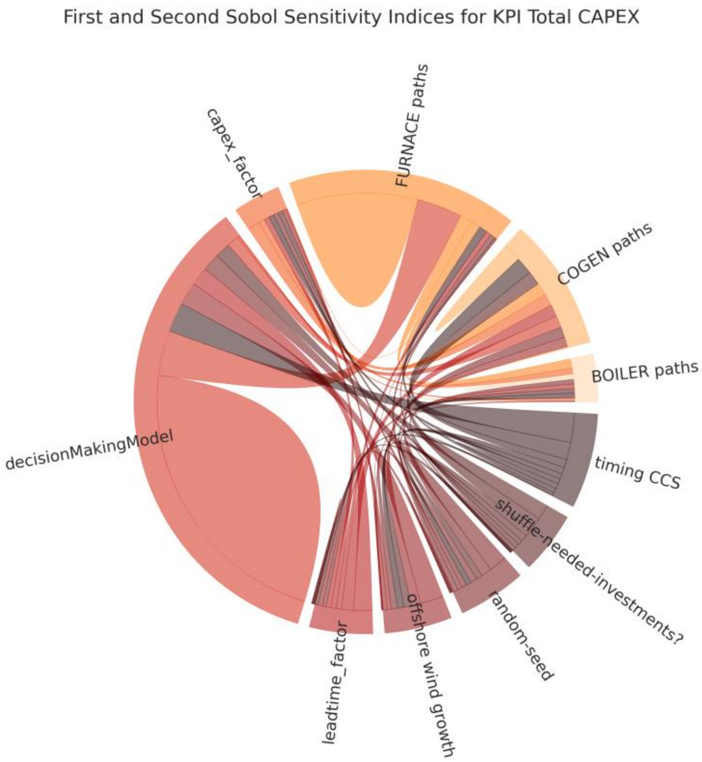


Figure 23 First- and second-order sensitivity indices for each KPI visualized in a chord diagram. The chords between the uncertainties indicate the degree of interaction, while a chord connecting an arc with itself indicates the first-order sensitivity index. The size of the arcs indicates the sensitivity of the KPI on this uncertainty and consists of the sum of the first- and second-order sensitivity indices. It may be considered an approximation of the total-order sensitivity index.

Investments

We now focus on analysing the Sobol scores on the chosen investments. For this analysis, we take into account the development of the different investments over time, categorized into five groups. The general development is shown in Figure 24. Especially in the 150 kV category, many investments are made.

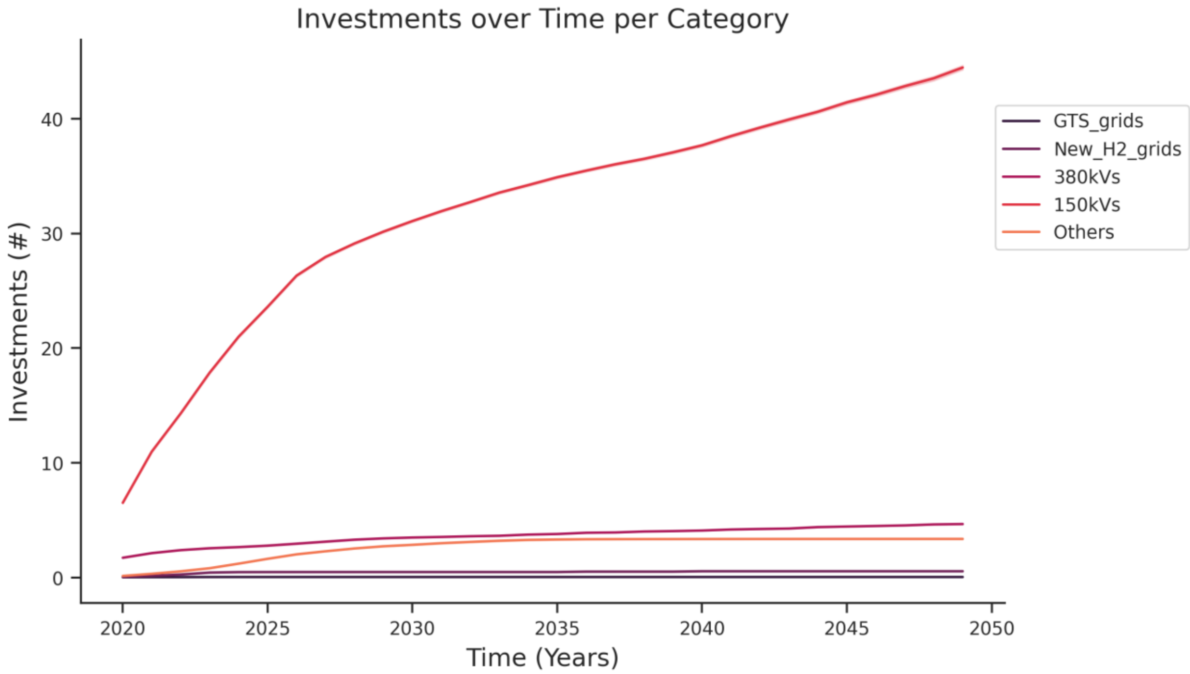


Figure 24 Investments over time. The different lines reflect the cumulative number of investments in different categories.

In order to get a sense of which uncertainties mainly influence the choice for certain (categories of) investments, we perform a Sobol analysis. The first- and total-order indices are shown in Figure 25. What is striking is that a large reliability interval is shown for furnace and boiler paths in the GTS grids investments. It is interesting to see how strongly the influence of uncertainties on the different categories differs. The decision strategy is strongly determinant in all categories. We focus on investments in new H2 grids and expansion of the 380 kV network to limit the number of analysis displayed. The analyses on the other categories of investments are included in appendix B.

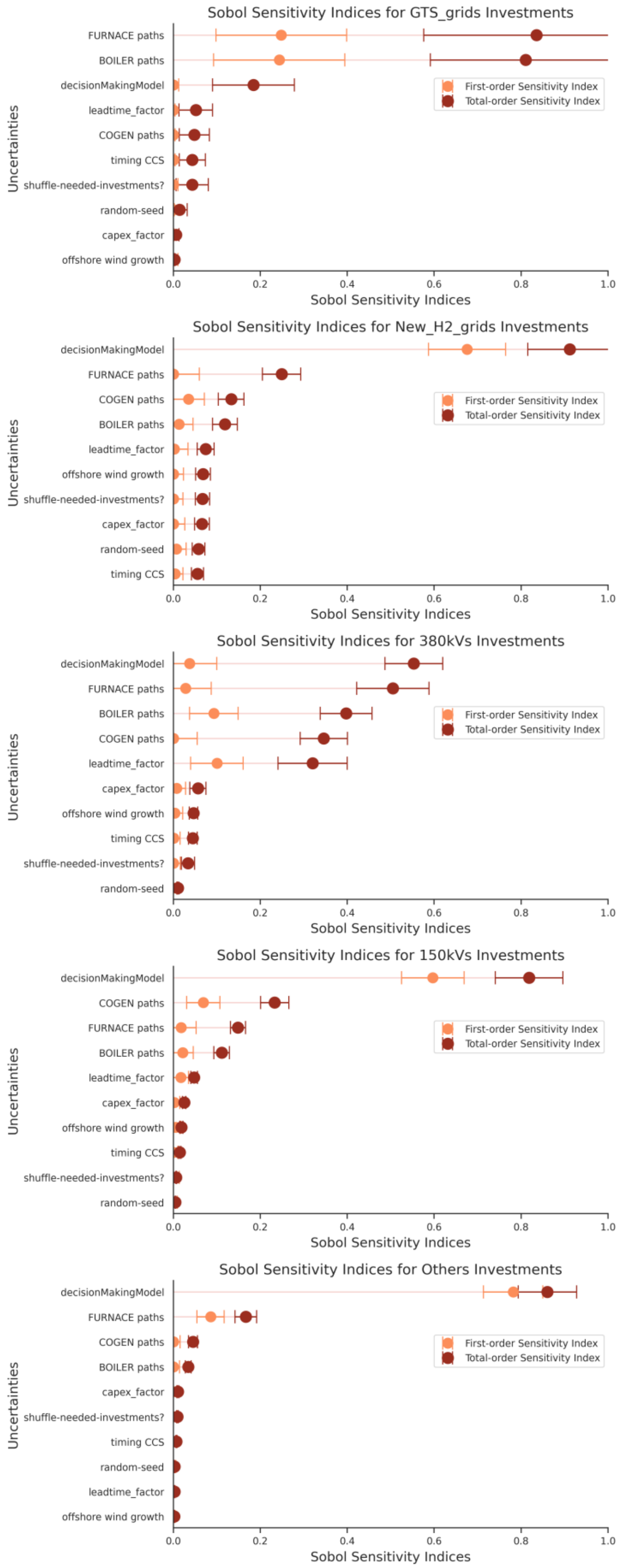


Figure 25 Sobol first- and total-order sensitivity indices for the different categories of investment alternatives. The total-order sensitivity index indicates the total contribution to explained variance, thus

includes the first-, second-, third-, and higher-order sensitivity indices. The first-order sensitivity index indicates the main effect of the uncertainty on the variance of the various investment categories. The further apart these two indices are, the higher the indirect (higher-order) interaction effects with other variables. For each index, the confidence interval is indicated by an error bar.

Again, we look at the changing influence of uncertainties over time. The heatmap with the total-order scores is shown in Figure 26. Especially with the new H2 grids, we see a strongly varying influence of the various uncertainties. From the start, there is a strong influence of the chosen technologies and the growth of offshore wind energy.

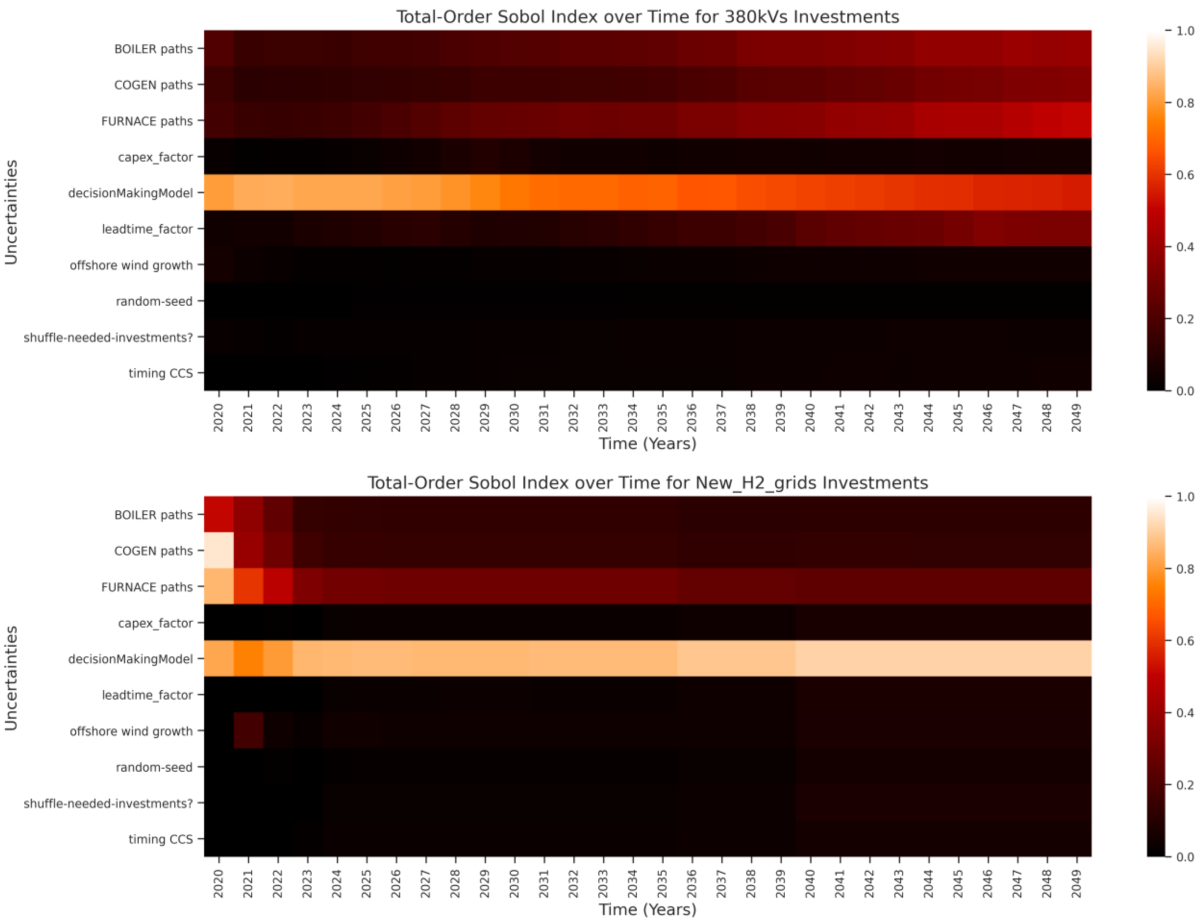
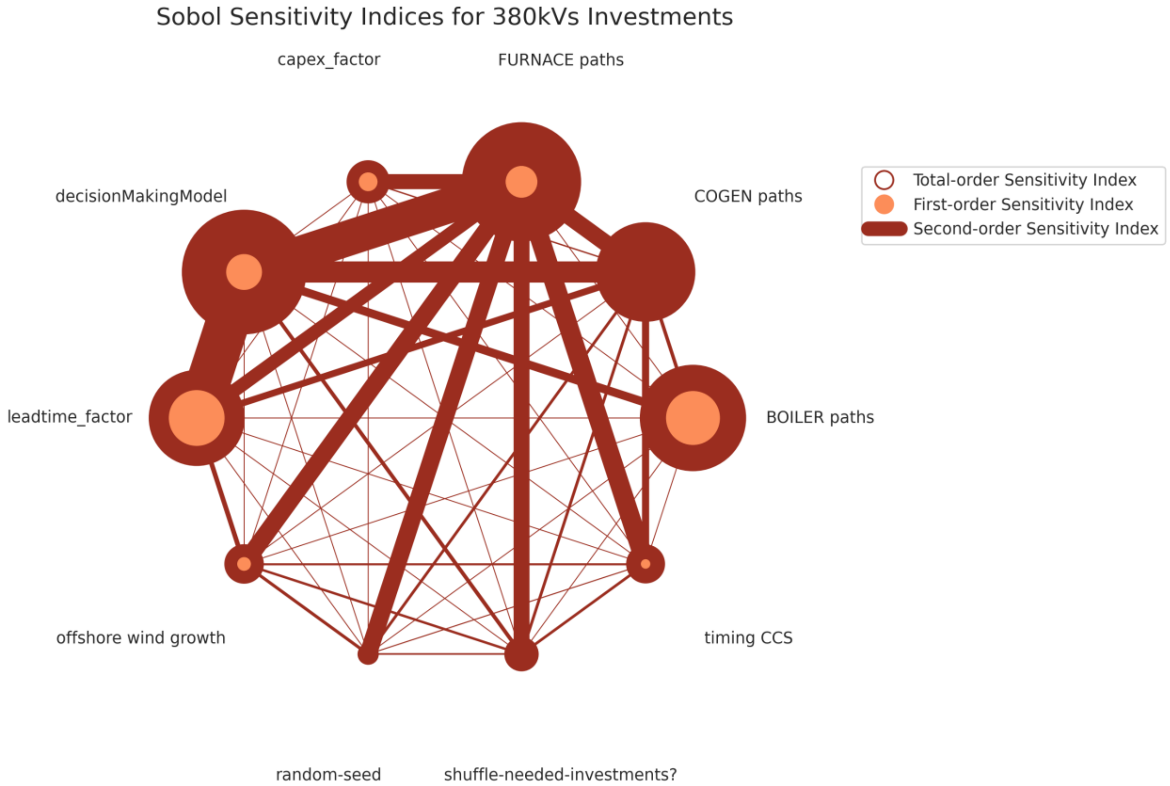


Figure 26 Heatmap of the total-order Sobol indices over time. The total-order Sobol index indicates the part of the variance in the amount of realized investment that is explained by the variance of the uncertainty, including the interaction effects with other uncertainties. For each year, the explained variance of the uncertainty on the realization of the investment category has been recalculated. No influence is indicated by a black square, while a strong influence is indicated by a white square. Some uncertainties only influence the variance of the concerning category of investments in specific moments in time, while others remain relatively constant.

To take a closer look at the composition of the effects of the uncertainties, we again use the circle plots in Figure 27. Interesting is that it is mainly indirect effects that seem to influence the choice of investments in the 380 kV segment. New hydrogen infrastructure is mainly driven by a combination of direct influence from the decision strategy, combined with the choice of furnace paths. The furnace paths almost exclusively have a total effect, consisting of interactions with many of the other uncertainties. This is also shown in Figure 28. Whereas for the choice of new hydrogen infrastructure, there is a significant relationship between the decision model and itself, in the expansions in the 380 kV network, we hardly see any relationships with themselves.



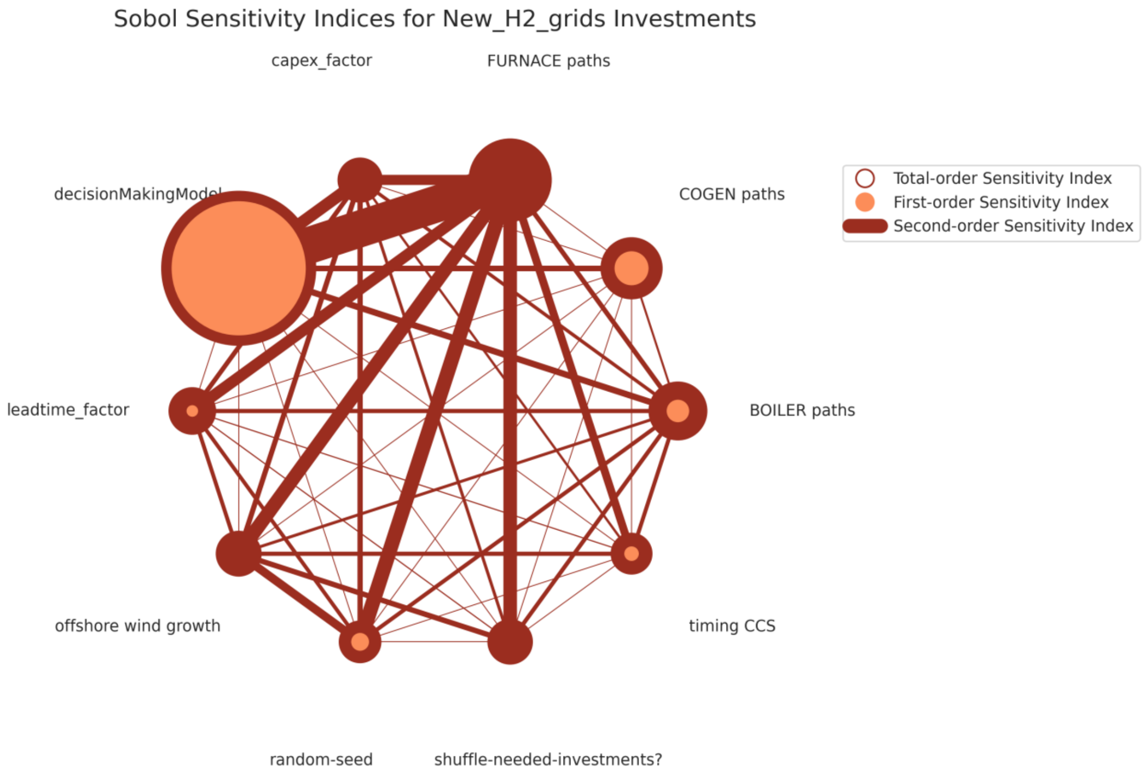
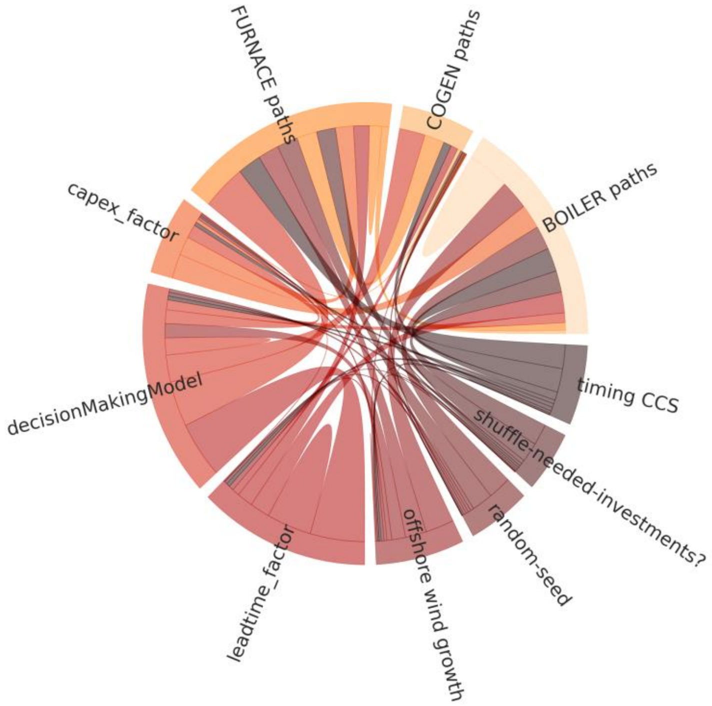


Figure 27 First, second, and total sensitivity indices for investments in the 380 kV segment and expansions in the amount of H_2 grids. The total-order SI is indicated by the outer (dark red) ring per uncertainty. The degree of (orange) padding indicates the proportion of the first-order sensitivity index. The connecting lines between the circles indicate the degree of interaction between the connected uncertainties: ticker means a higher degree of interaction. Only uncertainties with a minimum total-order sensitivity index of 0.01 are included.

First and Second Sobol Sensitivity Indices for 380kVs Investments



First and Second Sobol Sensitivity Indices for New_H2_grids Investments

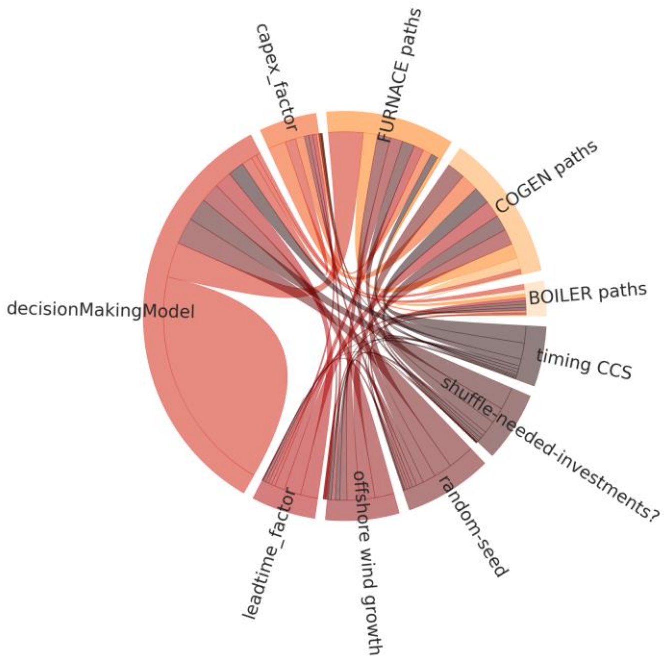


Figure 28 **First- and second-order sensitivity indices for investments in the 380 kV segment and expansions in the amount of H_2 grids, visualized in a chord diagram. The chords between the uncertainties indicate the degree of interaction. A chord which connects an arc with itself indicates the first-order sensitivity index. The size of the arcs indicates the sensitivity of the investment category on this uncertainty and consists of the sum of the first- and second-order sensitivity indices and may be considered an approximation of the total-order sensitivity index.**

Expansions in the hydrogen network are thus mainly driven by choice for a decision strategy. The number of realised 380 kV investments over time is also influenced by the decision model but mainly by a combination of factors. To show the influence of this strategy, we show the investments in these categories for the different strategies in Figure 29.

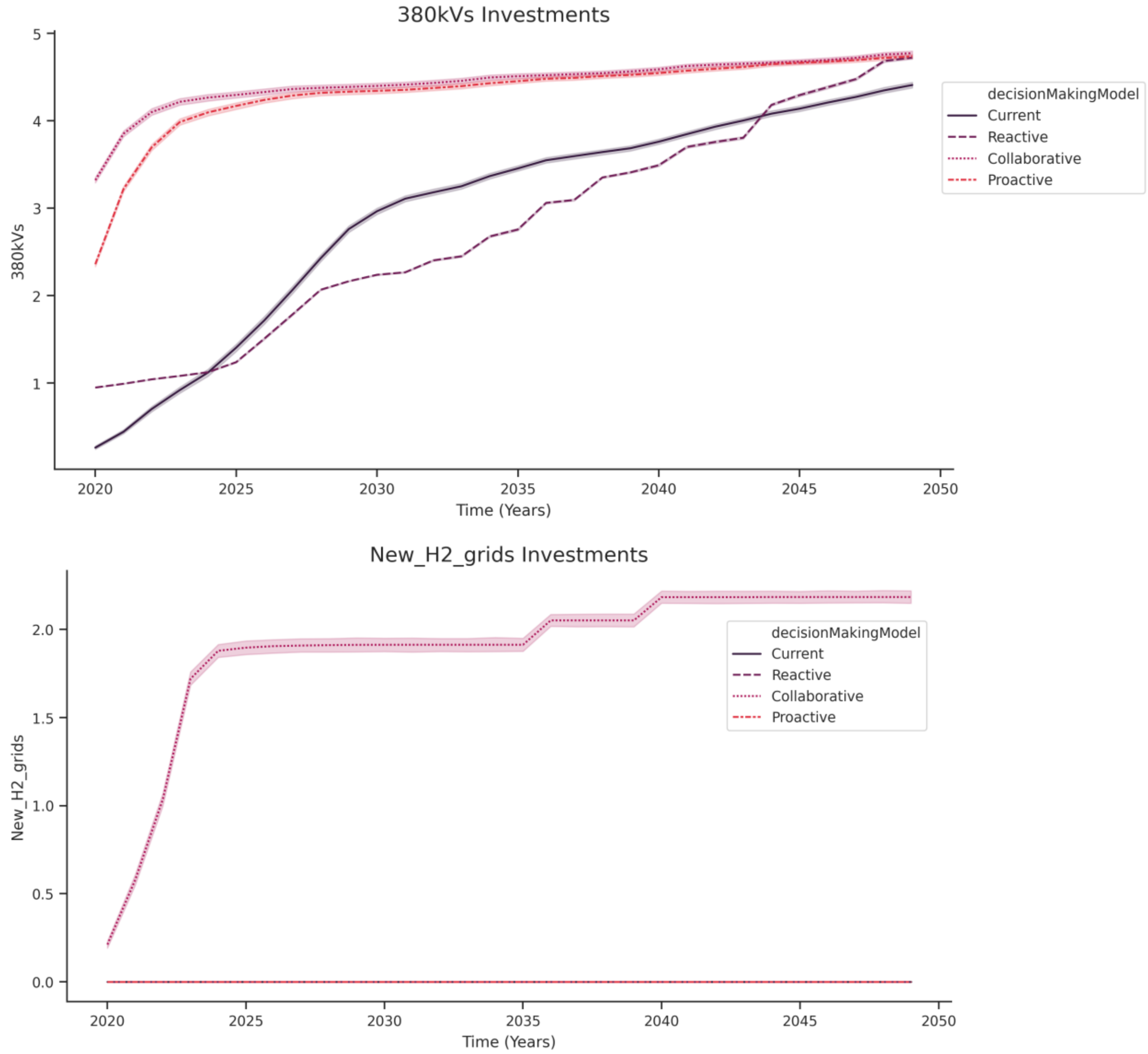


Figure 29 Development of investments in the 380 kV segment and new H_2 grids under the influence of the different applied decision-making models.

5.4.2 Main findings

The application of Sobol shows that interaction effects between the uncertainties play a role in the Windmaster model. These interactions occur with the KPI used capacity TenneT and, to a lesser extent, with total CAPEX. We see many interactions, particularly in the choice of investment alternatives. Some uncertainties, such as the choice of technologies, have virtually only interaction effects in some investment categories and no direct effect. By including the investments over time to the analysis, we gain a better understanding of the development of investments and the underlying uncertainties. We see that for new H_2 grids the decision model and furnace path technology together largely steer the choice for new hydrogen gas grids.

From the uncertainties at the interface, we conclude that the results of the model are not sensitive to permutations of the investment options and the random-seed used. The influence of manipulating development time and costs per investment is limited.

5.5 Scenario discovery: DREAM

We use the DREAM algorithm to search in the uncertainty space for areas resulting in the number of missed transition events exceeding the average value of 150. Based on the sensitivity analysis, we include the following uncertainties in the analysis. The other uncertainties, including the random-seed, are fixed so that they do not affect the model results.

- Boiler paths
- Cogeneration paths
- Decision-making model
- Furnace paths
- Offshore wind growth
- Timing CCS
- Lead time factor

5.5.1 Results

For the burn-in period, we discard the first 20% of each chain, so we look at three times 400 samples. In Figure 30, the three chains are plotted in a line plot for every uncertainty. Please note that here we use the original, raw data, to give an idea of the convergence of the individual chains, but also of the chains together. For the samples, it is noticeable that for the offshore wind growth, most samples are 0.5 or higher. In the model, this is interpreted as an increase in the supply of wind energy. For the technology pathways, the lower options seem to be chosen. With the lead time factor, we see that in the burn-in period (up to 100 samples) many samples are still being sampled at the same levels and that this fluctuates more afterwards. The univariate distribution of the raw values for the uncertainties is included in Appendix C.

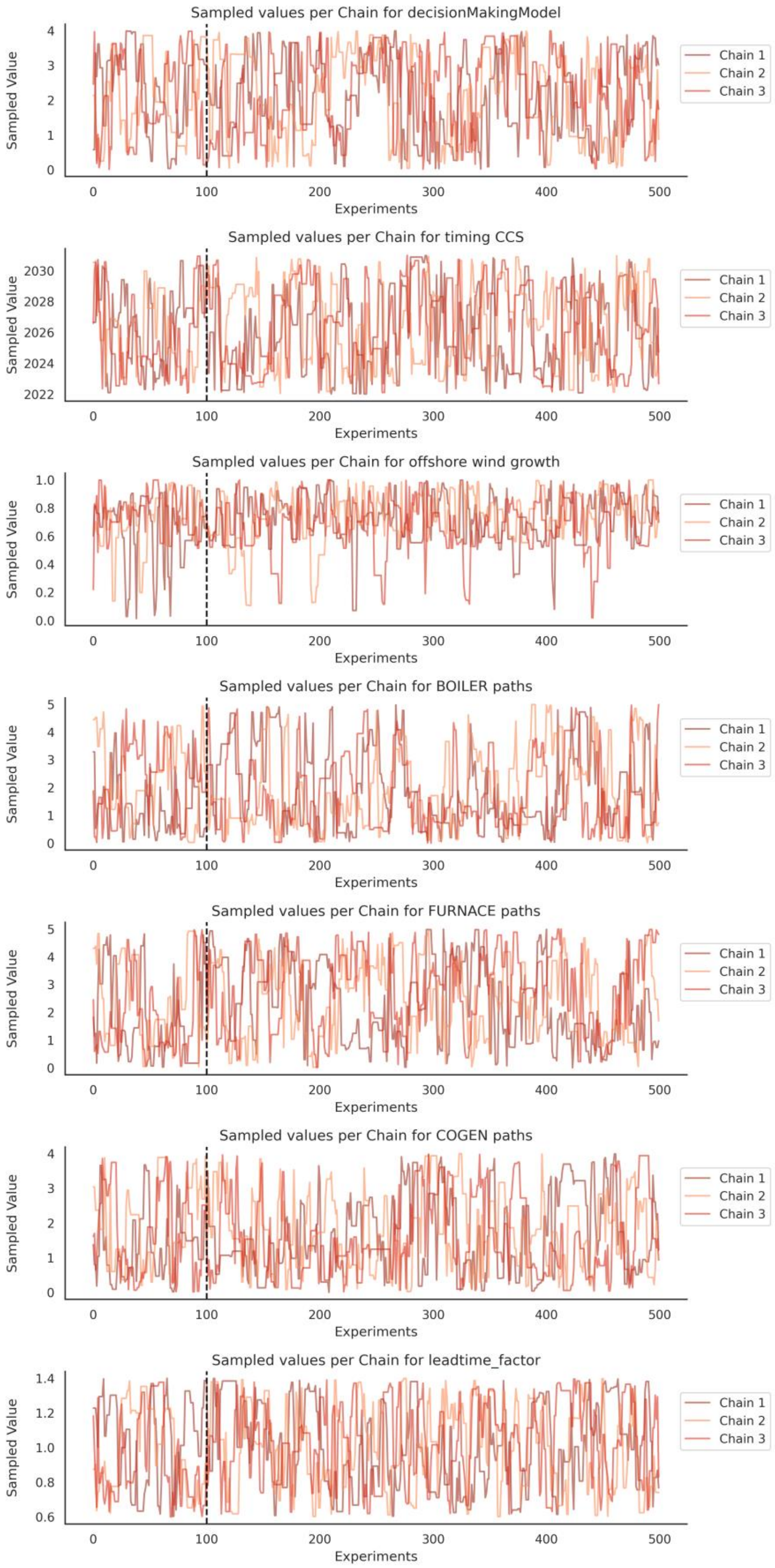


Figure 30 Sampled values per chain. Per MCMC chain, the drawn samples have been plotted. The first 20% of the points are discarded because they still depend too much on the three randomly chosen starting points. After that, the chains converge to certain areas in the domains of the different uncertainties. This is particularly visible in the case of offshore wind growth. Please note that the points have not yet been

completed and that these are only points that have been accepted by the DREAM algorithm. In reality, more samples were drawn and calculated, but mainly the points for which the outcome of the KPI was higher than 150 were actually taken into account and shown here. The points have not yet been rounded off either, as was done during the calculation of the Windmaster model.

In order to provide further insight into the sampled values, they are shown in different ways in Figure 31. We now round them off with the floor function, as is done in the model to make clear how the sampled value is factored into the outcome of the model. On the diagonal, it shows the univariate histograms for each uncertainty, with a KDE plot. This shows which values occurred mainly to generate results above 150 missed events. Some uncertainties stand out sharply. The rest of the uncertainties seem to have a limited influence. The offshore wind growth parameter has a strong tendency to be 'True'. For the boiler technology paths, the first two technologies seem to contribute strongly to many missed events, just like the first two technologies of the cogeneration paths. In terms of timing CCS, the earlier and later years 2022 and 2029 seem to have been sampled. For the lead time factor, values 1 and 1.4 appear to be relatively common. The investment strategy and the FURNACE paths appear to have virtually no influence.

By also visualizing the uncertainties in pairs, we gain into which combinations of uncertainties influence the exceeding of the average number of missed events. This is shown in the lower triangle by a KDE plot that indicates where a higher concentration of points is. On the upper triangle, this is visualized by a normal scatterplot with a high degree of transparency. The more points, the stronger this point is shown.

Looking for example, at the cogeneration and boiler paths, the choice for the first boiler path technology and the second cogeneration path technology seems to be decisive. For the cogeneration paths, the interactions with the lead time factor and the decision-making model are also significant, since these areas, in particular, are darker coloured.

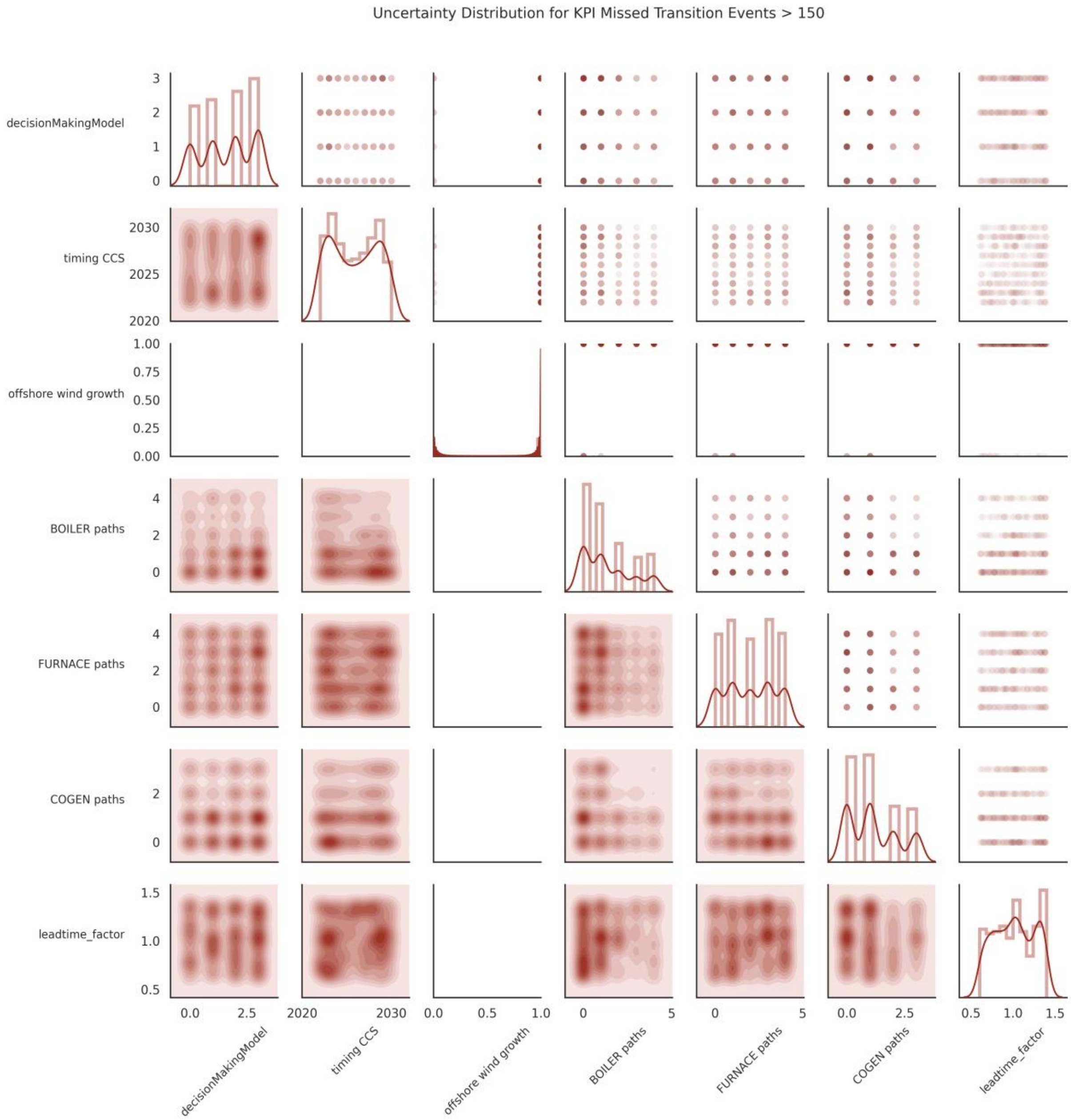


Figure 31 Sampled values during the DREAM algorithm, rounded using the mathematical floor function. On the diagonal, we see the histogram with a normalized distribution indication. The uncertainties that are more or less uniformly distributed, we conclude that these uncertainties do not negatively influence the results. Where there are one or more bins that stand out, this is, however, the case. In these bins, there is a higher density of points. This means that the number of missed events for these bins is higher than others. In the lower and upper triangle, the density of the pairwise distribution is indicated. Darker points mean more samples, which means that here are the points that are mainly responsible for the outcome in which we are interested. Bottom left is a so-called KDE plot, which plots the density of the dots. Top right is a normal scatter plot with a high degree of transparency. More sampled points in the same spot increase the intensity of this point. Since the uncertainties are mainly categorical or integer, this is necessary to give an idea of the number of points plotted there.

To look further into the relationship in sampled uncertainties, we visualize the sampled values in a parallel coordinates plot in Figure 32. The degree of opacity indicates that many samples have been drawn in this correlation.

For example, on the left axis, we see the samples drawn for offshore wind growth. Due to the high degree of opacity from 1 (which stands for an increase in offshore wind growth), we see that many samples have been drawn here. Based on this uncertainty, there does not seem to be a specific cogeneration path that leads to a high number of missed events. If there is no growth in the supply of offshore wind, in combination with a choice for the second cogeneration path technology, we still see a high number of missed events.

With the cogeneration and boiler paths, we also see a correlation between the choice for technologies that lead to a high number of missed events. We also see this connection, albeit to a lesser extent, with the furnace paths. For the decision strategy and timing CCS, we see a specific relationship: if the second decision strategy is chosen, there is a relationship with the conversion of CCS in 2023, where for the fourth strategy there will be a high number of missed events, mainly in the year 2029. With the lead time factor, there is a less clear connection. The density seems to be higher from the previous years for timing CCS to a higher lead time factor.

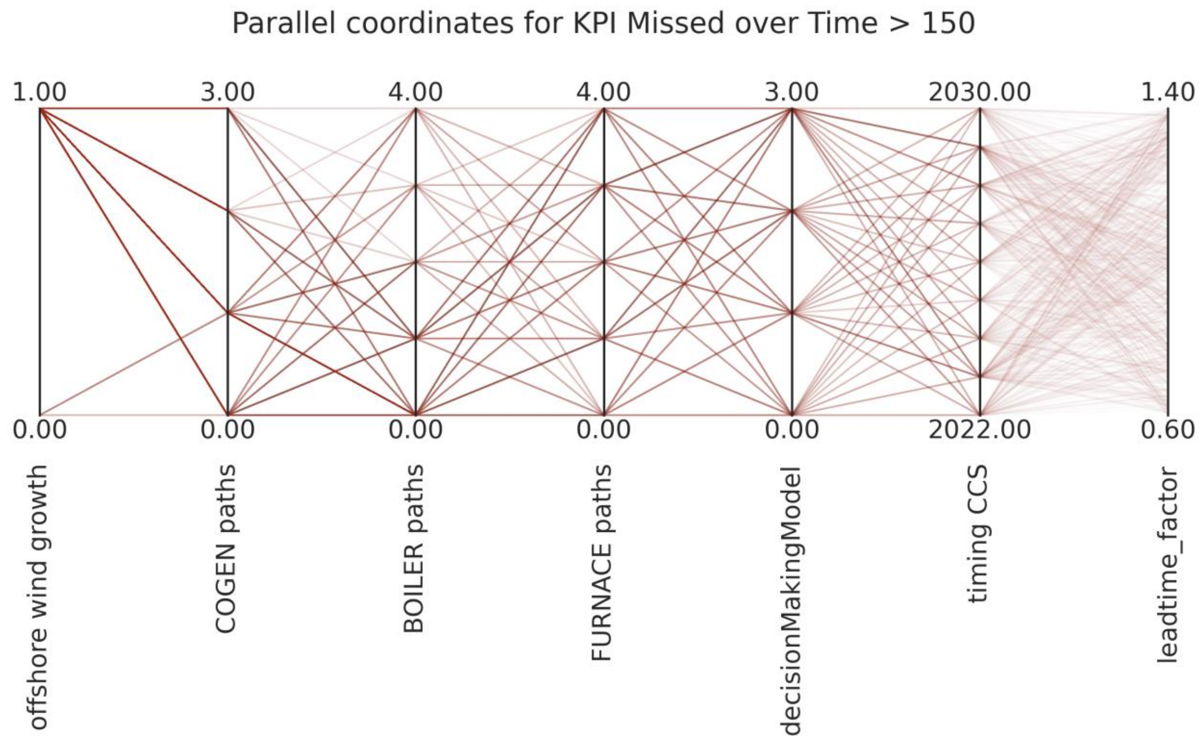


Figure 32 This plot is suitable to compare the various uncertainties and to be able to see the relationship between the points. Each uncertainty is plotted on its own axis with its own lower and upper boundary. All experiments are plotted over the different sampled values of uncertainties. A relatively high degree of transparency has been chosen, so that more frequent samples stand out better. It is striking, for example, that for offshore wind growth, many samples have been drawn with value 1 and that from there, there is little difference in the choice of a technology path for cogeneration plants. The relatively fewer samples drawn with 0 are always combined with the first or second technology option.

5.5.2 Main findings

Using the DREAM algorithm, we find the (combination of) uncertainties that lead to a higher number of missed transition events. In particular, the uncertainty surrounding offshore wind energy in combination with the various technologies leads to more missed events. Specific years in the introduction of CCS seem to emerge in combination with a lead time factor of 1 or 1.4.

6.

Uncertainty Analysis of Multi-model Ecologies

As the application on the Windmaster model shows, the methods available for uncertainty analysis on single models are suitable to be used for multi-models and to include uncertainties on the interface. This section discusses how to deal with additional uncertainties in multi-models and aims to answer the third sub-question: "How can these methods be applied to analyse a whole multi-model ecology?"

6.1 Interface

The Windmaster model is an example of an undirected graph with feedback loops over the sub-models. These loops relate to the changing context and emerging path dependencies. In the case of Windmaster, the investments proposed by the technical-economic model depend on prior investments and expected transition events. The context includes the current infrastructure, implemented energy technologies, the related supply and demand of feedstocks, and the (planned) investments. For undirected multi-models in general, it is not useful to perform uncertainty analysis on the sub-models, because in that case, we cannot take the changing context into account.

We gain insight into the sensitivity to sub-model outcomes exchanged on the interface by including them as uncertainty in the uncertainty analysis. By setting a bandwidth around these values, we determine whether small changes in the exchanged parameters will result in just small changes in the model outcome or will show significantly different outcomes. Since these are outcomes of the model components, there will always be interaction effects with other uncertainties. In the Windmaster model, for example, the CAPEX and the development time of an investment depends on the chosen decision model. These uncertainties must be analysed with a method that is able to deal with interactions and also indicates the direct and total effect on the multi-model outcomes. The second-order sensitivity index indicates which uncertainties affect the partial outcome, while the first-order sensitivity index indicates to what extent this intermediate outcome affects the multi-model results. Sobol is, therefore, more suitable for this than, for example, Extra-Trees feature scoring.

6.2 Epistemic opacity

Suppose we know nothing about the operation of the model, is it still possible to perform an uncertainty analysis? Quantitative uncertainty analysis methods are based on a black-box approach. This also becomes clear from the XPIROV framework, where the model is considered as a function of uncertainties and policy options. Therefore, an analyst can perform uncertainty analysis without being able to access the inner workings of a model. However, it is essential to take a couple of factors into account.

It is only possible to perform an accurate uncertainty analysis if a simulation model is deterministic. This means that the model returns the same results for the same given parameters. If a simulation model contains stochastic components, a model returns different results based on, for instance, the random seed used. Since the working and input to the model might be unknown, it is also possible that the reasons for the different given results are unclear for the analyst. This is harmful to the reproducibility of the research. Hence, it is advisable to run a few replications for a large number of experiments, without specifying other parameter values. If the results for the different replications are different, this indicates that results are generated based on parameters or input from other sources.

For sensitivity analysis methods, such as Extra-Trees feature scoring and Sobol global sensitivity analysis, the scores indicate the extent to which the uncertainties play a role in the difference in model outcomes. For Sobol, this is an approximation of the proportionally explained variance of an uncertainty. A high variance in the outcomes not explained by the parameter values provided could indicate that factors outside the uncertainty analysis mainly generate the model outcomes.

Another aspect is the (technical) transfer of information on the interface. It can be useful to perform permutations of the exchanged data and determine the effect on the multi-model outcomes. These permutations can lead to insight into the sensitivity of the subcomponents to, for example, small adjustments of numerical values, the transmission order of the information, or the descriptive statistic used. These uncertainties relate to the technical model implementation or the interface of the multi-model.

6.3 Computational expense

It is possible to divide the allocation of computer power between different input dimensions, for example, deep and stochastic uncertainties, and different system states. In the case of electricity networks, these system states can, for example, consist of different supply and demand states per asset per minute, hour, day, month, season, or year. It is necessary to know the added value of including more different scenarios, more stochastic influences, and more system states. To achieve a proper trade-off, an analyst can explore both deep and stochastic uncertainty spaces, using a global All-At-a-Time sampling method for deep uncertainties, and using an amount of fixed, random draws for stochastic uncertainties.

To determine a reasonable number of samples, an analyst can look at smaller subsets of the samples and repeat the analysis to analyse the convergence in terms of absolute score and bandwidth around this score. Note that the convergence of the sensitivity indices can only be determined based on an experiment set, in which the sampled values do not correlate with each other. This is the case with, for instance, a Monte Carlo sampling technique. The replications can initially be averaged in the analysis for each experiment. Then, the analysis can be repeated for the individual replications. This gives insight into the sensitivity to the stochastic uncertainties of the model. By exploring the added value of adding more samples to the analysis, a profound choice can be made for the distribution of computer power. It may seem contradictory to have to do more samples in order to conclude that one could have sufficed with fewer samples. However, this analysis should be seen as a first step in the exploration of the dimensions of uncertainty. If it turns out that it is possible to draw a reliable conclusion based on, for instance, 3,000 experiments and ten or twenty replications, this knowledge can then be used to calculate the model for different system states. It might even have been more convenient first to run the experiments on a limited number of replications to determine the number of experiments needed and then repeat the experiment set for additional replications.

To decrease the number of required experiments, we can lower the uncertainty space dimensionality. Based on a relatively low number of runs, a factor fixing technique can be applied. Feature screening reduces the dimensionality for possible subsequent analyses, such as a Sobol global sensitivity analysis, calibration, or scenario discovery method, without compromising the original variability of the model outcomes. For these subsequent analyses, the number of required samples is typically strongly related to the number of uncertainties included.

7.

CONCLUSION

Large-scale complex systems are increasingly part of our society. Simulation models can be used to get a better understanding of these systems. Since such systems are often too complex to analyse from the view of just one modelling paradigm, there is a growing need for integrating multiple modelling paradigms. Multiple models with different modelling formalisms and focal points can be connected in multi-model ecologies. Developing a multi-model ecology involves many challenges. One of these challenges is the validity of such models. Because the outcomes of models within the multi-model ecology are used and permutated in other models, there arise cascades and loops of variability, uncertainty and noise propagation, which are hard to identify, trace and consider. To investigate how uncertainty propagation in multi-models can be analysed, the following research question is defined:

"To what extent can we apply existing uncertainty analysis methods to multi-models?"

Based on the following sub-questions, an attempt is made to answer this research question.

1. What additional sources of uncertainties exist in multi-model ecologies in comparison to single models?

We describe uncertainty as *"a lack of precise knowledge as to what the truth is, whether qualitative or quantitative"* (National Research Council, 1994, p. 161). Uncertainty has three dimensions: location, level, and nature. Location refers to which part of the model the uncertainty occurs. We distinguish five locations for uncertainties: the conceptual model, the computer model, input data, the technical model implementation, and the processed output data. The level indicates the degree of uncertainty. We use two extreme levels. On one side, we have deterministic knowledge where there is no uncertainty. On the other side, we have total ignorance. In this level, there are uncertainties that we cannot say anything about because we do not know that we do not know them. Between these two extreme forms, we define five ascending levels, making it increasingly difficult to determine the probability of future scenarios. Nature indicates whether the uncertainty is caused by variability in the real system (ontic) or by lack of knowledge (epistemic).

We distinguish one additional location for uncertainty in multi-model ecologies: the model interface. Here, the exchange of parameters between submodels takes place without the intervention of an analyst or modeller. Uncertainty can quickly increase here, especially if there are many run-time interactions. Including uncertainties on this location in the analysis shows how sensitive model results are to small changes in the exchanged parameter values, the order in which this takes place, or the descriptive statistic used. The multi-model uncertainty matrix can be used to describe and assess uncertainty in a multi-model across the three dimensions.

Various aspects of multi-models contribute to uncertainty in multi-models or make it harder to analyse uncertainties. Epistemic opacity and computational expense limit the analyst's knowledge of the simulation model and set limits on the feasibility of extensive uncertainty analyses.

2. What methods for uncertainty analysis exist to analyse uncertainty propagation in single-models?

A variety of methods in the domain of uncertainty analysis is available. We distinguish methods for sensitivity analysis and calibration. Sensitivity analysis quantifies the contribution of specific uncertainties to uncertainty in the model outcomes. They are used for the identification of (non-) influential uncertainties and factor mapping. Each method has its application, limitations, and assumptions about the input and output variables. Two well-known methods for sensitivity analysis are extra-trees feature scoring and Sobol. Extra-trees feature scoring uses a set of decision trees to assign scores based on the contribution that the various uncertainties make to predicting the outcomes. Sobol determines which part of the variance in outcomes is determined by variance in the various uncertainties, distinguishing the direct effect and total effect.

Calibration methods are based on the concept of equifinality, which means that several parameter combinations result in the same outcomes. Frequently used methods use a Markov Chain Monte Carlo approach. These methods actively search for points in the uncertainty space that result in a high likelihood, using a defined likelihood function and Monte Carlo sampling. The sequence of points creates a Markov chain. Extensive methods, such as DREAM, have been developed. These methods deploy, for instance, multiple chains at the same time, select multiple candidate points, take a different subset of uncertainties with each step, dynamically adjust the step size, and exchange information between the different chains.

3. How can these methods be applied to analyse a whole multi-model ecology?

The way to perform an uncertainty analysis is related to the configuration of the multi-model. We distinguish two archetypes of multi-models with a different network structure: directed and undirected graphs. With directed graphs, there are no run-time interactions between the sub-models. The uncertainty analysis can be carried out on both the individual models and the multi-model as a whole. The findings of both analyses should lead to similar results. In the case of undirected graphs and feedback loops across the models, we only see added value in doing a sensitivity analysis on the multi-model as a whole. In doing so, we take into account the changing context and emerging path dependencies exchanged on the interface.

As proof of principle, we applied extra-trees feature scoring, Sobol, and DREAM on a case study multi-model. The application of these methods to the Windmaster model shows that it is possible to perform an uncertainty analysis on a multi-model taking into account the additional uncertainty aspects.

By including the uncertainties on the interface, we can use the methods developed for single models. Furthermore, we can use applications of analysis methods to reduce computational costs. This can be done by determining the added value of adding more samples by assessing the convergence of the method, reducing the dimensionality of the multi-model by factor fixing and by distinguishing between deep and stochastic uncertainties. To deal with epistemic uncertainty, multiple replications can be used without changing parameter values. If these replications give different outcomes, this indicates that the sub-model is not deterministic. Also, sensitivity methods indicate the contribution of the uncertainties to the explanation of the outcomes. If this contribution is low, other variables outside the analysis play a more significant role. By adding uncertainties regarding the technical implementation of the multi-model, the influence of communicating parameter values in different ways can be determined.

8.

Discussion

8.1 Implications

8.1.1 Implications for the use of uncertainty analysis on multi-models

The premise of this research is that uncertainty analysis on multi-models differs from the analysis on single simulation models. With the inclusion of additional uncertainties on the interface, uncertainty analysis methods can be used on multi-models as well. The methods do not have to be adapted for this.

Directed graphs

The performed uncertainty analysis on the Windmaster model concerns a multi-model as an undirected graph. The other distinct form is a directed graph. The findings on undirected multi-model interfaces can also be used for this type of multi-model. Compared to regular simulation models, there are few or no additional uncertainties on the interface because there is no run-time model interaction. However, complicating aspects as epistemic opacity and computational costs can still play a role. If there is a high degree of epistemic opacity, it may be useful to perform an uncertainty analysis on the subcomponents to determine whether the model is deterministic for the included uncertainties. It may be better to perform the analysis on the multi-model as a whole to reduce computational costs. However, if the uncertainty space is high dimensional, the curse of dimensionality may play a role, and it may still be better to perform the analysis on the sub-models. Since there is no run-time interaction, the findings of both analyses should lead to similar results.

Undirected graphs

The Windmaster model is an example of a tightly coupled multi-model. The submodels are developed side by side, and the number of interactions per time step is high. It is not immediately possible to perform an uncertainty analysis on the sub-models, because the model steps depend on each other's input. This makes it relatively easy to ensure that the context is the same in both models. In addition, the multi-model is designed to calculate the uncertainties based on the logic of the EMA workbench.

If multi-models are loosely coupled and not designed using the EMA workbench logic, it can be more challenging to perform an uncertainty analysis. However, by approaching the whole multi-model as a function based on uncertainties and policy options, the findings of the analysis on the Windmaster model should also be applicable to other undirected multi-models. We already see this in the number of applications using the EMA workbench or another implementation of the methods.

Identifying uncertainties on the interface will be an essential part of uncertainty analyses. The methods are based on a black-box approach, making the methods model-independent. The notion of epistemic opacity and computational costs can also be applied to these multi-models. The described approach to deal with these additional aspects of multi-models can also be used.

Sampling

There are two main approaches in sampling techniques: dependent and independent sampling. Although the choice for one or the other depends on the intended purpose, it may be useful in the context of multi-modelling to look at independent sampling techniques and how they can be applied. In contrast to dependent sampling techniques, independent sampling techniques actively search for areas in the outcome space that meet predefined requirements. Markov Chain Monte Carlo applications show that the number of required model evaluations can be significantly lower than when using dependent sampling techniques.

Since calibration approaches are based on specific target functions, it might be possible to use them for more than just calibration, being their primarily use nowadays. Our proposition for a different application is scenario discovery. Independent sampling uses a large number of experiments, of which usually only a small part contains results matching the predefined requirement. MCMC can be used to search directly for subspaces in the input space that generate these results. Because the sampling method lingers in these areas, this ensures more samples are being performed there. Therefore, it is possible to make a statement about the driving uncertainties in these sub-areas, based on more samples, without wasting computer power on samples in areas that are not of interest. However, more research is needed in this field.

8.1.2 Implications for the use of uncertainty analysis in policy decision support

Based on the uncertainty analysis and the obtained insights, what can we learn from these analyses? How can the results be used in a broader context?

From sensitivity analysis, we learn which uncertainties are primarily responsible for the variation in possible future states of the system. By linking these uncertainties to observations of the model outcomes over time, it becomes clear at what point these uncertainties start to play an important role. Based on these insights, we can look further into the development of robust policy pathways. Uncertainty analysis can provide a start for the identification of 'trigger points' or 'adaptation tipping points' for adaptive policy paths. These trigger points indicate when system outcomes become undesirable, and when (additional) measures are needed for the outcome to remain desirable. It is, therefore, useful to identify the uncertainties most affecting this outcome and which (combinations of) uncertainties lead to this outcome to become undesirable. We obtain these insights from a sensitivity analysis and scenario discovery method.

The strength of the Windmaster model, combining an investment decision model and a technical-economic model, is the possibility to compare different strategies to investigate which investment alternatives are opted for in response to different values of the uncertainties. The combination of a technical model and a decision model can be interesting for other applications, where, based on observations, a choice has to be made between a large number of possible investments. This is valuable for developing adaption pathways to identify the actions needed to obtain a positive result.

The aggregation level of the Windmaster model and the associated uncertainty analysis is not yet suitable to provide sufficient insight in developing these adaption pathways. Regardless, it is possible to determine which general investments are influenced over time by which uncertainties. Based on these identifications, it is possible to conclude that the choice for specific investment strategies is mainly determined by the type of technologies used for the boilers, furnaces, cogeneration plants, and by the opted-for decision-making strategy. To be able to generate adaption pathways, actions need to be at a more concrete level, such as the specific location of the investment. Other than that, it is essential to study the relation between uncertainties and identifying which specific investments lead to a low number of missed events.

8.2 Limitations

8.2.1 Research setup

This research focusses on uncertainty analysis on multi-models in general. There are many different configurations of multi-models possible of which the Windmaster model is just one example. The Windmaster model is tightly coupled and relatively constraint because of a limited set of investment options, physical factors, and decision strategies. Although these may be considered good reasons to carry out an initial exploration of uncertainty analysis on multi-models, this does limit the generalisability of the results. Regardless, the theoretical approach of the two proposed different archetypes and the application of this approach to an example multi-model provides a handle on how to deal with uncertainty analysis on multi-models. The additional challenges and interface do apply to both archetypes.

The discussed approaches to uncertainty analysis on a multi-model deal with multi-models consisting of two submodels. The challenge of multi-modelling in general and uncertainty analysis in particular also lies in the multiple interactions and feedback loops across different models. This research can offer limited insight into performing uncertainty analyses on multi-models with multiple sub-models. However, it should be noted that the most common implementations of multi-models can be described by one or more combinations of the described archetypes.

8.2.2 Uncertainty analysis

Windmaster model

Regarding the uncertainty analysis of the Windmaster model, there are a number of limitations. During the sensitivity analysis, many different parameter values are applied to the model. When calculating some experiments, the model returned an error message. Since the uncertainty analysis was performed as a proof-of-principle, and the Windmaster model will be further developed, it was decided not to solve these bugs. The concerning experiments were excluded from the analysis. The error messages gave no reason to assume that the successful experiments would be affected by these bugs.

There are a few imperfections in the Windmaster model. For example, the load flow calculations for the infrastructure are not yet or not correctly calculated. Compared to the realized capacity of, for example, the TenneT grid, the load flow is low. If these are divided, the capacity used is approximately 0.2%. Although the shape and behaviour of both the load flow and the capacity are most important, it is clear that the exact numbers are not in line with expectations. It is also striking that the used investment strategy affects the realization of various investments. Although it seems plausible that hydrogen gas investments will only get off the ground with a collaborative decision strategy, it also becomes clear that with a reactive strategy, less 150 kV investments will be realized. It could well be that this says more about the exact implementation of the decision models in the Windmaster model than it does about the real system. This immediately indicates an essential challenge in modelling: the dividing line between the conclusions of the implementation of the model versus the real-world system. The knowledge of experts and other actors is often required to interpret this.

Choice of KPI's

The focus is on four specified outcomes: realized investments (#), missed events over time (#), used capacity TenneT (%), and the total CAPEX (€). Although these outcomes give a broad picture of the system, particular uncertainties may play a significant role within the system, but not on these specified outcomes. This is in line with what Saltelli et al. (2019) say about carrying out uncertainty analyses: the analysis should focus on the purpose of the model and not on the model itself.

Stochastic influence

For each experiment, ten replications were performed for stochastic uncertainties, which influence the lead time per conversion asset. The uncertainty analysis on the individual replications implied that one outcome of interest is sensitive to these stochastic components. This could entail that the number of replications carried out is not sufficient. It might be useful here to repeat the experiments for more replications or repeat the analysis for 3,000 experiments and 20 replications.

It may be useful to use multiple statistical descriptions of the outcomes in the replications to gain more insight into the influence of stochasticity on model outcomes. As an example, we mention the use of the interquartile distance. These indicate a degree of dispersion in the model results and can be used as a supplement to, for example, plotting model results. Scenario discovery, for example, can be used to determine whether the influence of stochasticity is more severe in certain cases than in others.

Feature fixing

Extra-Trees feature scoring allows interaction effects, but these effects cannot be calculated. If there are factors with a high total effect but a limited main effect, these factors are labelled as not influential. Whether this is a bad thing will, of course, depend on the model and analysis concerned. In the case of Windmaster, we do not expect strong second-order interaction effects from the uncertainties that have been labelled as not influential. We do see, for example, that there are strong second-order interaction effects between uncertainties that have come through the screening. In addition, we see interaction effects between the lead time factors and CAPEX factor and the decision model on the investment outcomes, while these were deemed non-influential during the screening phase. Since these three uncertainties have some degree of overlap, it is not strange that interaction effects occur here. If these uncertainties were wrongly removed from the analysis, this would still have little effect on the reliability of the uncertainty analysis. It might be useful to take these factors further into the analyses if higher-order interaction effects are likely to play a role, even though they do not score high during the feature-scoring.

During the factor fixing, the investments over time were not recorded, due to limitations in the dimensionality of the results handler of the EMA-workbench. As a result, a feature scoring over time for the different (categories of) investment could not be carried out in this phase. Storing the investments over time is later solved by defining separate outcomes for each time step. Therefore, the investments over time can still be included in the Sobol analysis. As the uncertainty analysis over time has been performed on the outcomes and the identified uncertainties on the investments match those of the other three KPIs, we do not expect that we missed a critical uncertainty. By still showing the effect over time for the most critical uncertainties, a nearly complete picture can be created of the influence of uncertainties on the choice for investment alternatives.

The factor fixing is based on uncertainties that did not have a high feature score. This choice could be better substantiated by introducing a dummy parameter. This dummy does not influence anything in the model but can be assigned (low) scores by the methods. After all, the scores are an approach based on experiments and not analytically calculated scores. Based on the assigned score for the dummy variable, a cut-off point can be determined. In addition, the dummy score may give a better insight into the number of samples that need to be included.

Sobol

For the calculation of Sobol sensitivity indices, some SI's turned out to be slightly negative or come with a high confidence interval. This could indicate that more model evaluations need to be performed or that the primary conditions for a Sobol sensitivity analysis are not met: the Sobol method assumes a unimodal, symmetrically shaped distribution of the outcome. However, based on the visualization of the distribution of the results, we state that this basic condition has been met. However, it may help to choose a moment-independent method like PAWN, as it is not based on assumptions about the form of output distribution.

8.3 Future Research

Based on the limitations and findings of this research, we make some recommendations for future research. For a single case study model is used, we recommend to apply uncertainty analysis on multi-models with different network structures and a higher number of sub-components. The influence of uncertainties and feedback mechanisms on the interface can then be further investigated in different (combinations of) network types. The role of epistemic opacity can also be further investigated, especially when multi-models are developed in cross-institutional collaborations. The limitations that computational costs impose and how they can be reduced can also be further investigated by including more dimensions of uncertainty into the analysis.

Finally, we recommend doing more research on how other methods for uncertainty analysis can be applied to multi-models. Especially the use of moment-independent sensitivity analysis add value since multi-modal outcomes could arise due to interaction between heterogeneous models. In the context of scenario discovery, we recommend looking further into the application of MCMC methods, especially if the input space becomes highly dimensional.

9.

Recommendations

Developing multi-models and performing uncertainty analyses brings additional challenges that go far beyond what we have discussed in this research. Nevertheless, we acknowledge that the development of multi-models is interesting to perform a broader analysis of increasingly complex systems. There are certainly still many challenges and things to discover in this branch of simulation models. These will only be tackled further by trying out different combinations, implementations, and new methods. We make some recommendations for the use of uncertainty analysis in the context of multi-models.

9.1 Version Tracking

With the proliferation of involved software and hardware for multi-models, one can quickly lose the overview. Specific operating systems, versions of software, python packages, and other components of the technical model implementation may be incompatible. It is essential to apply version tracking so that bugs can be fixed quickly. For the Windmaster model, this would be:

- Python: 3.6
  - Python package NetworkX: 2.1
  - Python package JPype1: 0.7.0
- Netlogo: 6.0.4
- Java (JRE): 1.8.0

It would be even better if a container-like approach were used to ensure that all necessary applications, settings, and packages are delivered with the model, and the model is offered as a plug-and-play component.

9.2 Methods

9.2.1 Moment-independent sensitivity analysis

Some uncertainty analysis methods, such as the Sobol global sensitivity analysis, assume a specific outcome distribution. Newer approaches, such as PAWN, consider the whole distribution function and can therefore also deal with multimodal outcomes. They can deal with other sampling methods as well, so no additional experiments need to be performed. This analysis can be done in addition to Sobol. At this moment, there is no python implementation of PAWN available.

9.2.2 DREAM

This research used a DREAM implementation for scenario discovery. Although there are already examples available for the use of DREAM for calibration purposes, the use for scenario discovery has not been applied before. It may be interesting to investigate this approach further. The possibility to make use of processor cores would also make this method even more interesting.

9.3 Decision-making under Deep Uncertainty

Uncertainty analysis in itself adds little value. It is essential to interpret the results of the uncertainty analysis to make clear to policymakers what we can get from the analyses. An unambiguous answer is still often expected from science. Dealing with deep uncertainties is for many decision-makers, but also scientists and model-owners, not yet far developed. There is still a tendency to assume a minimal number of future scenarios. Although this makes the system tame in a sense, which is ideal for decision making, we must not forget that the influence of uncertainty on the real system is not affected by the assumptions a policy analyst might make in determining these scenarios. It is vital to investigate scenarios that are not driven by limited assumptions about the future development of the system. Collaboration between different actors, modellers, model-owners, analysts, and experts will be necessary to identify various future scenarios, determine the effects on the system and develop policies that will work well regardless of the future scenarios.

10.

BIBLIOGRAPHY

Abu-Hanna, A., Nannings, B., Dongelmans, D., & Hasman, A. (2010). PRIM versus CART in subgroup discovery: When patience is harmful. *Journal of Biomedical Informatics*, *43*(5), 701–708.

Agusdinata, D. B. (2008). *Exploratory modeling and analysis: A promising method to deal with deep uncertainty*. TU Delft.

Agusdinata, D. B., & DeLaurentis, D. (2008). Specification of system of systems for policymaking in the energy sector. *The Integrated Assessment Journal*, *8*(2), 1–24.

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298.

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1–4), 11–29.

Bollinger, L. A., Nikolic, I., Davis, C. B., & Dijkema, G. P. J. (2015). Multimodel Ecologies: Cultivating Model Ecosystems in Industrial Ecology. *Journal of Industrial Ecology*, *19*(2), 252–263.

Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 791.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, *77*(1), 34–49.

Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, *22*(10), 1509–1518.

Cariboni, J., Gatelli, D., Liska, R., & Saltelli, A. (2007). The role of sensitivity analysis in ecological modelling. *Ecological Modelling*, *203*(1–2), 167–182.

Carley, K. M., Morgan, G., Lanham, M., & Pfeffer, J. (2012). Multi-modeling and sociocultural complexity: Reuse and validation. *Advances in Design for Cross-Cultural Activities Part II*, (July), 128–137.

Castrejon-Campos, O., Aye, L., & Hui, F. K. P. (2020). Making policy mixes more robust: An integrative and interdisciplinary approach for clean energy transitions. *Energy Research & Social Science*, *64*(January), 101425.

Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, *59*(8), 3873–3878.

Davis, P. K., & Tolk, A. (2007). Observations on new developments in Composability and multi-resolution modeling. *Proceedings - Winter Simulation Conference*, 859–870.

Feinstein, A. H., & Cannon, H. M. (2001). Fidelity, verifiability, and validity of simulation: Constructs for evaluation. *Developments in Business Simulation and Experiential Learning*, *28*, 57–67.

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, *9*, 123–143.

Funtowicz, S. O., & Ravetz, J. R. (1990). *Uncertainty and quality in science for policy* (Vol. 15). Springer Science & Business Media.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42.

Ghanem, R., Owhadi, H., & Higdon, D. (2017). Handbook of Uncertainty Quantification. In R. Ghanem, D. Higdon, & H. Owhadi (Eds.), *Handbook of Uncertainty Quantification*. Cham: Springer International Publishing.

Herder, P. M., Bouwmans, I., Dijkema, G. P. J., Stikkelman, R. M., & Weijnen, M. P. C. (2008). Designing infrastructures using a complex systems perspective. *Journal of Design Research*, *7*(1), 17–34.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, *169*(3), 615–626.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York, NY: Springer New York.

Jaxa-Rozen, M., & Kwakkel, J. H. (2018). Tree-based ensemble methods for sensitivity analysis of environmental models: A performance comparison with Sobol and Morris techniques. *Environmental Modelling & Software*, *107*(June), 245–266.

Khorashadi Zadeh, F., Nossent, J., Sarrazin, F., Pianosi, F., Van Griensven, A., Wagener, T., & Bauwens, W. (2017). Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model. *Environmental Modelling and Software*, *91*, 210–222.

Kwakkel, J. H. (2017). The Exploratory Modeling Workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling and Software*, *96*, 239–250.

Kwakkel, J. H., Auping, W. L., & Pruyt, E. (2013). Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting and Social Change*, *80*(4), 789–800.

Kwakkel, J. H., & Jaxa-Rozen, M. (2016). Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. *Environmental Modelling and Software*, *79*, 311–321.

Kwakkel, J. H., Walker, W. E., & Marchau, V. A. W. J. (2010). Classifying and communicating uncertainties in model-based policy analysis. *International Journal of Technology, Policy and Management*, *10*(4), 299–315.

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis. In *Academy of Management Learning & Education*. Rand Corporation.

Makridakis, S., Hogarth, R. M., & Gaba, A. (2009). Forecasting and uncertainty in the economic and business world. *International Journal of Forecasting*, *25*(4), 794–812.

Mikulecky, D. C. (2001). The emergence of complexity: Science coming of age or science growing old? *Computers and Chemistry*, *25*(4), 341–348.

Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, *33*(2), 161.

National Research Council. (1994). Science and Judgment in Risk Assessment. In *Environmental Health Perspectives* (Vol. 102). Washington, D.C.: National Academies Press.

Nederlandse Emissieautoriteit. (2015). Klimaatakkoord van Parijs. Retrieved from https://www.emissieautoriteit.nl/onderwerpen/klimaatakkoord-van-parijs

Nikolic, I., Kwakkel, J. H., & Bletsis, A. (2019). *Multi- Modelling and Simulation for Large-Scale Sociotechnical Systems : Extending the Multimodel Ecology Perspective*. Delft.

Nikolic, I., Warnier, M., Kwakkel, J. H., Chappin, E., Lukszo, Z., Brazier, F., … Palensky, P. (2019). *Principles, challenges and guidelines for a multi-model ecology*. Delft.

Pace, D. K. (2015). Fidelity, Resolution, Accuracy, and Uncertainty. In M. L. Loper (Ed.), *Modeling and Simulation in the Systems Engineering Life Cycle: Core Concepts and Accompanying Lectures* (pp. 29–37). London: Springer London.

Pennock, M. J., & Gaffney, C. (2018). Managing Epistemic Uncertainty for Multimodels of Sociotechnical Systems for Decision Support. *IEEE Systems Journal*, *12*(1), 184–195.

Petersen, A. C. (2006). *Simulating nature: a philosophical study of computer-simulation uncertainties and their role in climate science and policy advice*. Amsterdam: Het Spinhuis.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232.

Pianosi, F., & Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, *67*, 1–11.

Pianosi, F., & Wagener, T. (2018). Distribution-based sensitivity analysis from a generic input-output sample. *Environmental Modelling & Software*, *108*, 197–207.

Plomp, A. j., Wetzels, W., Seebregts, A., & Kroon, P. (2013). *Verkenning voor Rotterdam Climate Initiative - CO2-emissies tot 2030*. (May 2015), 29.

Pruyt, E., & Kwakkel, J. H. (2014). Radicalization under deep uncertainty: a multi-model exploration of activism, extremism, and terrorism. *System Dynamics Review*, *30*(1–2), 1–28.

Ratto, M., Tarantola, S., & Saltelli, A. (2001). Sensitivity analysis in model calibration: GSA-GLUE approach. *Computer Physics Communications*, *136*(3), 212–224.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., … Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*(March 2018), 29–39.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., … Tarantola, S. (2008). Global Sensitivity Analysis: The Primer. In *Global Sensitivity Analysis. The Primer* (pp. 237–275). Chichester, UK: John Wiley & Sons, Ltd.

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity Analysis in Practice: a Guide to Assessing Scientific Models*. Chichester: John Wiley & Sons, Ltd.

Saltelli, A., Tarantola, S., & Chan, K. P. S. (1999). A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, *41*(1), 39–56.

Shockley, E. M., Vrugt, J. A., & Lopez, C. F. (2018). PyDREAM: High-dimensional parameter inference for biological models in python. *Bioinformatics*, *34*(4), 695–697.

Sobol, I. M. (2001). *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*. *55*, 271–280.

Taleb, N. N. (2015). *The Black Swan: The impact of the Highly Improbable*. New York: Random house.

Tarantola, S., Gatelli, D., & Mara, T. A. (2006). Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering and System Safety*, *91*(6), 717–727.

Tin Kam Ho. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278–282. IEEE Comput. Soc. Press.

Torn, K. (2019). *Windturbines Maasmond*. Port of Rotterdam Authority.

United Nations. (2018). About the Sustainable Development Goals. Retrieved from United Nations website: https://www.un.org/sustainabledevelopment/sustainable-development-goals/

van Dam, K. H., Nikolic, I., & Lukszo, Z. (2013). *Agent-Based Modelling of Socio-Technical Systems*. Dordrecht: Springer Netherlands.

Van den Hoek, R. E., Brugnach, M., Mulder, J. P. M., & Hoekstra, A. Y. (2014). Analysing the cascades of uncertainty in flood defence projects: How “not knowing enough” is related to “knowing differently.” *Global Environmental Change*, *24*(1), 373–388.

Verikas, A., Gelzinis, A., & Bacauskiene, M. (2011). Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, *44*(2), 330–349.

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*(12), 1–15.

Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., & Robinson, B. A. (2009). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, *23*(7), 1011–1026.

Walker, W. E., Harremoës, P., Rotmans, J., Van der Sluijs, J. P., Van Asselt, M. B. A., Janssen, P., & Krayer von Krauss, M. P. (2003). Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment, 4, 2003*.

Wilby, R. L., & Dessai, S. (2010). Robust adaptation to climate change. *Weather*, *65*(7), 180–185.

Wurth, T., Nikolic, I., Kwakkel, J. H., Sloot, M., Cuppen, E., & Quist, J. (2019). *Project Windmaster: De weg naar een adaptief investeringsbeleid*. Delft.

A.

Extra-Trees Feature Scoring

The first step of the analysis is to assess the distribution of the generated experiments. Since a Monte Carlo sampling was used to generate the experiments, it is important to check if the samples are evenly distributed over the whole range of each uncertainty. A pair plot of the included uncertainties is included in Figure 33 to assess the distribution. In the other figures, the distribution is given for each univariate distribution.

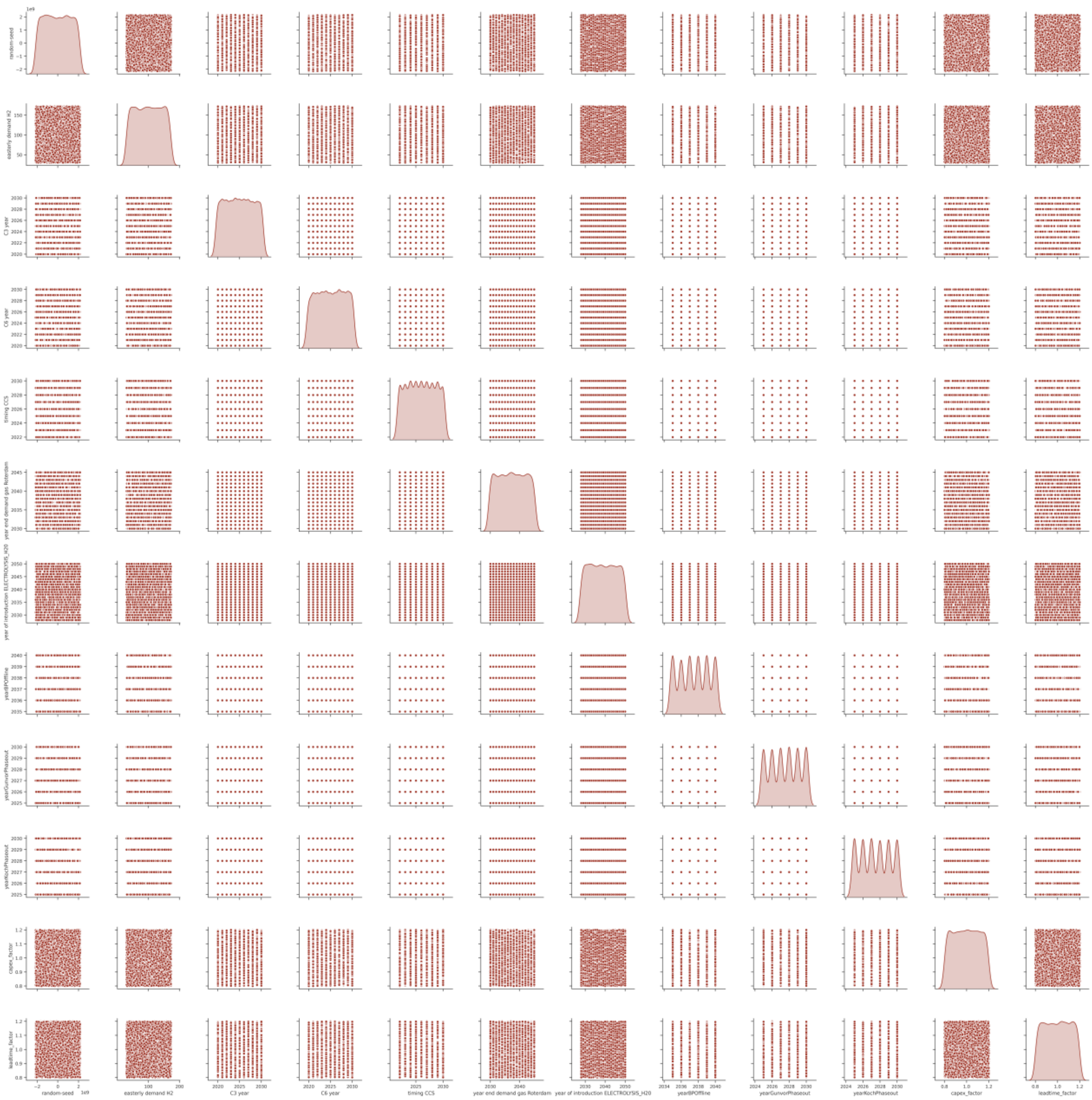


Figure 33 Pair plot of the distribution of the sampled integer, categorical, and real uncertainties. From left (and top) to right (and bottom): random-seed, easterly demand H2, C3 year, C6 year, timing CCS, year end demand gas Rotterdam, year of introduction electrolysis H2O, year BP Offline, year Gunvor

Phaseout, year Koch phaseout, CAPEX factor, and lead time factor. On the diagonal, a histogram is plotted to indicate the univariate distribution.

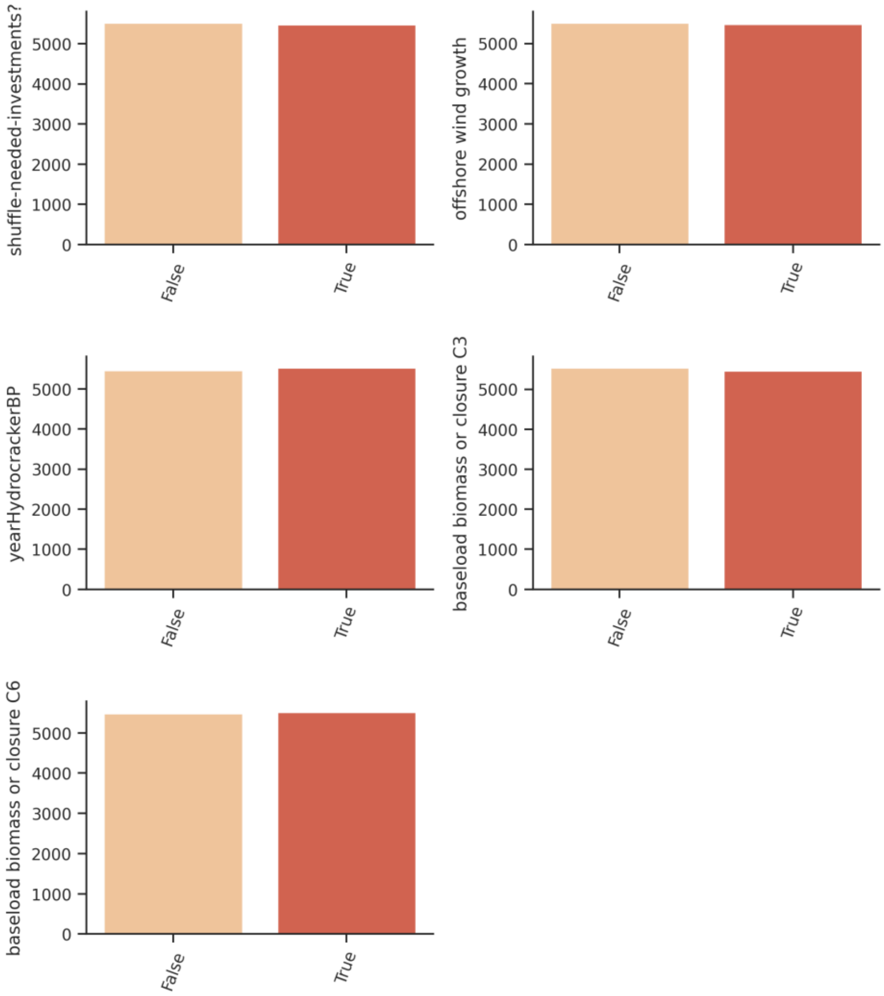


Figure 34 Distribution of samples of Boolean uncertainties

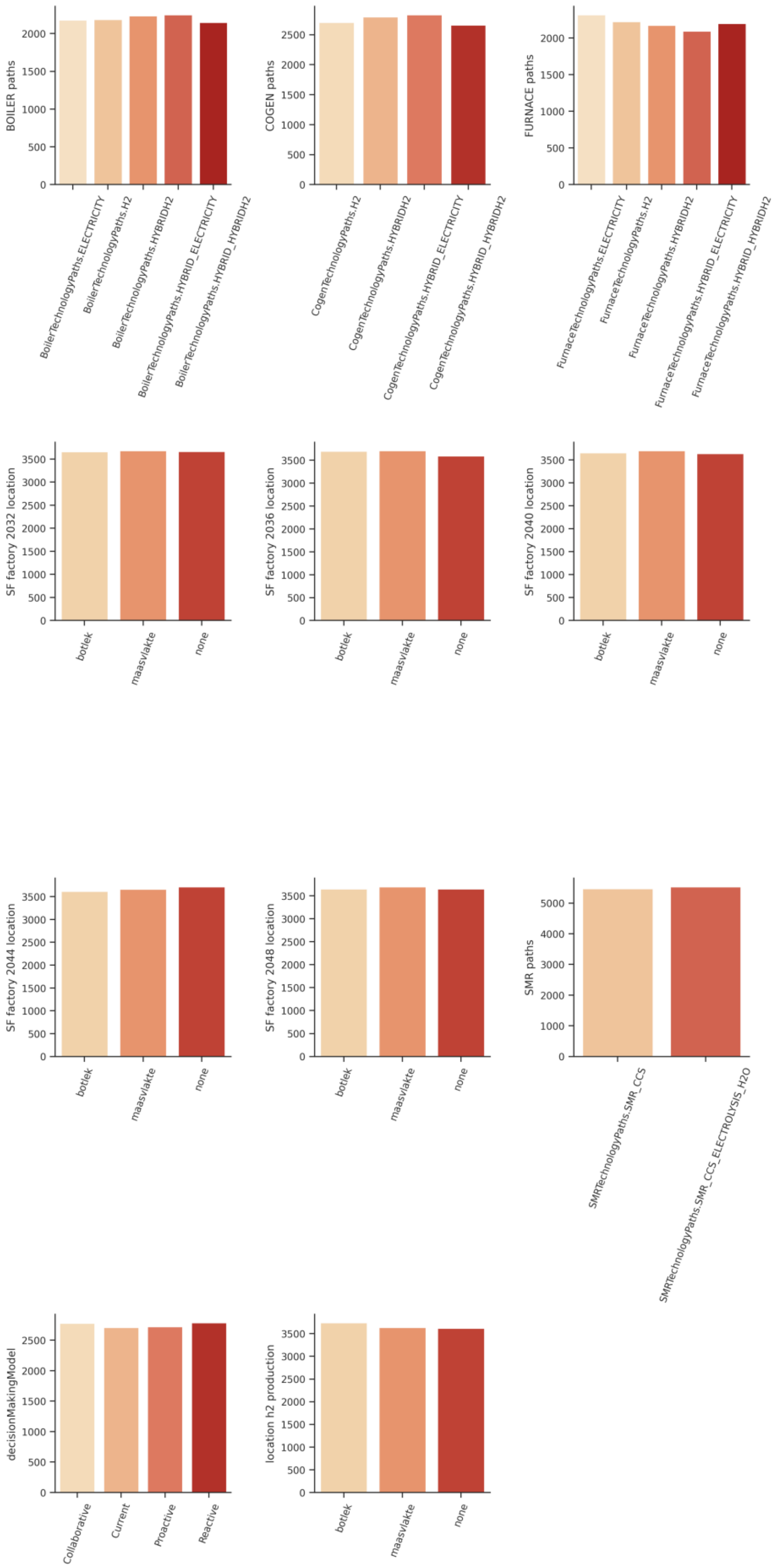


Figure 35 Distribution of samples of categorical uncertainties

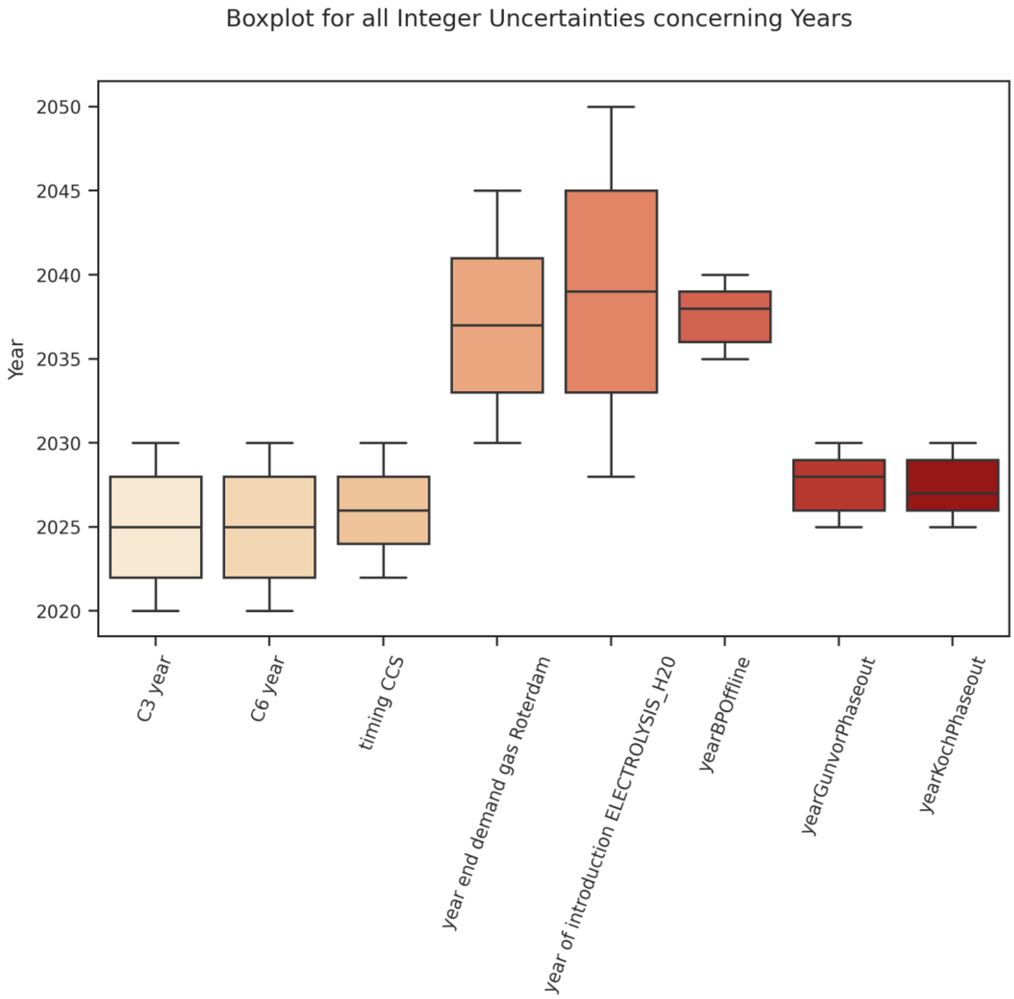


Figure 36 **Boxplot of samples of integer uncertainties**

We see that there are no values for uncertainties that have been sampled disproportionately often and therefore form a reliable sample to generate results. The next step is analysing the outcomes. It can help to look at the outcomes in relation to each other and the underlying uncertainties. Figure 37 shows the results broken down by the underlying decision strategies. Figure 38 shows the results in relation to each other. In Figure 39, we see the density over time in a KDE plot, also broken down by decision model.

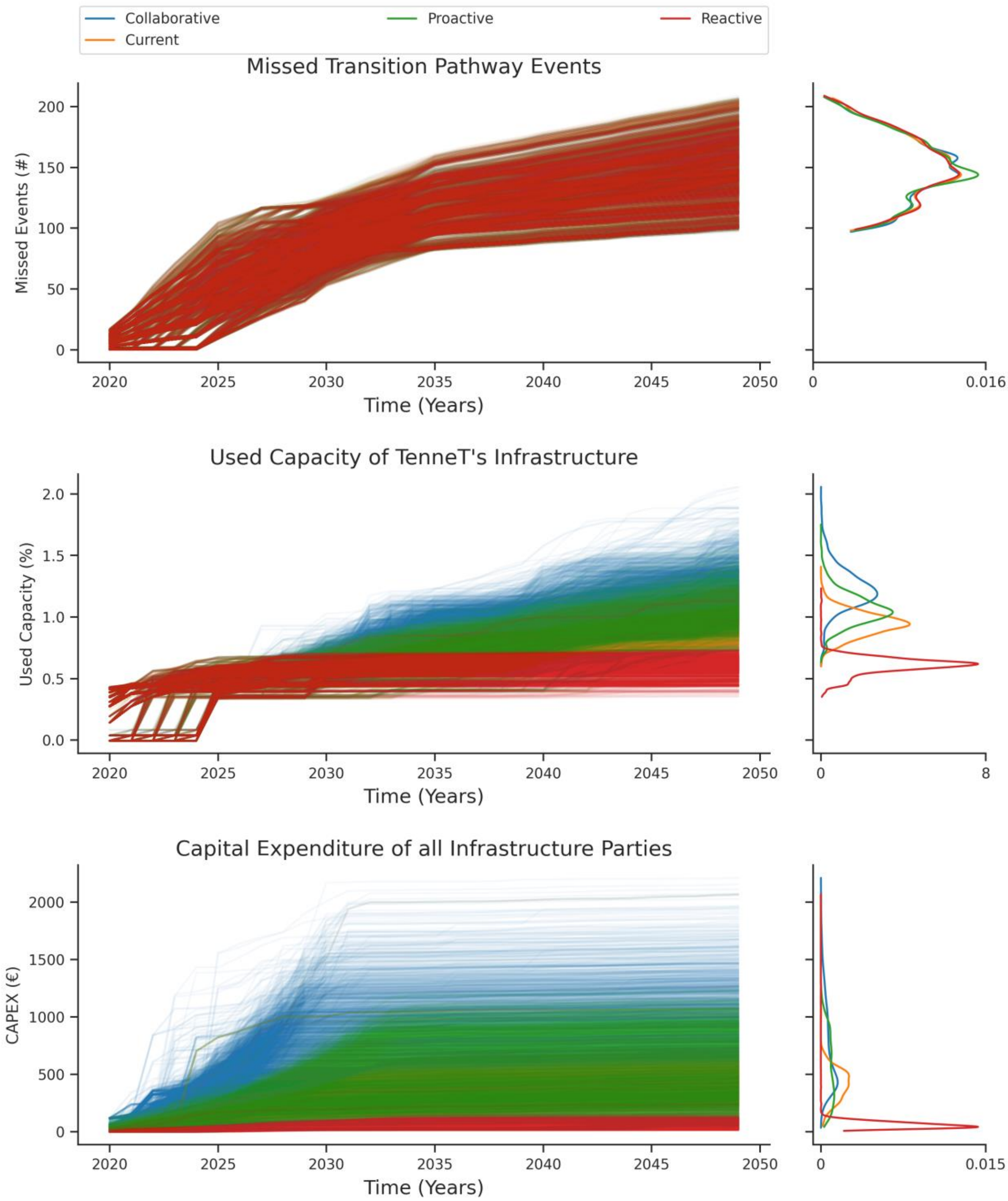


Figure 37 Line plots for realizations of KPI under different decision-making strategies

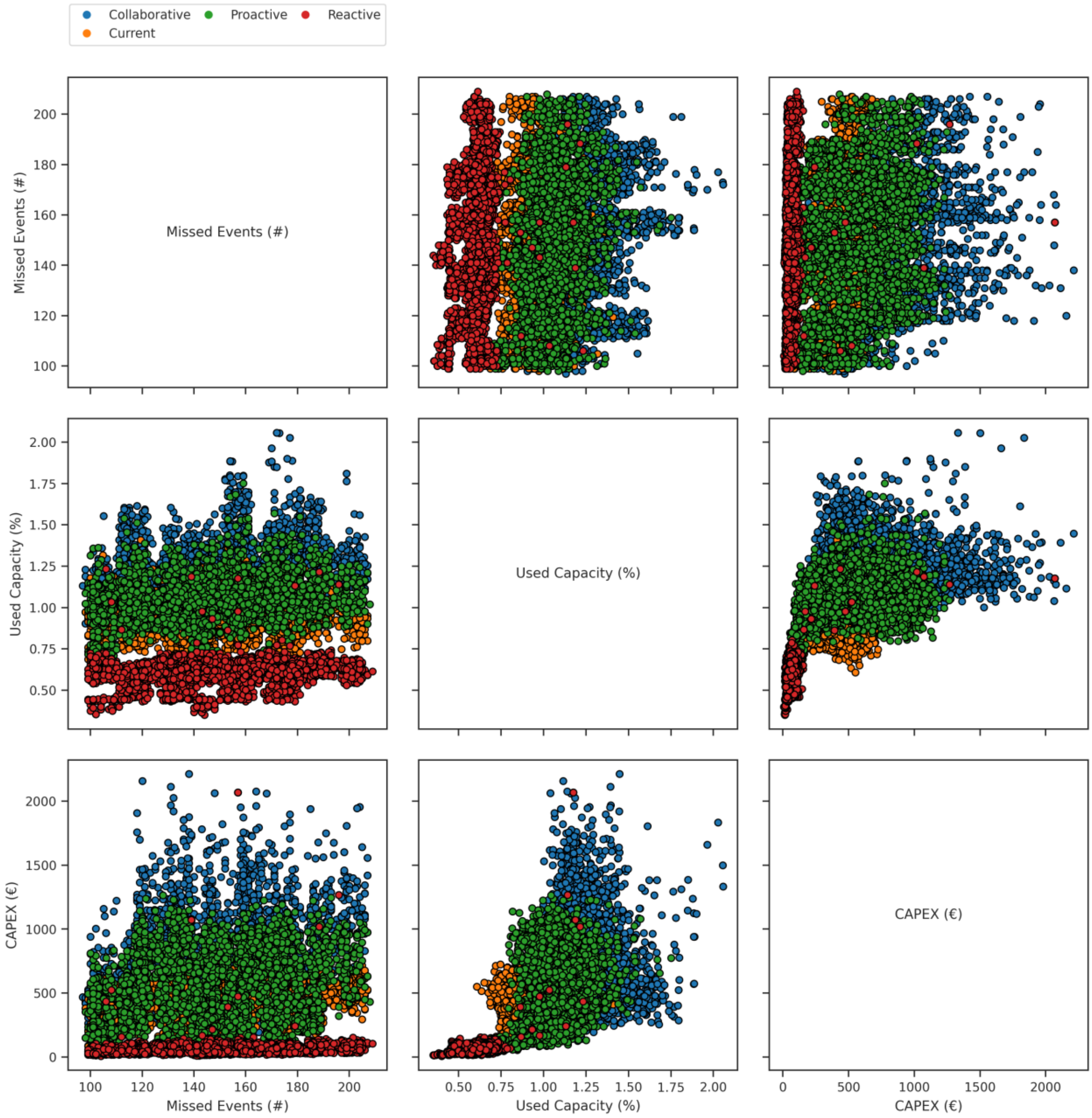
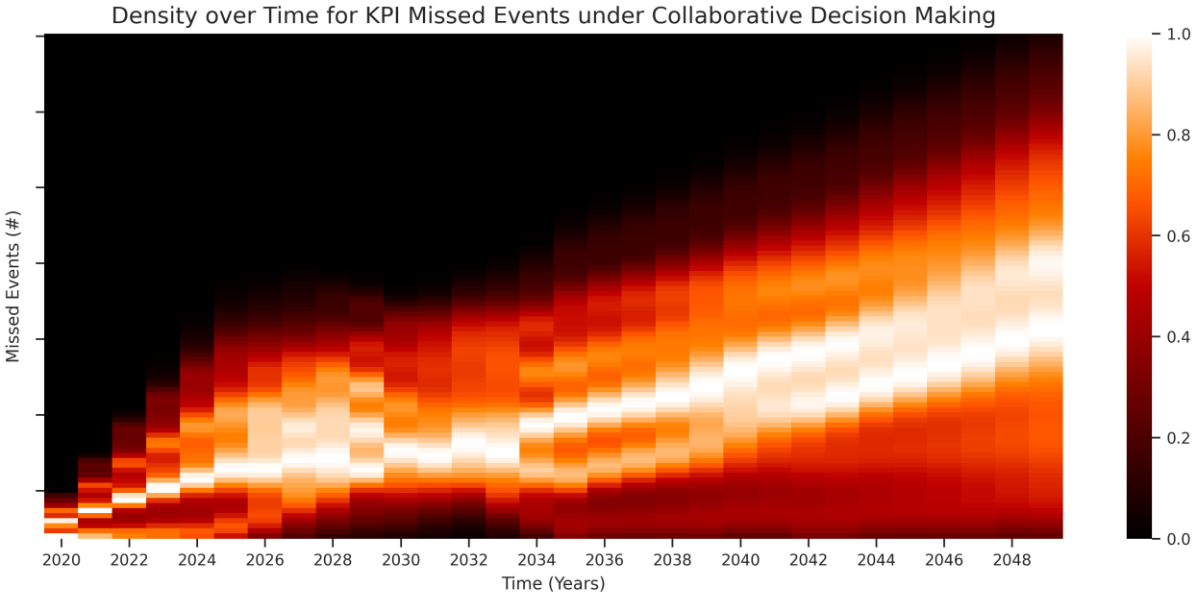
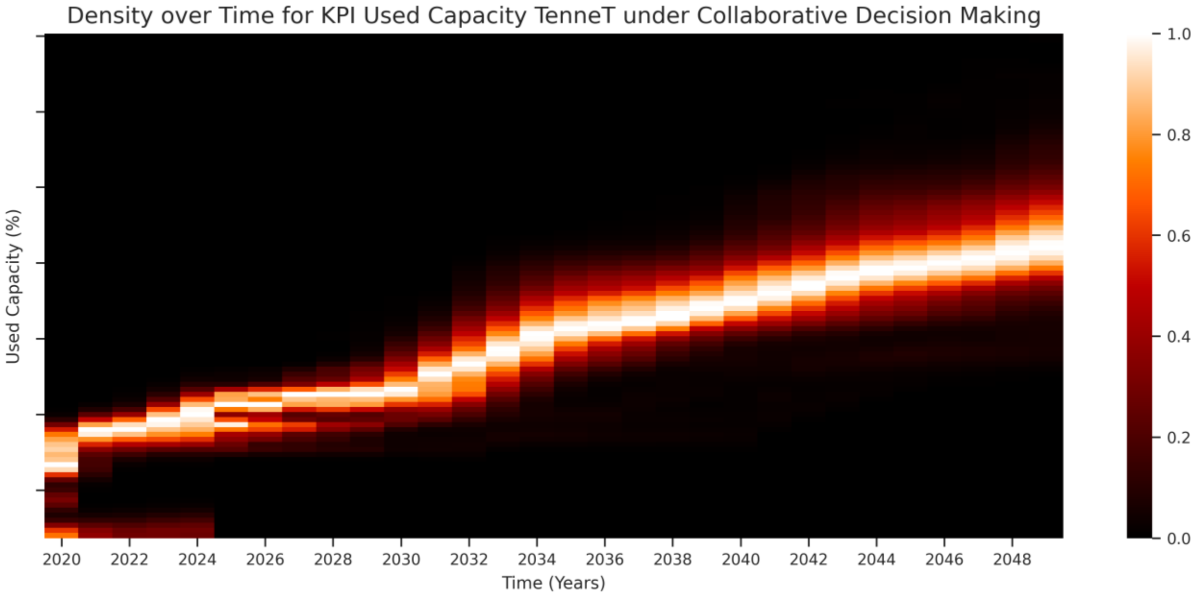
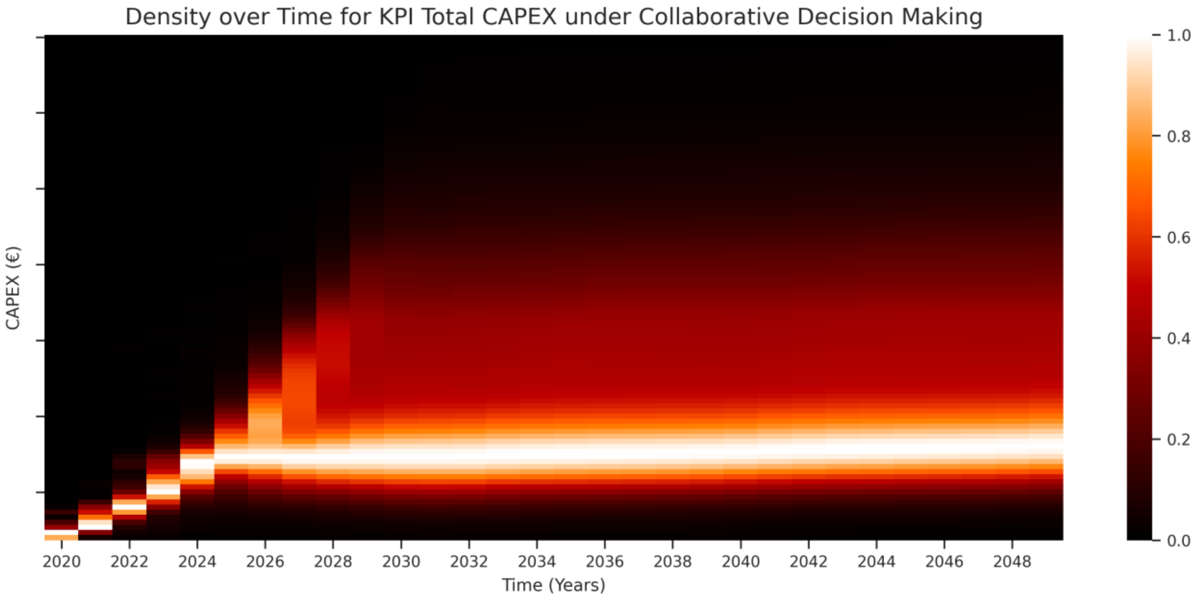


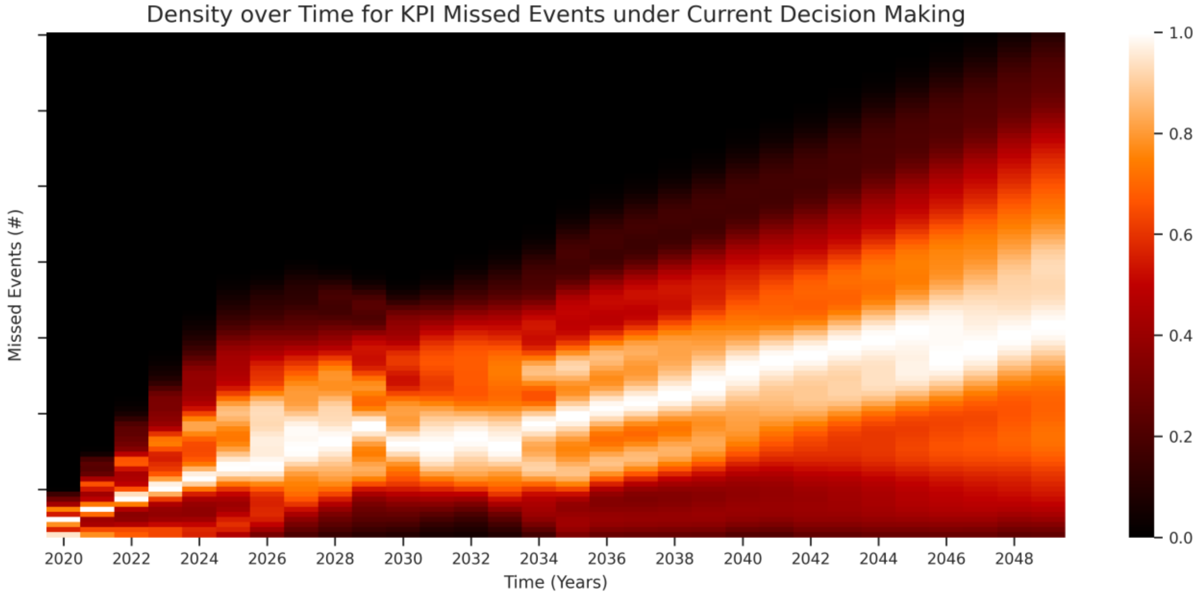
Figure 38 Coherence of the different outcomes, categorized per decision-making model


Density over Time for KPI Missed Events under Collaborative Decision Making
Missed Events (#)
Time (Years)
2020
2022
2024
2026
2028
2030
2032
2034
2036
2038
2040
2042
2044
2046
2048
0.0
0.2
0.4
0.6
0.8
1.0

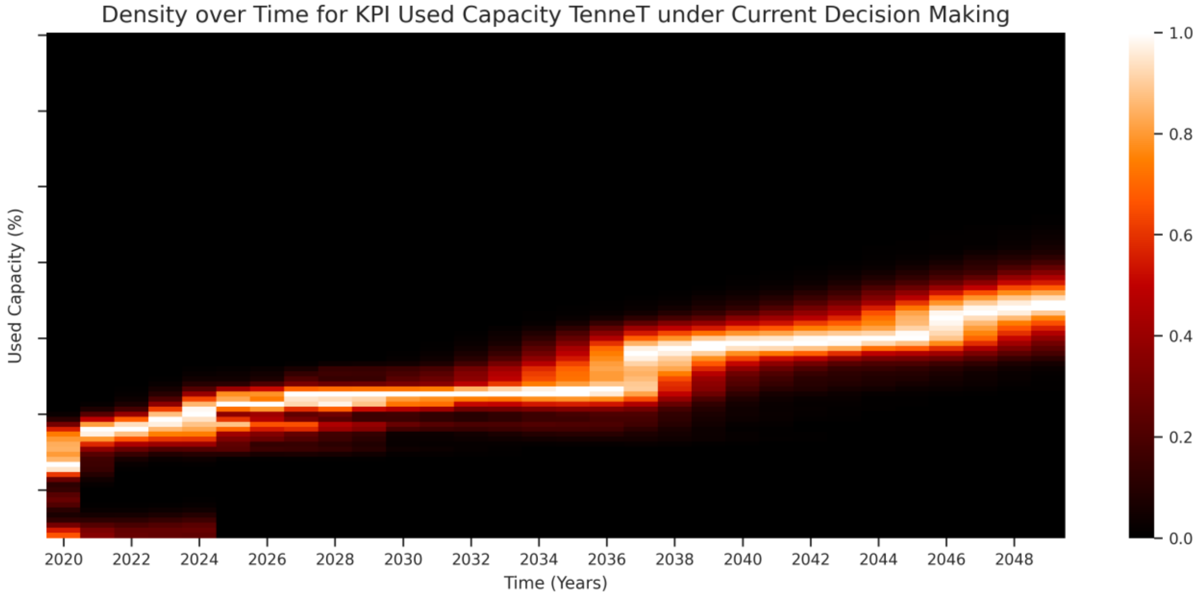

Density over Time for KPI Used Capacity TenneT under Collaborative Decision Making
Used Capacity (%)
Time (Years)
2020
2022
2024
2026
2028
2030
2032
2034
2036
2038
2040
2042
2044
2046
2048
0.0
0.2
0.4
0.6
0.8
1.0


Density over Time for KPI Total CAPEX under Collaborative Decision Making
CAPEX (€)
Time (Years)
2020
2022
2024
2026
2028
2030
2032
2034
2036
2038
2040
2042
2044
2046
2048
0.0
0.2
0.4
0.6
0.8
1.0

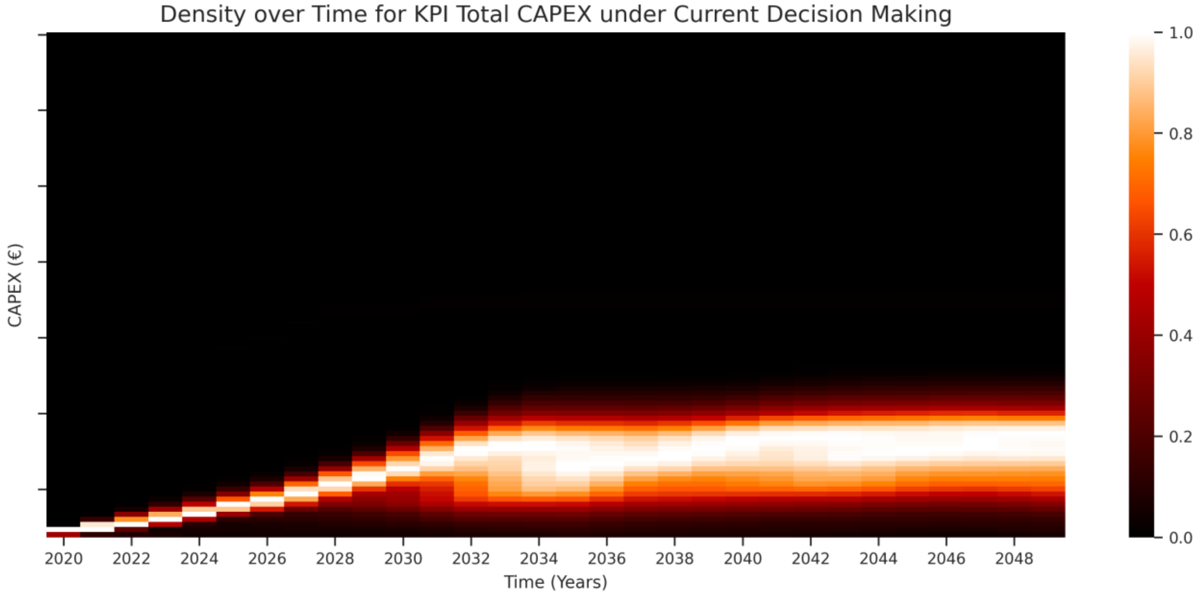
Density over Time for KPI Missed Events under Current Decision Making

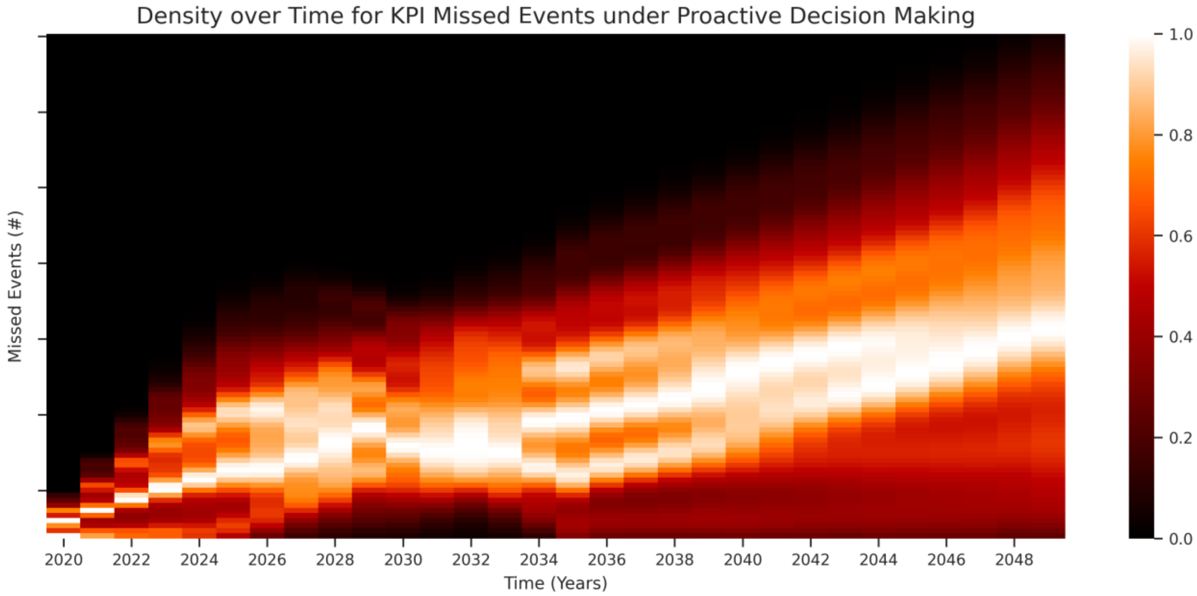


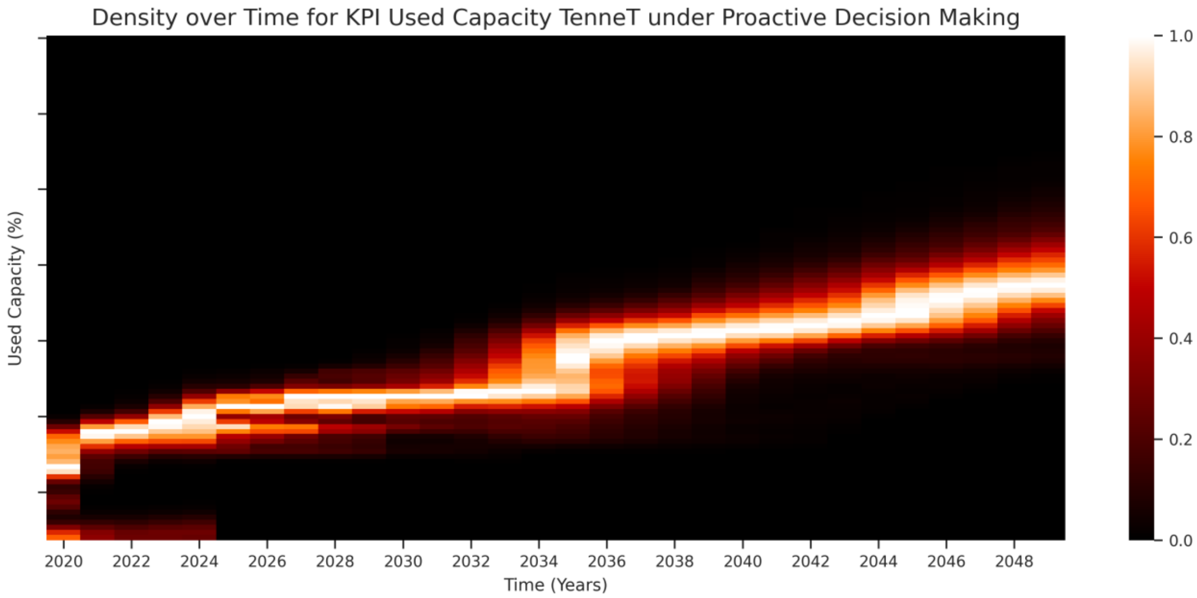
Density over Time for KPI Used Capacity TenneT under Current Decision Making

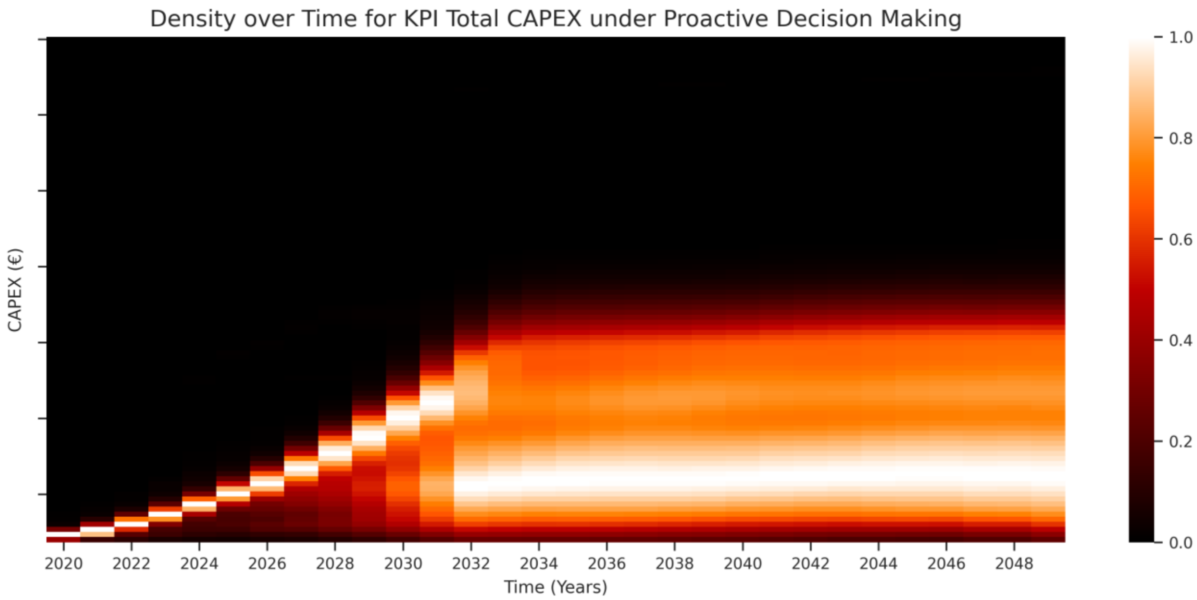


Density over Time for KPI Total CAPEX under Current Decision Making




Density over Time for KPI Missed Events under Proactive Decision Making
Missed Events (#)
2020
2022
2024
2026
2028
2030
2032
2034
2036
2038
2040
2042
2044
2046
2048
Time (Years)
1.0
0.8
0.6
0.4
0.2
0.0


Density over Time for KPI Used Capacity TenneT under Proactive Decision Making
Used Capacity (%)
2020
2022
2024
2026
2028
2030
2032
2034
2036
2038
2040
2042
2044
2046
2048
Time (Years)
1.0
0.8
0.6
0.4
0.2
0.0


Density over Time for KPI Total CAPEX under Proactive Decision Making
CAPEX (€)
2020
2022
2024
2026
2028
2030
2032
2034
2036
2038
2040
2042
2044
2046
2048
Time (Years)
1.0
0.8
0.6
0.4
0.2
0.0

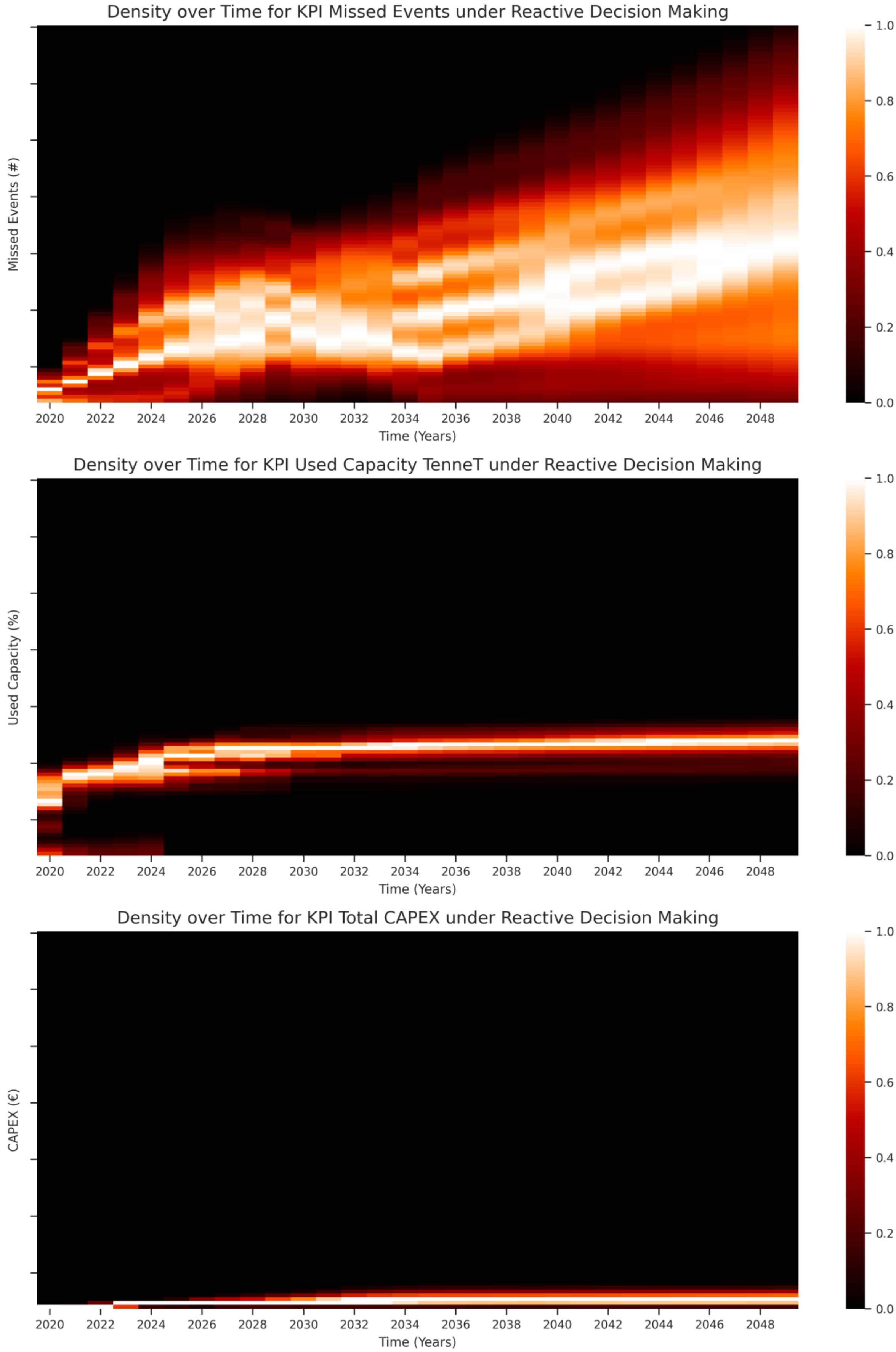


Figure 39 **Density over time for the different KPI's under the different decision-making strategies**

B.

Sobol Sensitivity Analysis

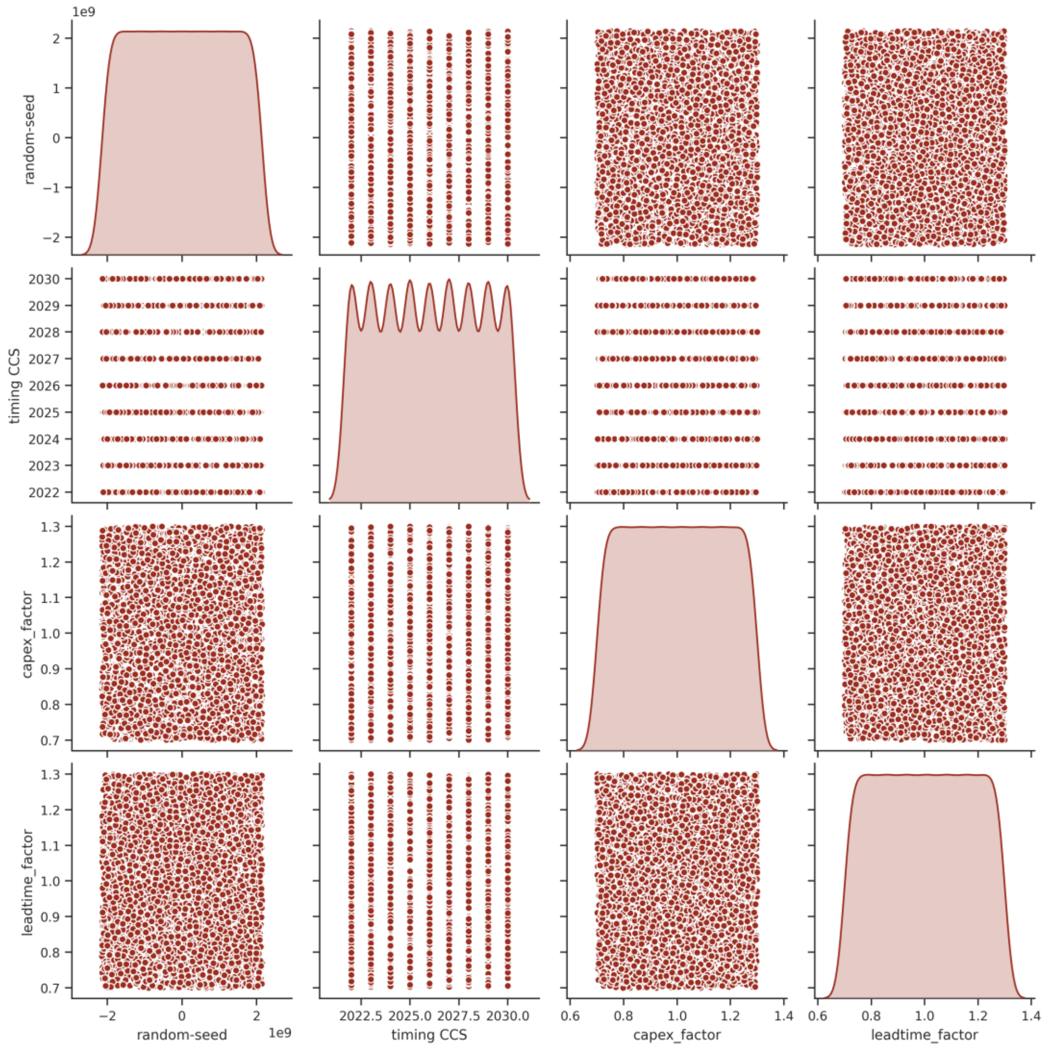
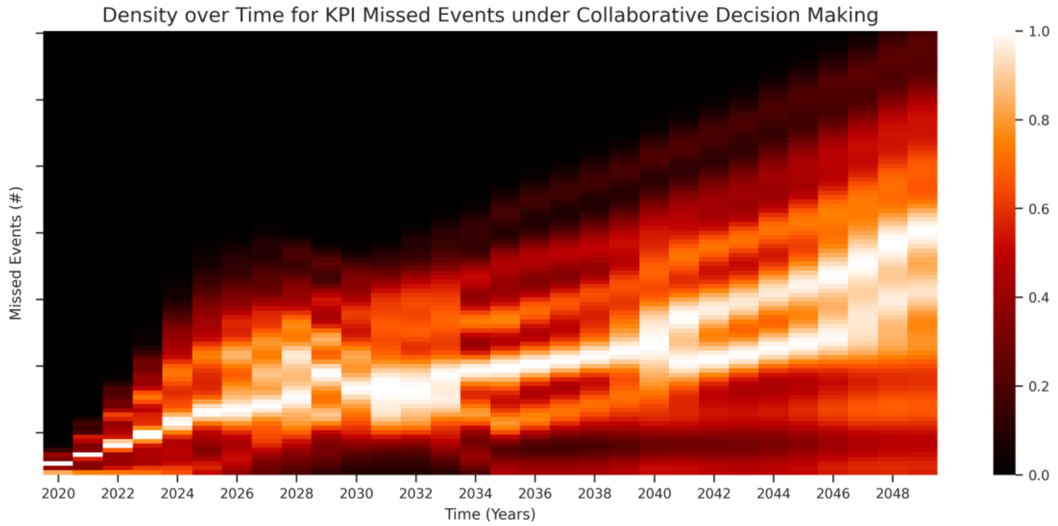
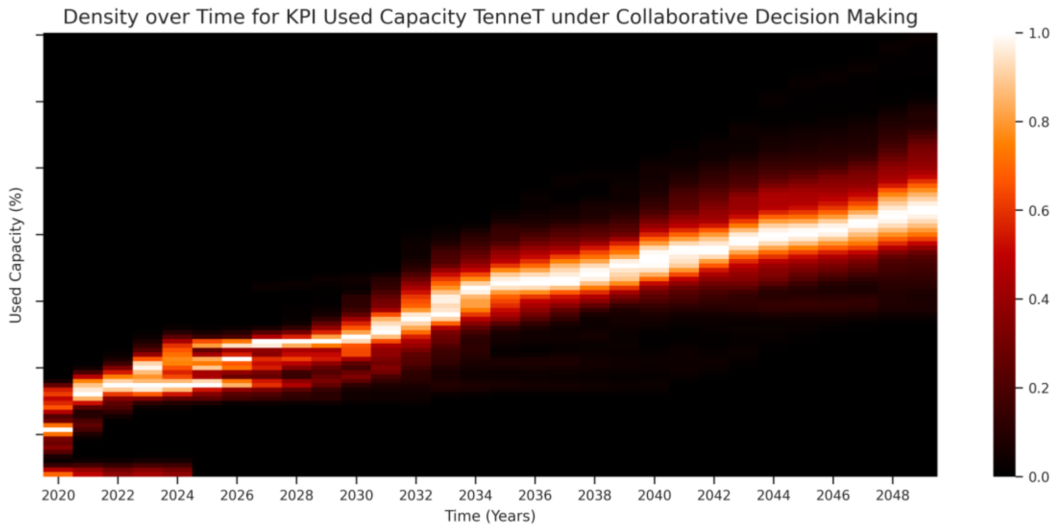
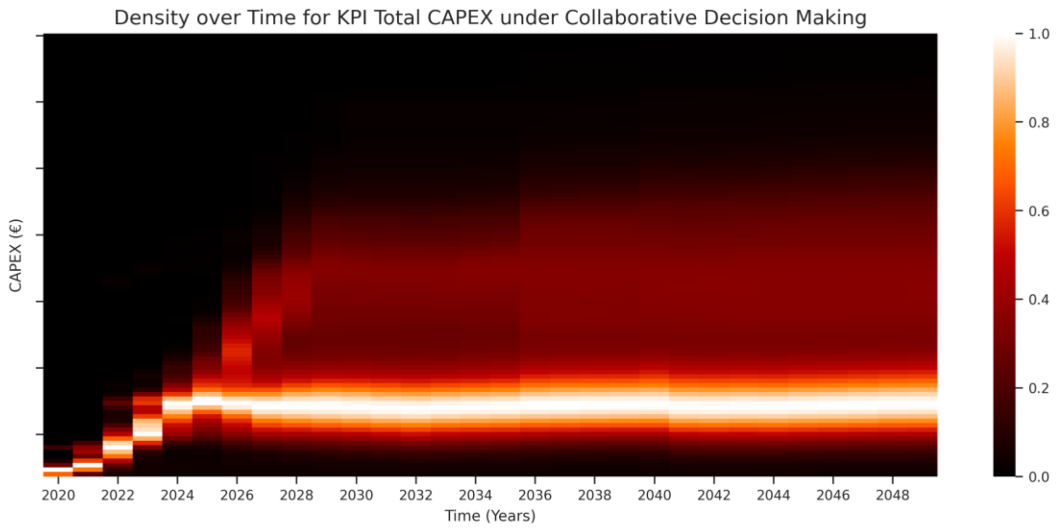
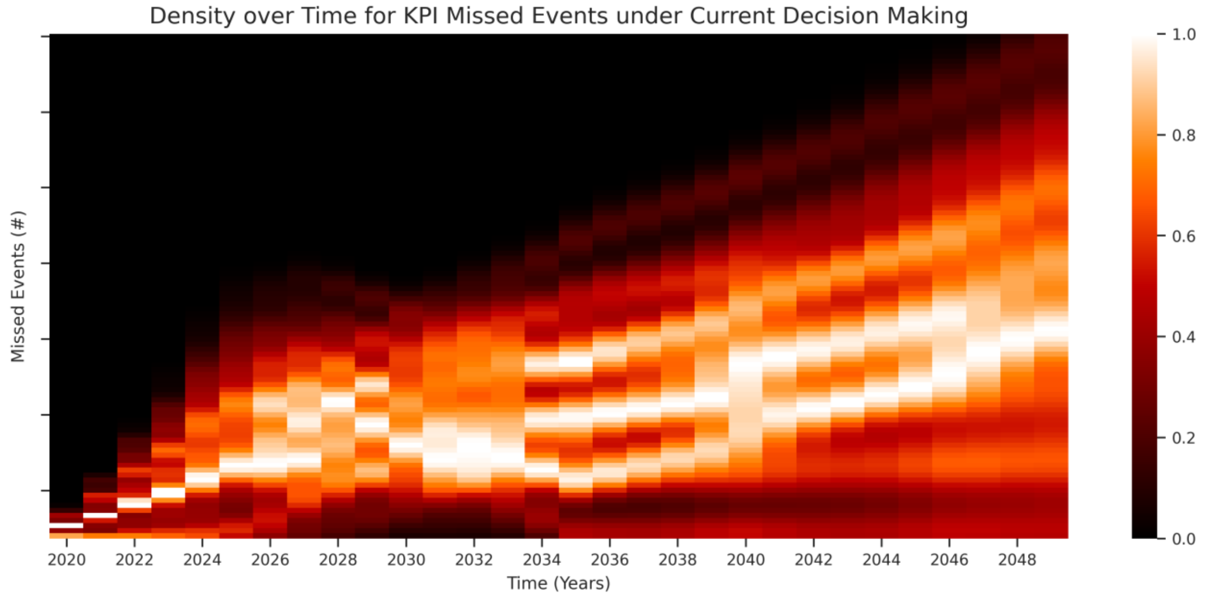


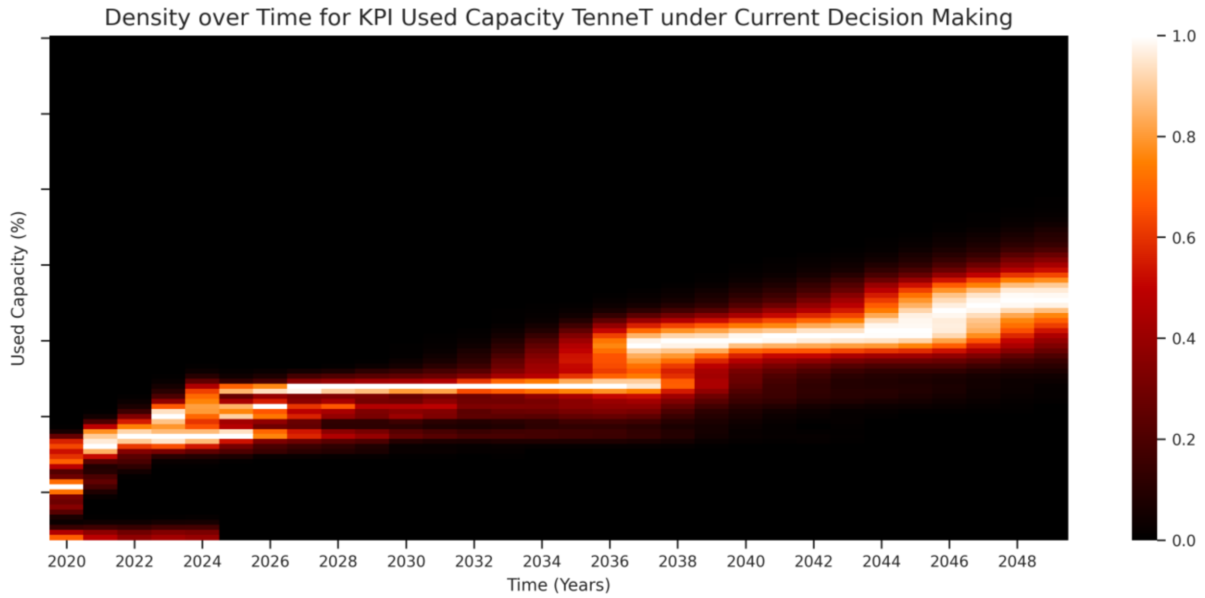
Figure 40 **Pair plot of the sampled values based on the Sobol sequence. From left (and top) to right (and bottom): Random seed, timing CCS, CAPEX factor, and lead time factor. Each point represents a sampled value in the domain of the various uncertainties. The KDE plot indicates the univariate distribution.**

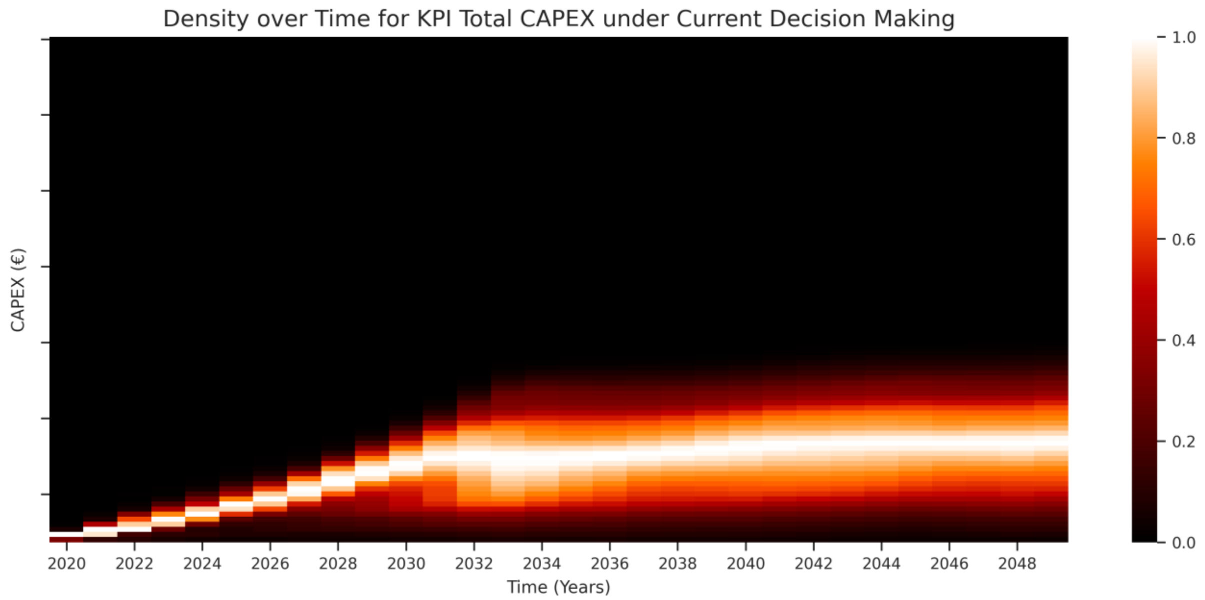

Density over Time for KPI Missed Events under Collaborative Decision Making
Missed Events (#)
Time (Years)
2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048
0.0 0.2 0.4 0.6 0.8 1.0

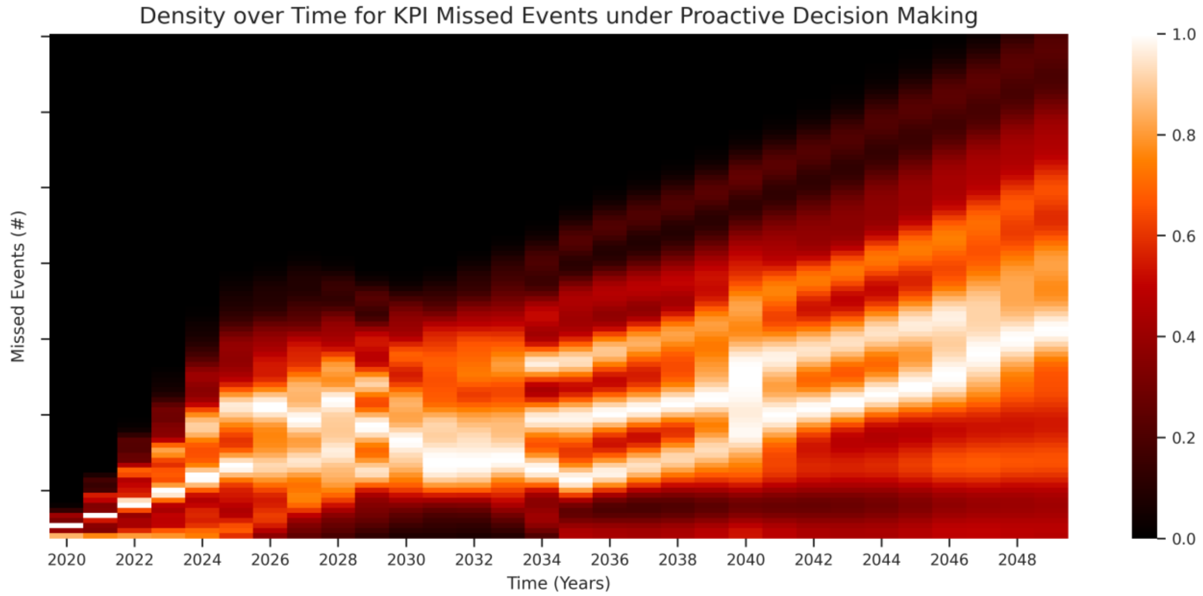

Density over Time for KPI Used Capacity TenneT under Collaborative Decision Making
Used Capacity (%)
Time (Years)
2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048
0.0 0.2 0.4 0.6 0.8 1.0

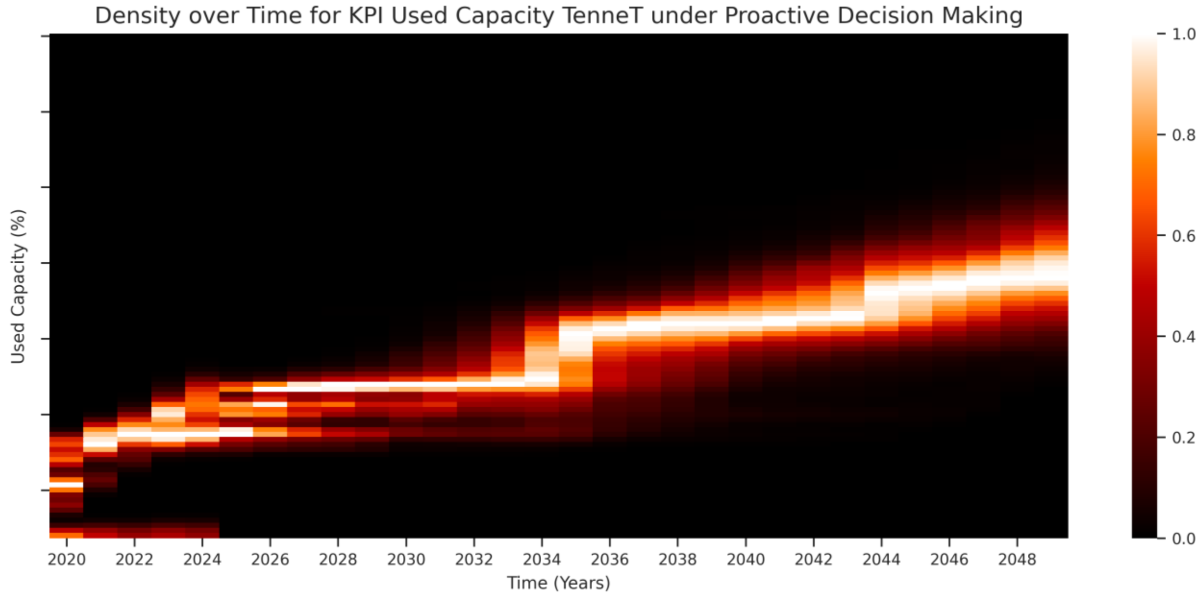

Density over Time for KPI Total CAPEX under Collaborative Decision Making
CAPEX (€)
Time (Years)
2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048
0.0 0.2 0.4 0.6 0.8 1.0

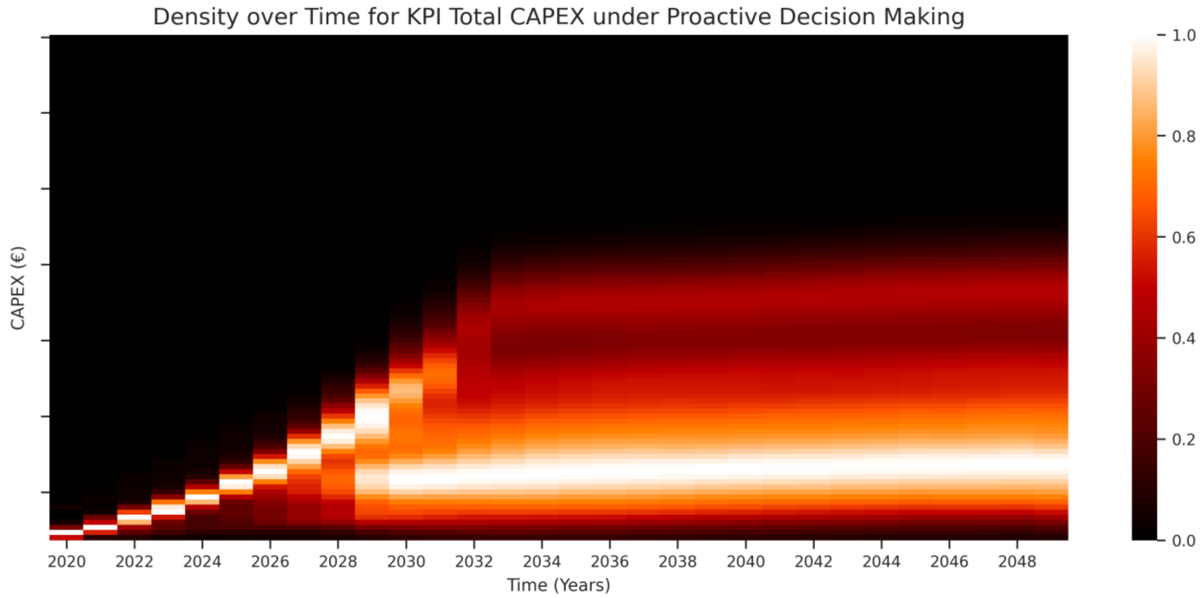

Density over Time for KPI Missed Events under Current Decision Making
Missed Events (#)
Time (Years)
2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048
0.0 0.2 0.4 0.6 0.8 1.0


Density over Time for KPI Used Capacity TenneT under Current Decision Making
Used Capacity (%)
Time (Years)
2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048
0.0 0.2 0.4 0.6 0.8 1.0


Density over Time for KPI Total CAPEX under Current Decision Making
CAPEX (€)
Time (Years)
2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048
0.0 0.2 0.4 0.6 0.8 1.0


Density over Time for KPI Missed Events under Proactive Decision Making
Missed Events (#)
2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048
Time (Years)
0.0 0.2 0.4 0.6 0.8 1.0


Density over Time for KPI Used Capacity TenneT under Proactive Decision Making
Used Capacity (%)
2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048
Time (Years)
0.0 0.2 0.4 0.6 0.8 1.0


Density over Time for KPI Total CAPEX under Proactive Decision Making
CAPEX (€)
2020 2022 2024 2026 2028 2030 2032 2034 2036 2038 2040 2042 2044 2046 2048
Time (Years)
0.0 0.2 0.4 0.6 0.8 1.0

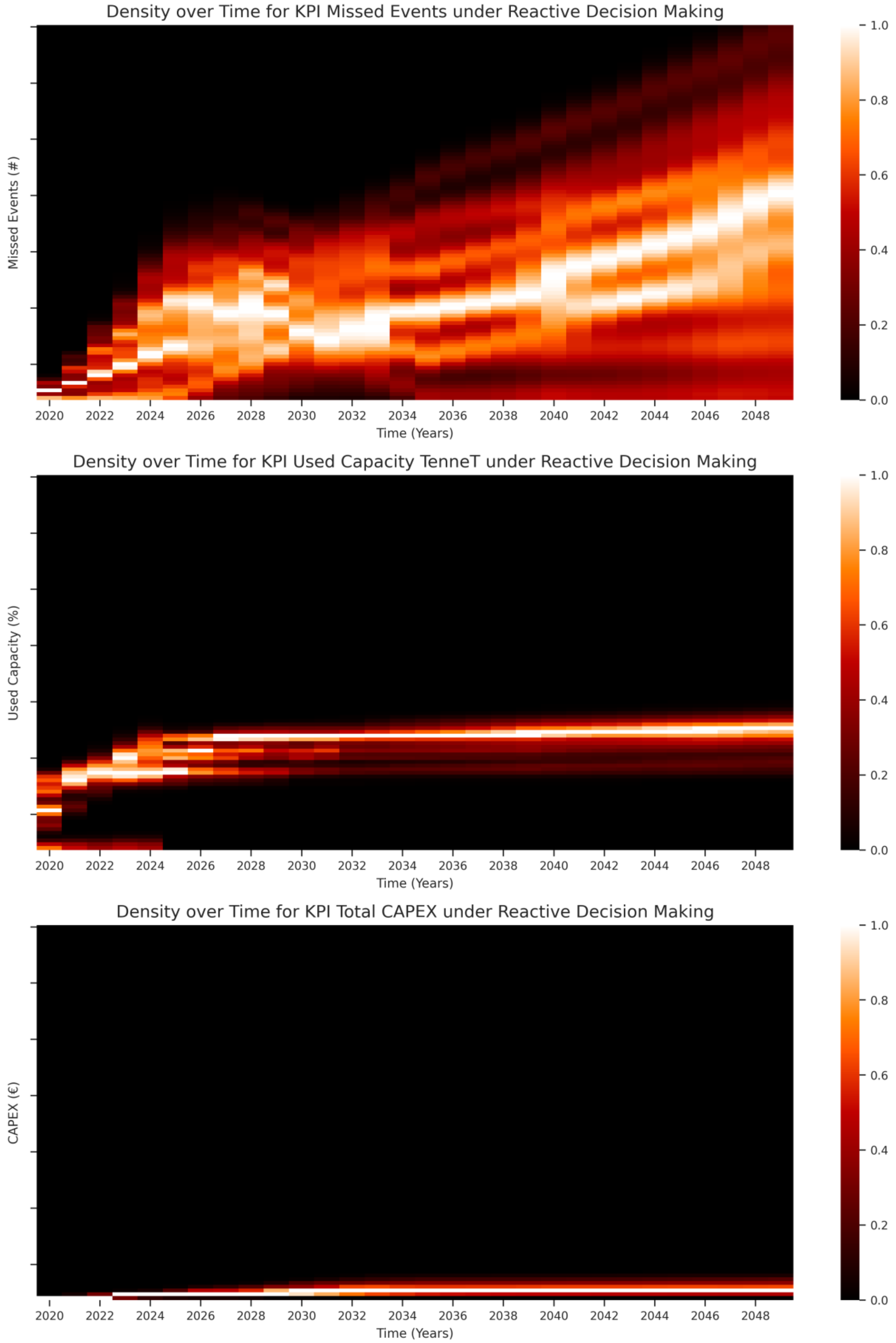
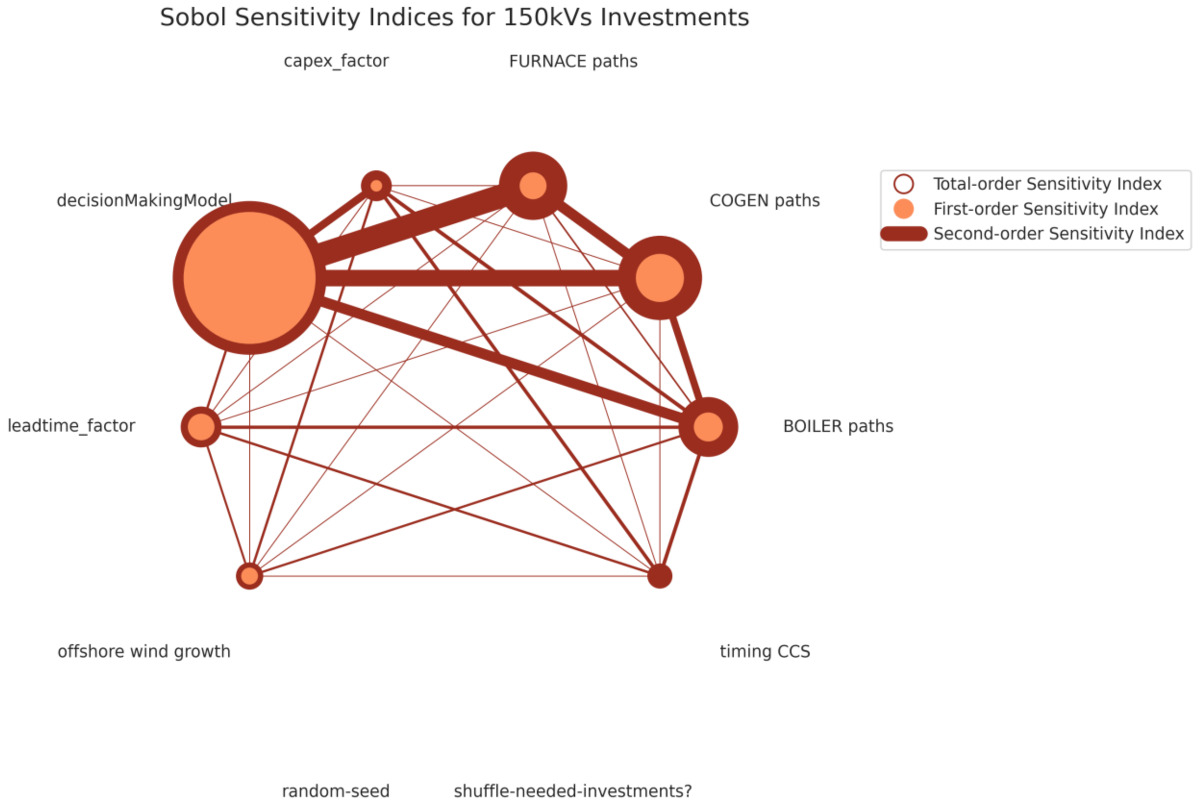
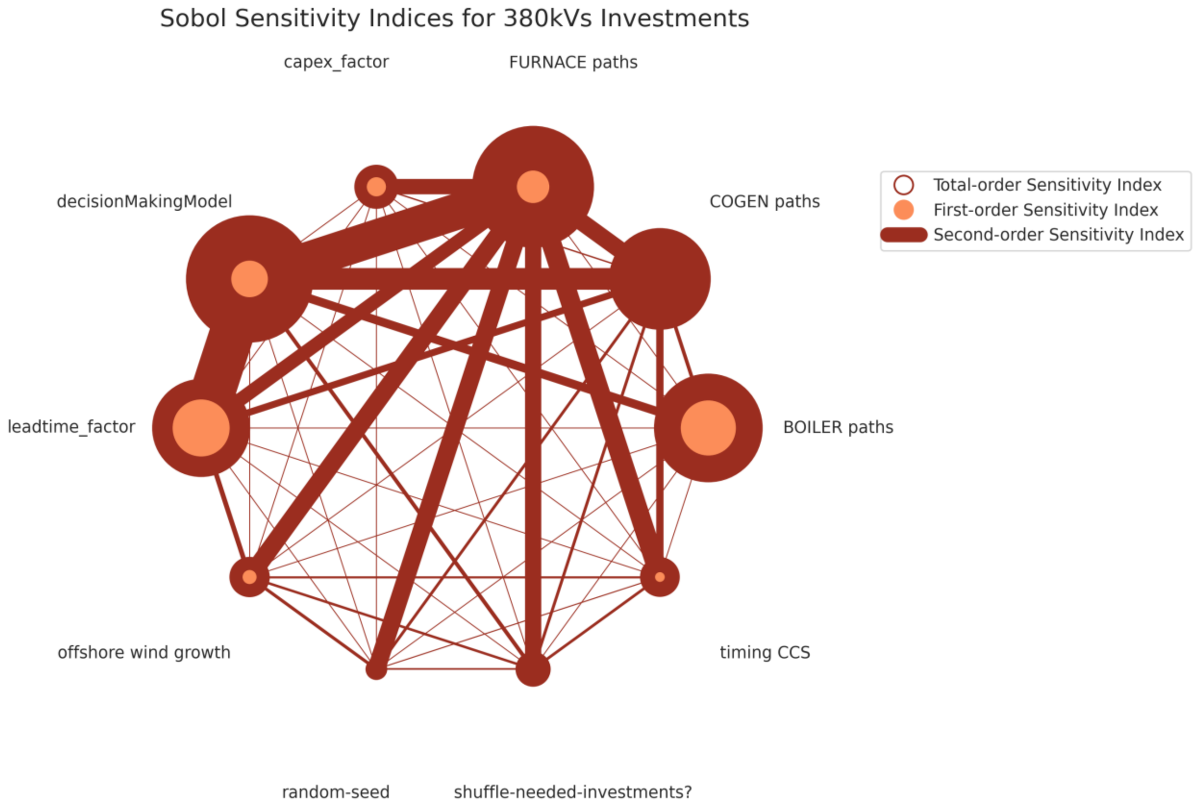
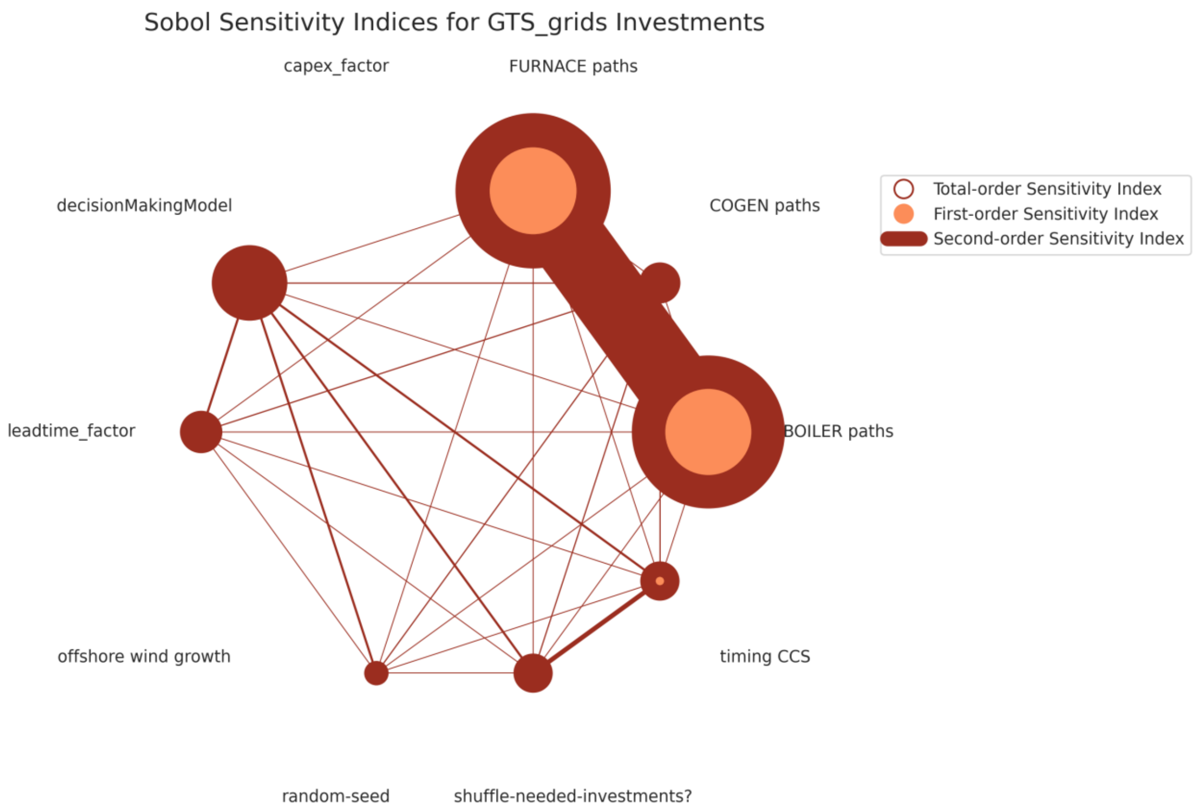
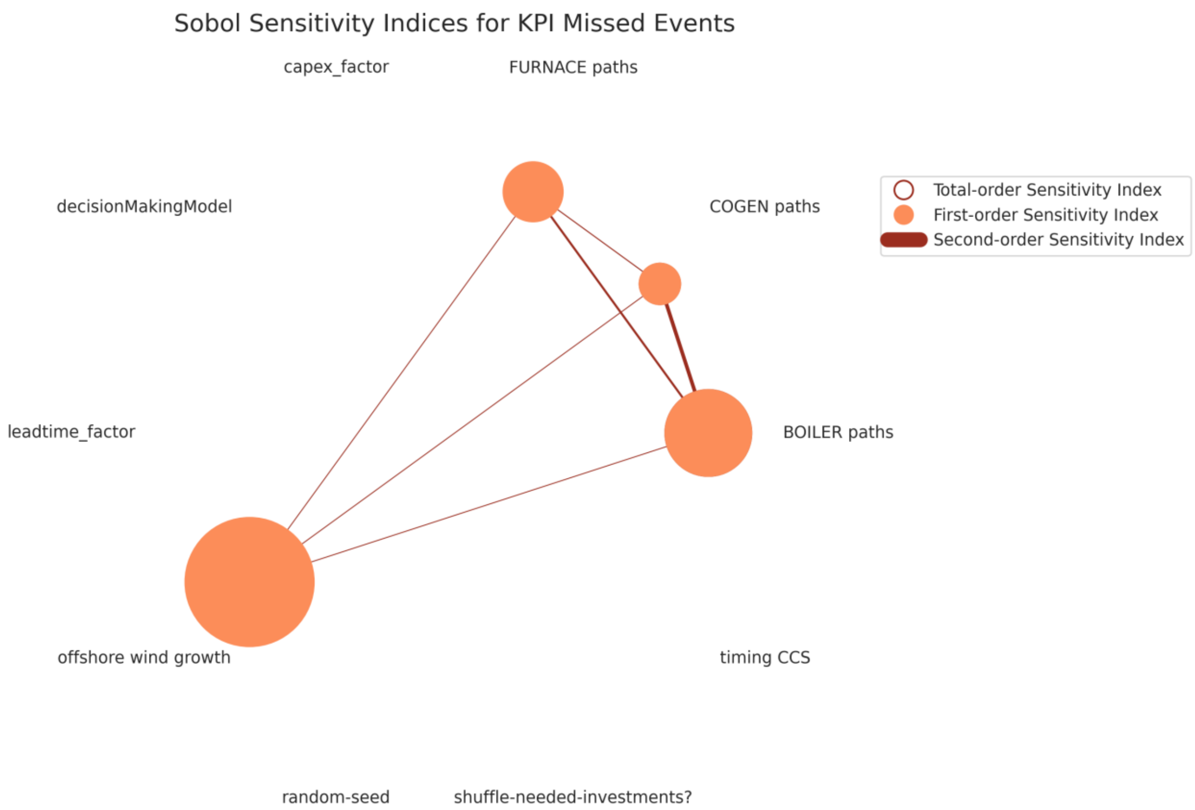


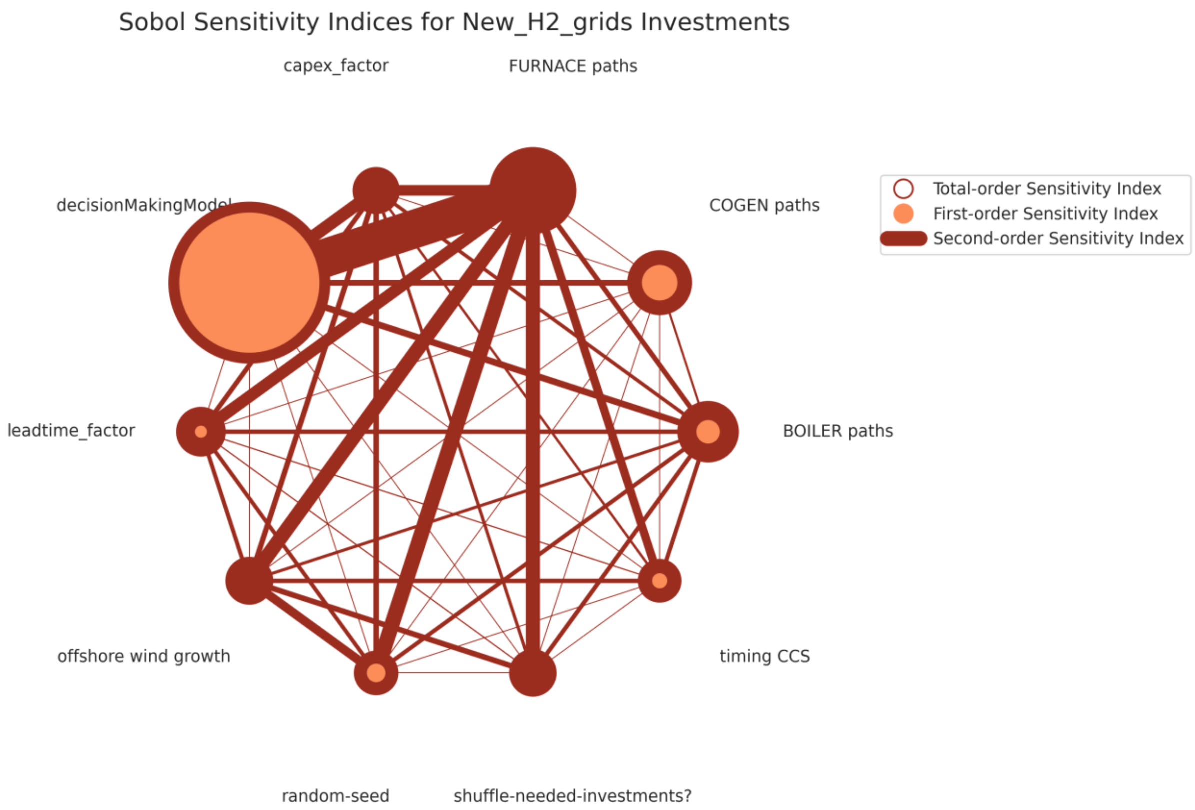
Figure 41 Density over time for KPI's per decision-making model

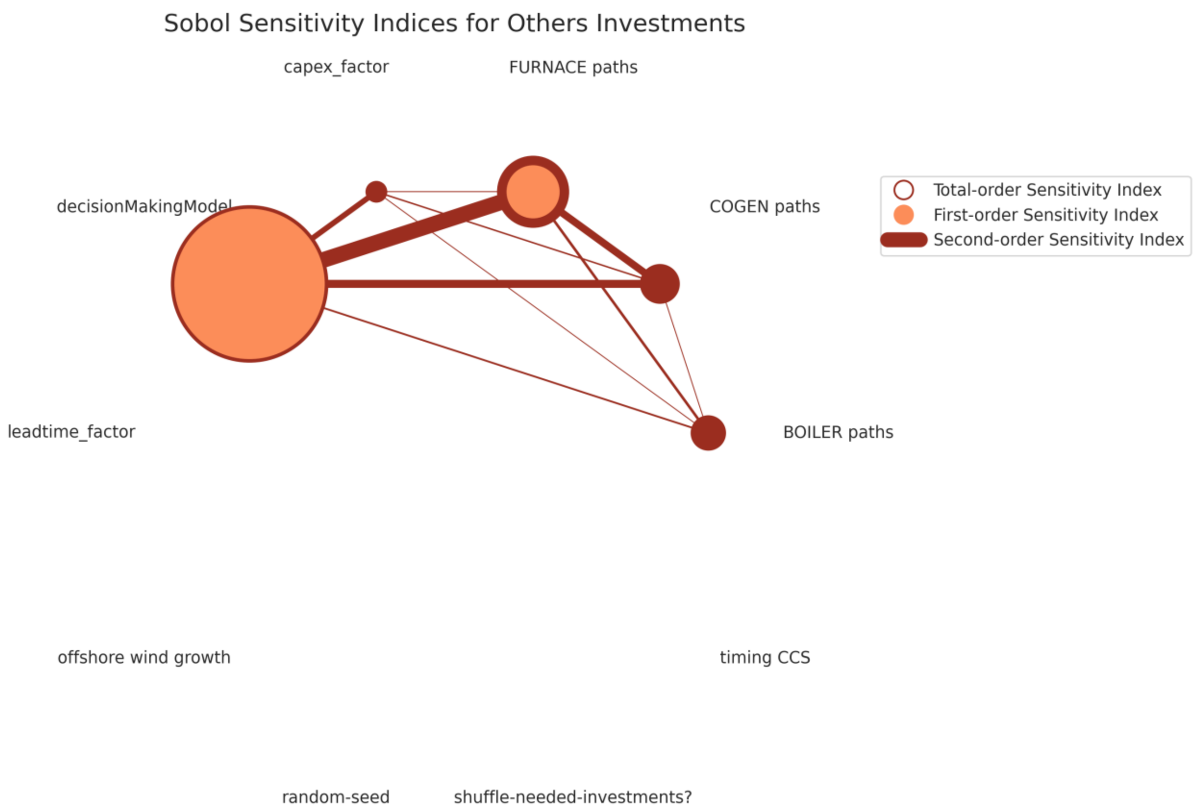

Sobol Sensitivity Indices for 150kVs Investments
capex_factor
FURNACE paths
decisionMakingModel
COGEN paths
leadtime_factor
BOILER paths
offshore wind growth
timing CCS
random-seed
shuffle-needed-investments?
Total-order Sensitivity Index
First-order Sensitivity Index
Second-order Sensitivity Index


Sobol Sensitivity Indices for 380kVs Investments
capex_factor
FURNACE paths
decisionMakingModel
COGEN paths
leadtime_factor
BOILER paths
offshore wind growth
timing CCS
random-seed
shuffle-needed-investments?
Total-order Sensitivity Index
First-order Sensitivity Index
Second-order Sensitivity Index


Sobol Sensitivity Indices for GTS_grids Investments
capex_factor
FURNACE paths
COGEN paths
BOILER paths
timing CCS
shuffle-needed-investments?
random-seed
offshore wind growth
leadtime_factor
decisionMakingModel
Total-order Sensitivity Index
First-order Sensitivity Index
Second-order Sensitivity Index


Sobol Sensitivity Indices for KPI Missed Events
capex_factor
FURNACE paths
COGEN paths
BOILER paths
timing CCS
shuffle-needed-investments?
random-seed
offshore wind growth
leadtime_factor
decisionMakingModel
Total-order Sensitivity Index
First-order Sensitivity Index
Second-order Sensitivity Index


Sobol Sensitivity Indices for New_H2_grids Investments
capex_factor
FURNACE paths
decisionMakingModel
COGEN paths
leadtime_factor
BOILER paths
offshore wind growth
timing CCS
random-seed
shuffle-needed-investments?
Total-order Sensitivity Index
First-order Sensitivity Index
Second-order Sensitivity Index


Sobol Sensitivity Indices for Others Investments
capex_factor
FURNACE paths
decisionMakingModel
COGEN paths
leadtime_factor
BOILER paths
offshore wind growth
timing CCS
random-seed
shuffle-needed-investments?
Total-order Sensitivity Index
First-order Sensitivity Index
Second-order Sensitivity Index

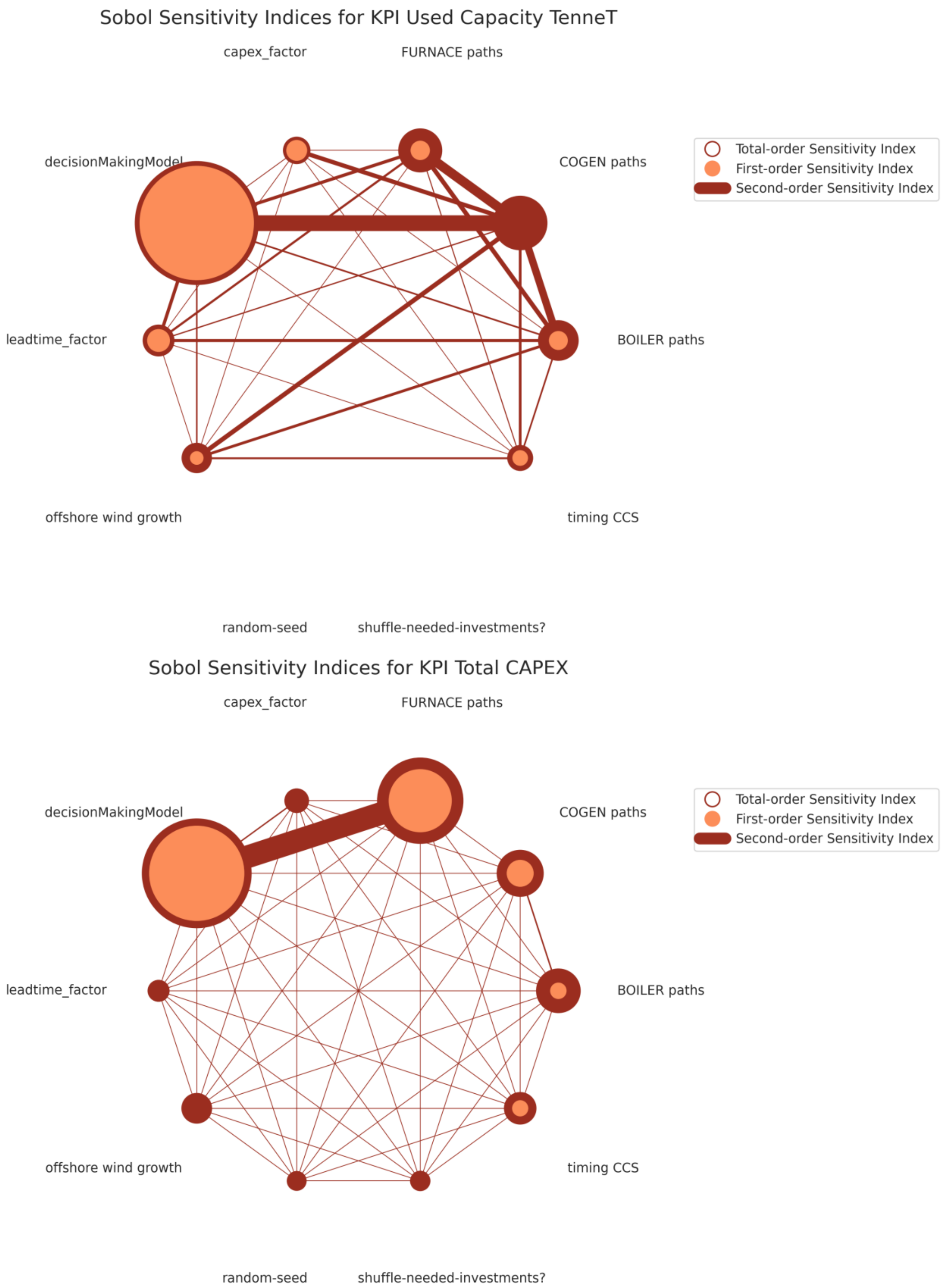
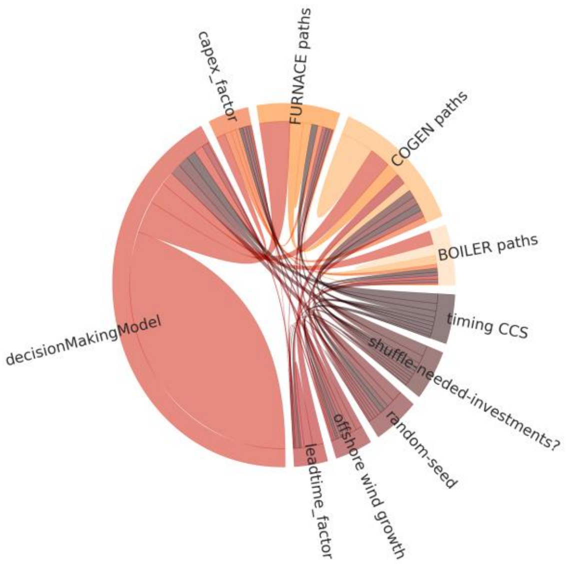
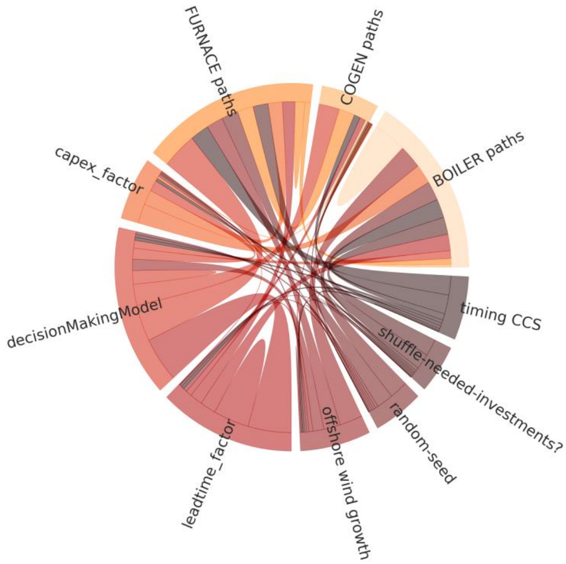


Figure 42 **Sobol Sensitivity Indices per KPI and investment category**

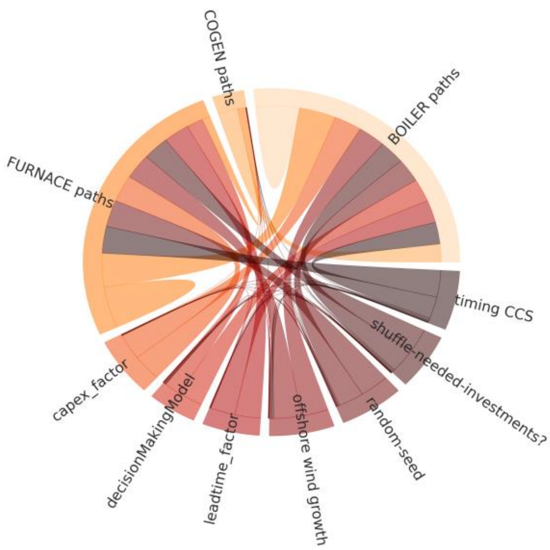
First and Second Sobol Sensitivity Indices for 150kVs Investments



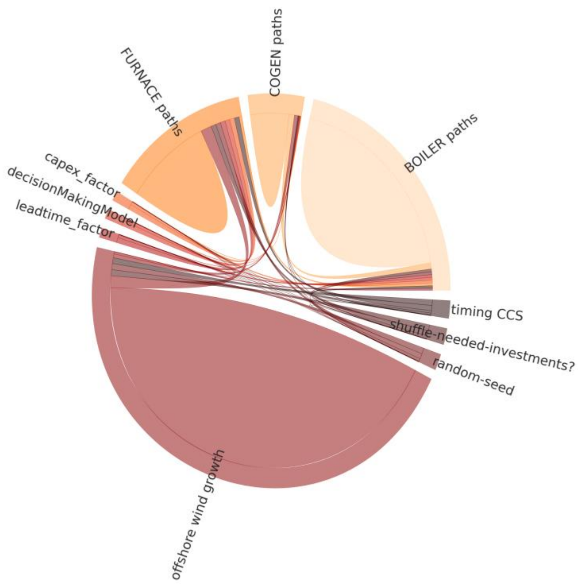
First and Second Sobol Sensitivity Indices for 380kVs Investments



First and Second Sobol Sensitivity Indices for GTS_grids Investments



First and Second Sobol Sensitivity Indices for KPI Missed Events



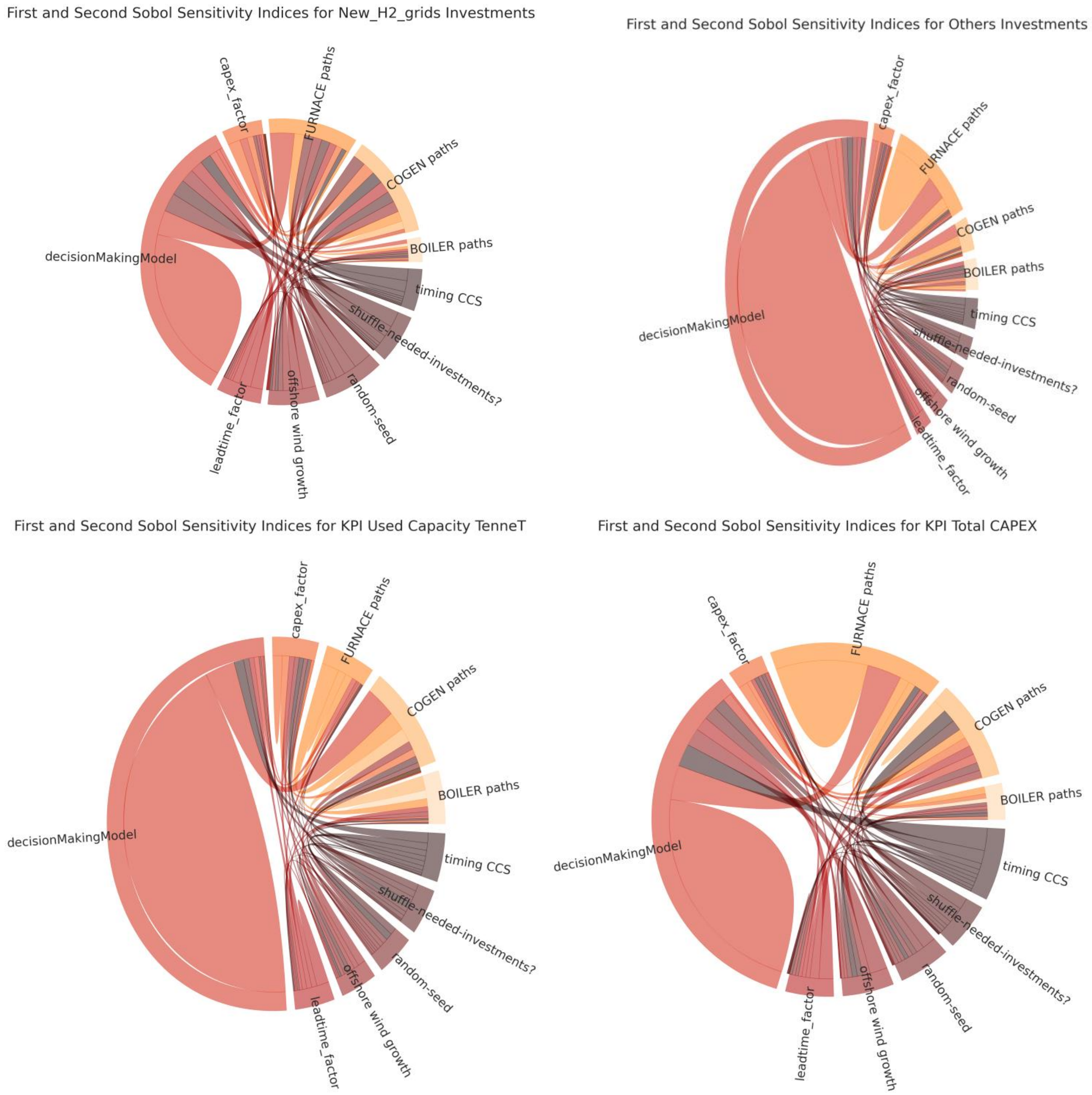
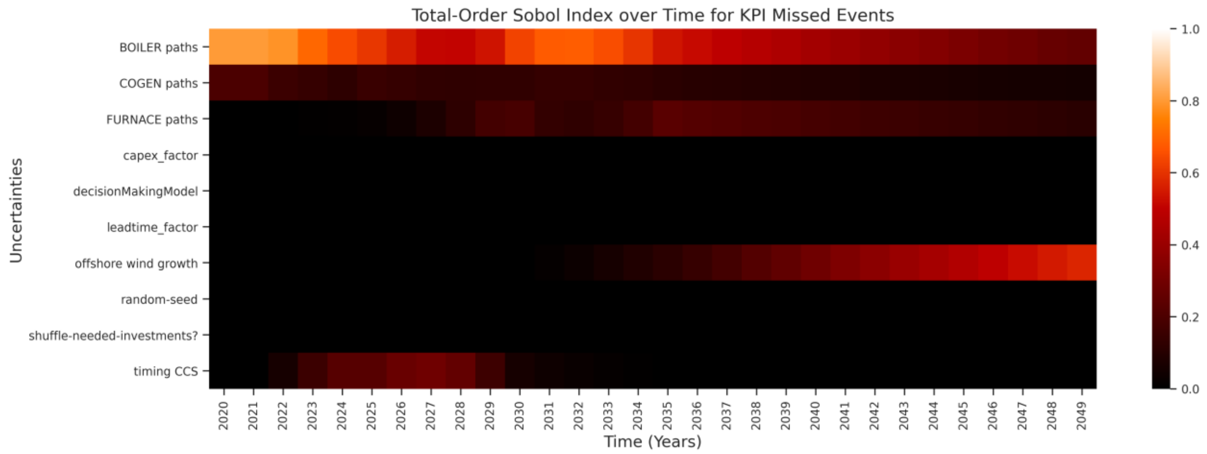
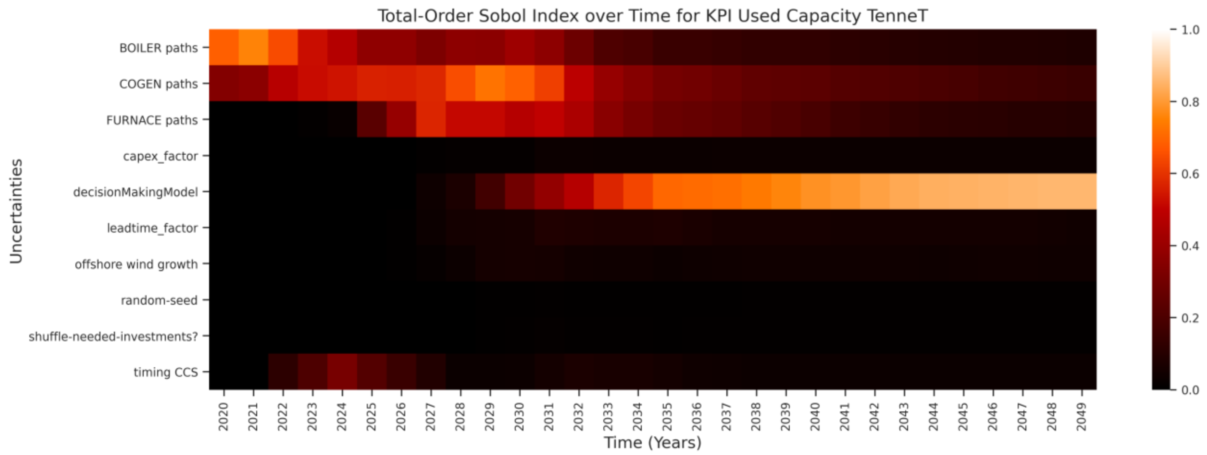
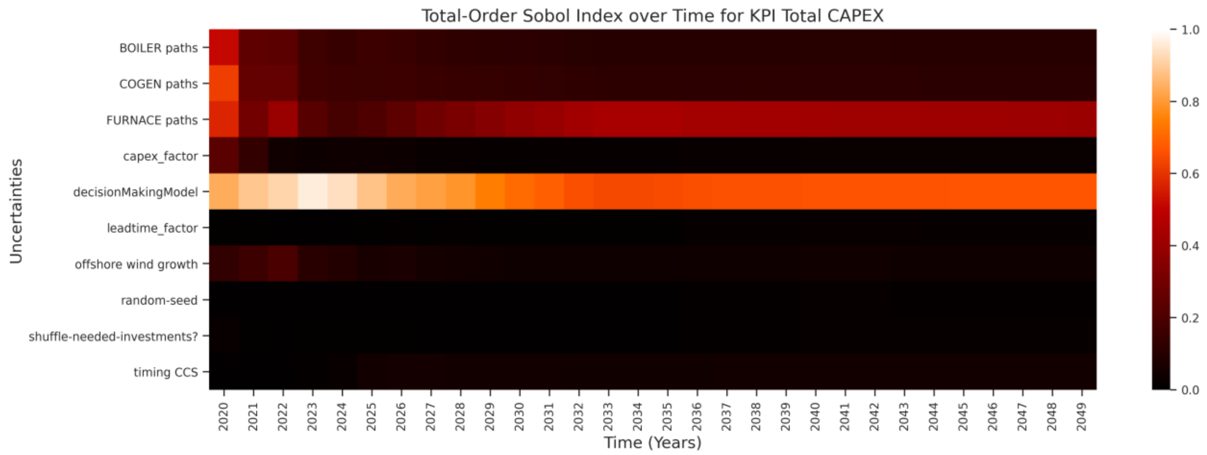
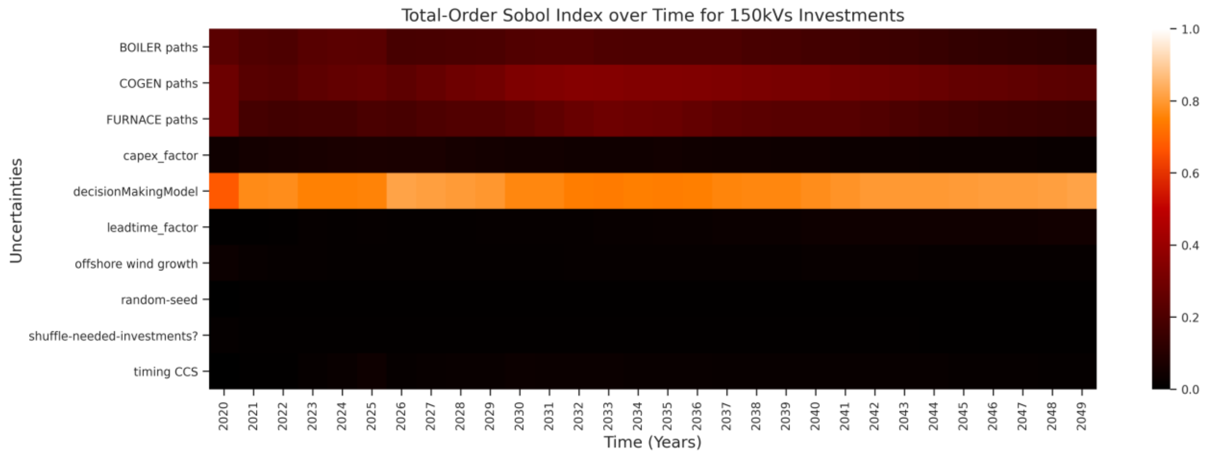


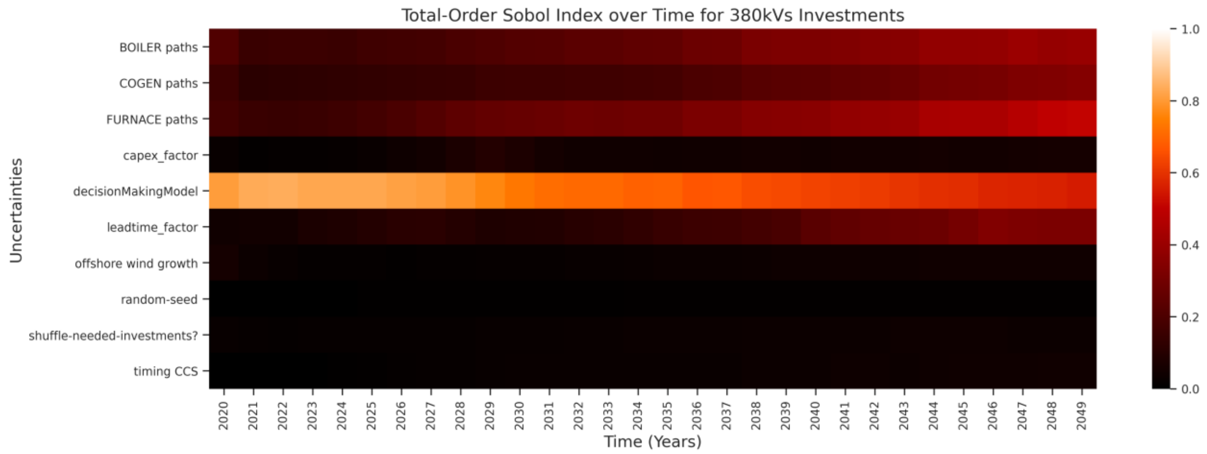
Figure 43 **First- and second-order sensitivity indices per KPI and investment category**

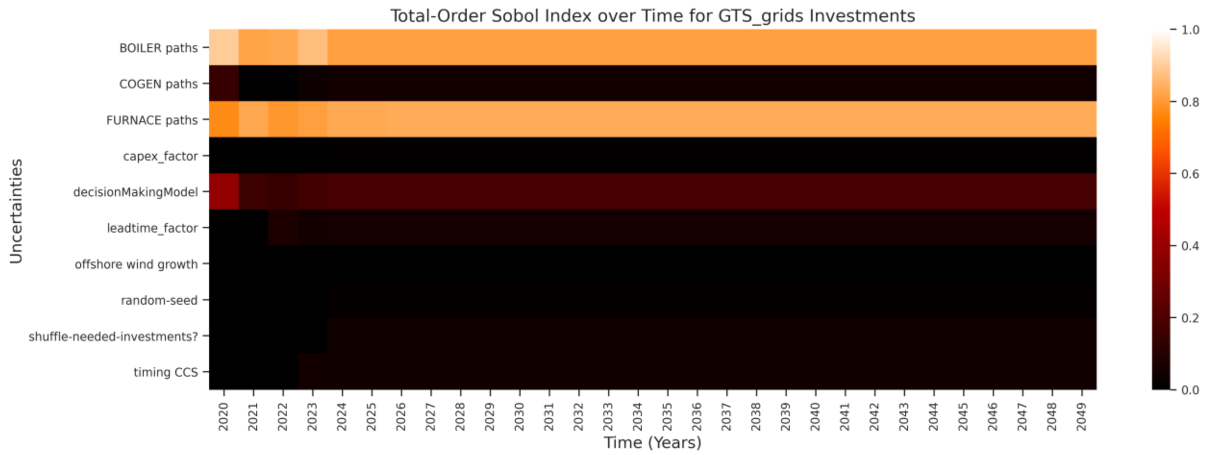

Total-Order Sobol Index over Time for KPI Missed Events
Uncertainties
BOILER paths
COGEN paths
FURNACE paths
capex_factor
decisionMakingModel
leadtime_factor
offshore wind growth
random-seed
shuffle-needed-investments?
timing CCS
2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049
Time (Years)
1.0 0.8 0.6 0.4 0.2 0.0


Total-Order Sobol Index over Time for KPI Used Capacity TenneT
Uncertainties
BOILER paths
COGEN paths
FURNACE paths
capex_factor
decisionMakingModel
leadtime_factor
offshore wind growth
random-seed
shuffle-needed-investments?
timing CCS
2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049
Time (Years)
1.0 0.8 0.6 0.4 0.2 0.0


Total-Order Sobol Index over Time for KPI Total CAPEX
Uncertainties
BOILER paths
COGEN paths
FURNACE paths
capex_factor
decisionMakingModel
leadtime_factor
offshore wind growth
random-seed
shuffle-needed-investments?
timing CCS
2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049
Time (Years)
1.0 0.8 0.6 0.4 0.2 0.0


Total-Order Sobol Index over Time for 150kVs Investments
Uncertainties
BOILER paths
COGEN paths
FURNACE paths
capex_factor
decisionMakingModel
leadtime_factor
offshore wind growth
random-seed
shuffle-needed-investments?
timing CCS
Time (Years)
2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049
0.0 0.2 0.4 0.6 0.8 1.0


Total-Order Sobol Index over Time for 380kVs Investments
Uncertainties
BOILER paths
COGEN paths
FURNACE paths
capex_factor
decisionMakingModel
leadtime_factor
offshore wind growth
random-seed
shuffle-needed-investments?
timing CCS
Time (Years)
2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049
0.0 0.2 0.4 0.6 0.8 1.0


Total-Order Sobol Index over Time for GTS_grids Investments
Uncertainties
BOILER paths
COGEN paths
FURNACE paths
capex_factor
decisionMakingModel
leadtime_factor
offshore wind growth
random-seed
shuffle-needed-investments?
timing CCS
Time (Years)
2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049
0.0 0.2 0.4 0.6 0.8 1.0

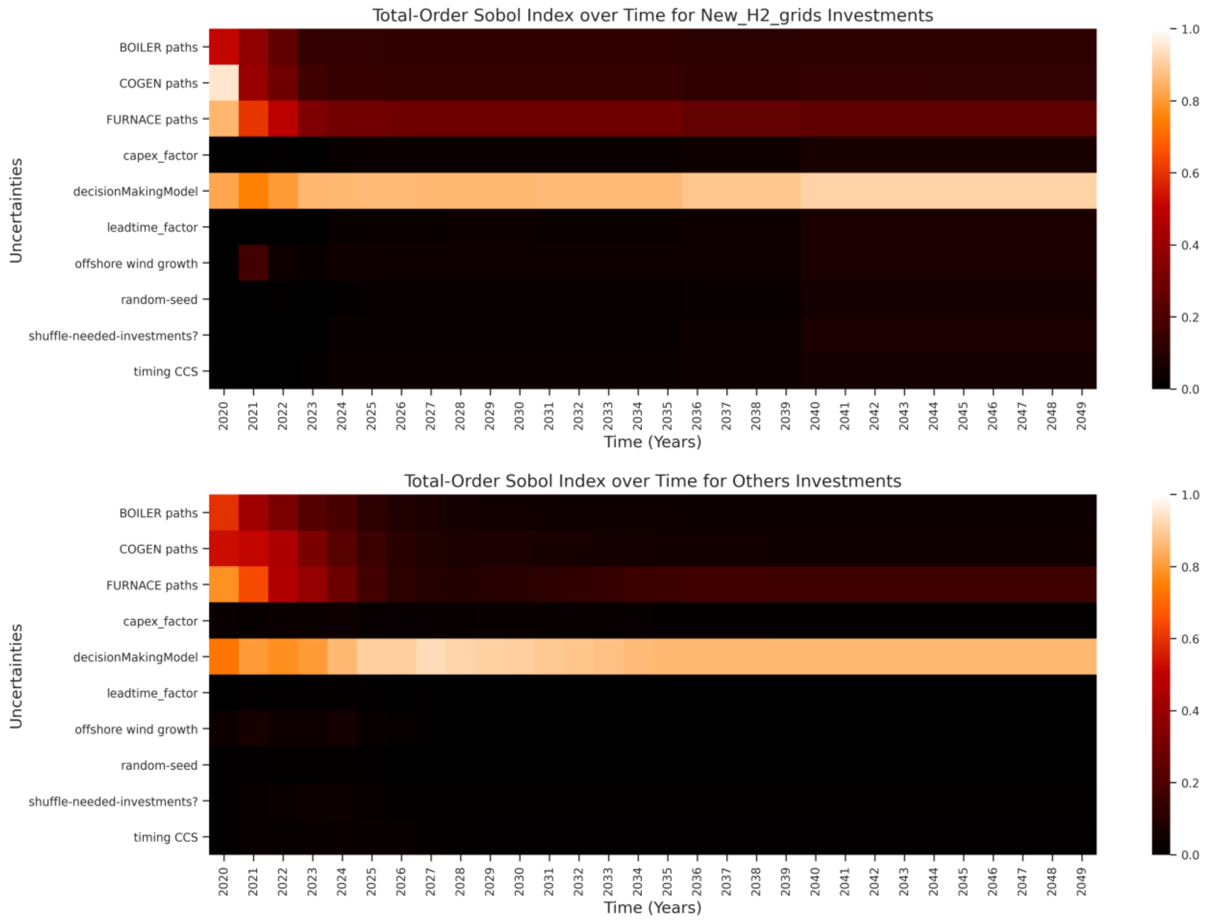
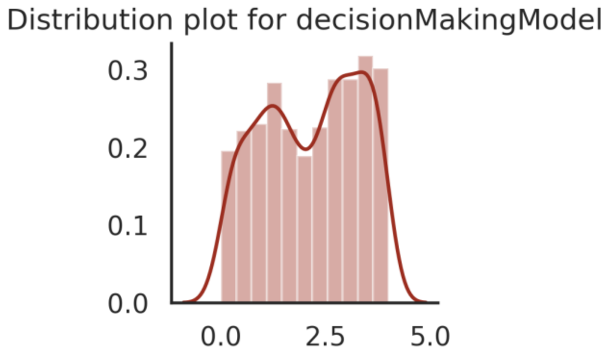
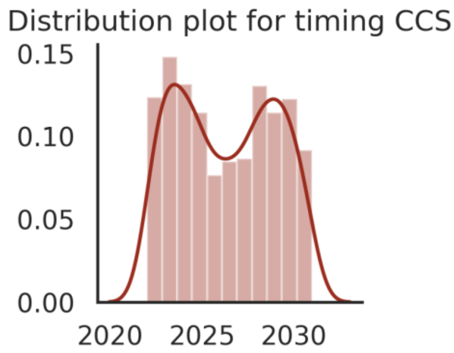


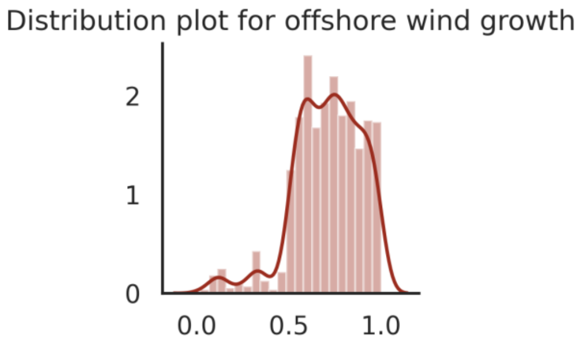
Figure 44 **Total-order sensitivity indices for KPI's and investment categories**

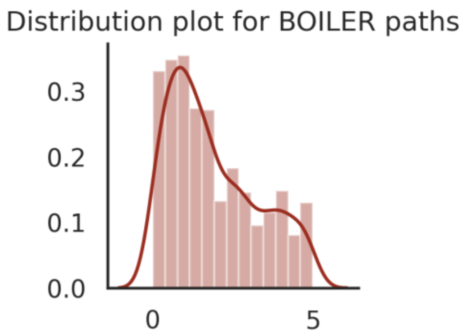
C.

DREAM


Distribution plot for decisionMakingModel
0.3
0.2
0.1
0.0
0.0
2.5
5.0


Distribution plot for timing CCS
0.15
0.10
0.05
0.00
2020
2025
2030


Distribution plot for offshore wind growth
2
1
0
0.0
0.5
1.0


Distribution plot for BOILER paths
0.3
0.2
0.1
0.0
0
5

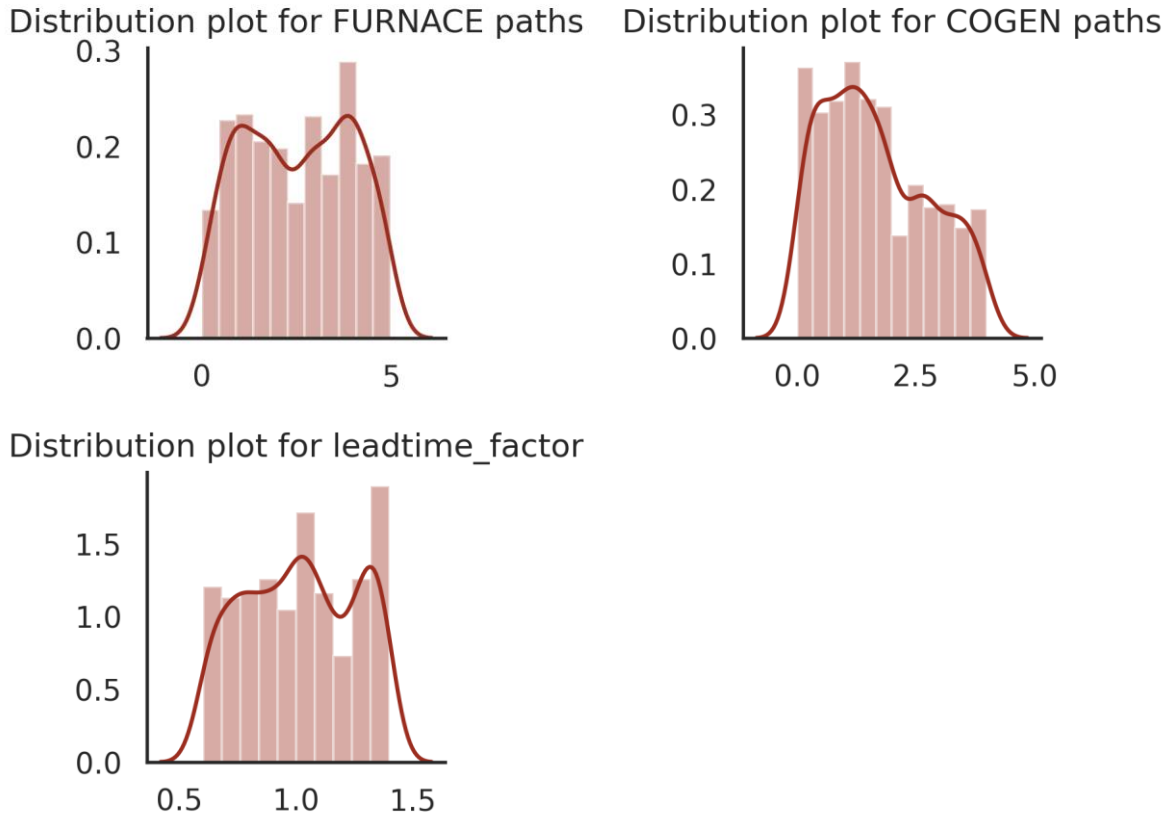


Figure 45 **Distribution plot of the raw sampled data from the DREAM algorithm.**

D.

CODE

The code for the analysis on the Windmaster model is available on GitHub:

https://github.com/alexanderdrent/Uncertainty-Analysis-Windmaster