Quantifying Resilience under Deep Uncertainty

Patrick Steinmann

Propositions

1. The use of single metrics will not progress our research on resilience. (this thesis)
2. Scenarios for decision support should be based on plausibility, not probability. (this thesis)
3. Discoveries should not be named after their discoverer(s).
4. The most significant questions in artificial intelligence research can only be answered by philosophers.
5. The widespread availability of Large Language Models such as ChatGPT will increase the rate of societal in-person interactions.
6. Uncertainties and simplifications should be emphasized more when communicating scientific advances to the general public.

Propositions belonging to the thesis, entitled

Quantifying Resilience under Deep Uncertainty

Patrick Steinmann

Wageningen, 30 January 2024

Quantifying Resilience under Deep Uncertainty

Patrick Steinmann

Thesis committee

Promotor
Prof. Dr J. Molenaar
Professor of Applied Mathematics
Wageningen University & Research

Co-promotor
Dr G. A. K. van Voorn
Associate Professor, Mathematical and Statistical Methods
Wageningen University & Research

Other members
Prof. Dr I. N. Athanasiadis, Wageningen University & Research
Prof. Dr G. J. Hofstede, Wageningen University & Research
Prof. Dr V. A. W. J. Marchau, Radboud University, Nijmegen
Dr E. Thompson, University College London, United Kingdom

This research was conducted under the auspices of the Graduate School for Production Ecology & Resource Conservation.

Quantifying Resilience under Deep Uncertainty

Patrick Steinmann

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A. P. J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 30 January 2024
at 4 p.m. in the Omnia Auditorium.

Patrick Steinmann
Quantifying Resilience under Deep Uncertainty,
140 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summary in English

ISBN 978-94-6447-991-1
DOI 10.18174/642842

Summary

Human society is dependent on a wide variety of complex systems, such as drinking water supply, communications infrastructure, and food logistics networks. These systems have become increasingly interconnected and -dependent with each other and society itself. Due to ongoing processes such as anthropogenic climate change, land use change, and digitization, they also face unprecedented levels of stress and disruption. Resilience - the ability of a system to withstand or recover from a disturbance - has been proposed as a framework for preparing systems for future disturbances. However, it is unclear how resilience can be quantified under conditions of so-called deep uncertainty - if the analyst does not know, or the stakeholders cannot agree on, the appropriate metric(s) to use for quantification. In this thesis, I examine how resilience can be quantified under deep uncertainty along two lines of research.

The first line of research, covered in Chapters 2, 3, and 4, investigates how resilience could be quantified if the uncertainty is present because there are many available metrics. In Chapter 2, we conduct a systematic scoping review of the peer-reviewed academic literature on resilience metrics for socio-technical and -ecological systems. We identify a number of resilience metrics, and classify a subset of them into ten distinct categories based on their underlying conceptual approaches. We also document what types of disturbances were considered, identifying four distinct types, including one not described in previously published disturbance frameworks. Finally, we study whether socio-ecological systems were investigated using a so-called "ecological resilience" approach, and socio-technical systems with a so-called "engineering resilience" approach, and vice versa. We find that the engineering resilience approach, which is conceptually and technically simpler, was more commonly used than the ecological resilience approach, even when studying socio-ecological systems.

In Chapter 3, we apply a variety of different resilience metrics and disturbances to a set of stable patterns in the Game of Life, a simple cellular automaton. We identify several predictive features of resilience, such as population size, the number of connected components, and pattern density. However, we also observe that the different resilience metrics rarely agree with one another, and that no pattern is highly resilient across multiple different disturbances.

In Chapter 4, we investigate whether using an ensemble of resilience metrics, rather than just a single metric, can improve the ability of a complex system to respond to disturbances. As in the previous chapter, we investigate this by applying a variety of metrics and disturbances to a nonlinear resource-consumer system. We find that conceptually distinct metrics sometimes behave quite similarly to one another, giving comparable resilience scores across different disturbances.

However, it is not possible to identify these similarities *ex ante*. We also find that it is not only possible to optimize a system to satisfy multiple resilience metrics for a given disturbance simultaneously, but that a system optimized in such a way is also more resilient to other disturbances, compared to a system optimized with just a single resilience metric.

The second line of research, covered in Chapters 5 and 6, investigates how resilience might be quantified if the uncertainty is present because there are no resilience metrics available at all. In the absence of quantitative metrics, we focus on qualitative scenario-based methods for exploring a system's behavior and vulnerabilities, with the intent of using these scenarios to identify analyst and stakeholder objectives, as well as decision-relevant dynamics and vulnerabilities, which may inform the future selection of resilience metrics. In Chapter 5, we introduce a novel method for identifying dynamic scenarios in a simulation model's time series outputs using time series clustering, and linking the identified clusters to underlying drivers using multi-class scenario discovery.

In Chapter 6, we present a second method for generating scenarios with a simulation model. Going beyond the mere analysis of outputs used in the previous chapter, we explore the targeted search for scenario sets using many-objective optimization. We compare our developed method against three other approaches, including conventional scenario axes techniques, and a static version of the clustering approach from Chapter 5. We find that our targeted search approach performs better than or equally to the other methods across three different criteria, and performs best overall, producing scenario sets that are maximally diverse, plausible, and comprehensive.

This thesis adds important knowledge in the fields of resilience and decision making under deep uncertainty. Advances at the intersection of these fields may result in more resilient socio-technical and -ecological systems, and ultimately a planet better prepared for an uncertain and volatile future.

Contents

Chapter 1

General Introduction

1.1 A Society of Systems

Modern society is dependent on a wide range of complex systems to ensure its continued functioning. From sending a text message to charging a phone, from commuting to grabbing a bottle of milk off a supermarket shelf - behind all these moments of daily life sits an array of systems working hard to create and deliver what we often take for granted. Many of these systems are closely interconnected, allowing them to be more efficient and effective. As the world population and global living standards have increased, tighter interconnections between the various social, technical, and ecological systems have become necessary to supply the resources for consumption, growth, and development (Helbing, 2013). These systems are now often referred to as socio-technical and socio-ecological systems, respectively, to highlight their integrated nature, and the fact that they have co-evolved - as society has shaped the built and living environment, those environments have shaped society. We are bound to these systems, as they are bound to us, for continued existence.

However, many of our planet's systems are experiencing unprecedented levels of stress and disruption. Over the past decades, anthropogenic climate change has increased the frequency and intensity of weather events such as droughts and extreme precipitation (Intergovernmental Panel On Climate Change, 2023). Global development and land use change have increased our exposure to non-climate hazards such as earthquakes and tornadoes (Smith & Katz, 2013; Weinkle et al., 2018). The increasingly blurry boundary between physical and digital worlds has introduced new modes of failure and disruption into our lives (Renn et al., 2022). All these factors pose grave challenges to the systems keeping us alive, fed, and happy.

Understanding how socio-technical and -ecological systems, from local water supply networks to global financial schemes, respond to disturbances is therefore a matter of paramount scientific and societal importance. How vulnerable are we to floods, ransomware, or pandemics? How will we respond to a solar storm, riot, or economic depression? What should we do to improve our capacity to handle mass migration, hybrid warfare, or earthquakes? These are urgent questions of local and global policy alike.

1.2 Resilience and Uncertainty

Resilience has been proposed as a framework for preparing our socio-technical and -ecological systems for current and future hazards. While the term has evolved through multiple generations of conceptual interpretation (Folke, 2006; Holling, 1973, 1996) across a variety of fields of science and practice, it is commonly understood to represent the ability of a system to withstand or recover from a disturbance (Walker et al., 2004). Exact definitions of the term abound, and it is questionable whether a single unifying definition could or even should be identified, as the term is also often used as a bridging or boundary object to facilitate understanding between different disciplines (Brand & Jax, 2007). Nevertheless, there is broad agreement that resilience is at least a desirable, if not necessary, component of global adaptation and development. Many international organizations which concern themselves with policy and investment see increasing resilience as a key feature of their approaches, including the World Bank, the United Nations Development Programme, and the North Atlantic Treaty Organization.

Measurement is a key challenge in the application of resilience. To date, there is no universally agreed-upon approach for translating the conceptual idea of resilience, the ability of a system to recover from a disturbance, into some quantifiable property of real-world systems. Dozens of metrics, indicators, and quantifications of resilience have been proposed across many different fields of science and practice (see e.g. Hosseini et al. (2016), Quinlan et al. (2016), and Sun et al. (2020)). The plethora of available resilience metrics may be a direct result of the multidisciplinary heritage and evolution of the term itself. However, this wealth of metrics has not made quantifying resilience any easier, on the contrary, it has introduced a significant uncertainty into the analytical process: which metric should be chosen for a given analysis?

The uncertainty of metric choice stems from two distinct sources. Firstly, resilience is not an inherent property of natural or man-made systems, it is a label we apply to a certain kind of macro-scale behavior of complex, multidimensional systems (Park et al., 2013) which is beneficial or positive in some broader sense. Thus, there is no single underlying attribute of those systems we can uniquely identify as being “the resilience” thereof (Carpenter et al., 2001; Cutter, 2016; Meerow & Newell, 2019), no matter how much research we do. Secondly, and more importantly, many of the socio-technical and -ecological systems we rely on are not controlled by a single actor, but exist in a web of shared and contested governance between different stakeholders with various perspectives, goals, and needs (Gotts et al., 2019). It is therefore likely that a single measurement approach will never have the required legitimacy among all stakeholders, or align with their diverse objectives. In this sense, the measurement of resilience in socio-technical and -ecological systems fulfills multiple criteria of being a wicked problem (Rittel & Webber, 1973), primarily the lack of a definitive problem formulation, the sensitivity to framing, and the necessity of correctness.

1.3 Research Question and Approach

In this thesis, I investigate approaches to dealing with the uncertainty surrounding the choice of metric when assessing the resilience of socio-technical and -ecological systems. The following question has guided my research: **How can we quantify resilience if we are uncertain about which metric to use?**

The conceptual framework underpinning my research is that of Decision Making under Deep Uncertainty, or DMDU. Deep uncertainty exists "when analysts do not know, or the parties to a decision cannot agree on, (1) the appropriate models to describe the interactions among a system's variables, (2) the probability distributions to represent uncertainty about key variables and parameters in the models, and/or (3) how to value the desirability of alternative outcomes." (Lempert et al., 2003). These three elements are sometimes referred to as structural, parameter, and metric uncertainty, respectively. The existence of these uncertainties is not necessarily novel, drawing upon literature at least as far back as Knight's (1921) work. However, the advent of high-performance computing, and the increasing acceptance of computational methods in science (Winsberg, 2010), have opened the door to a new and improved treatment of these uncertainties. Because uncertainty-based methods are especially suitable for dealing with wicked, multi-stakeholder problems (Funtowicz & Ravetz, 1994), they are a natural fit for quantifying resilience.

In conventional approaches to modelling complex systems, uncertainties are often dealt with through simplification and aggregation. For example, when modelling the spread of an infectious disease, a population-level contact rate might be used to describe how often individuals meet each other, rather than dealing with the uncertainty of how often every unique individual encounters someone. In effect, this homogenizes all individuals into a single population. But what if this aggregation is obscuring important behavior differences between people of different age, or with different family structures? Establishing the relevance of such questions, and the resulting trustworthiness of such aggregated models for decision support, is difficult (Walker et al., 2003). The conceptual keystone of Decision Making under Deep Uncertainty therefore is that analytical decisions such as simplification and aggregation should take place once the model has been completed, and the implications of the interacting uncertainties have been computationally evaluated, rather than during model creation. In practice, this requires that ensembles of alternative hypotheses and structures capturing the uncertainties are integrated into the model during construction. While this does require a different approach to making such models, the benefits, such as improved system understanding, more robust policy insights, and more reproducible research, may be worth the effort (Auping, 2018). This computational ensemble approach to uncertainties is commonly referred to as Exploratory Modelling & Analysis, or EMA (Bankes, 1993). In this thesis, I apply exploratory modelling to a variety of complex system models and resilience metrics to understand how the models behave under a variety of conditions, how the metrics interact with the models, and how different metrics compare.

1.4 Thesis Outline

DMDU researchers and practitioners commonly distinguish different levels of (un)certainty ranging from (near) certainty to total ignorance. Adapted from Walker et al. (2013) and Marchau et al. (2019), the levels are described in Table 1.1 in a generic sense.

Level	Name	Description
1	Clarity	The truth is known.
2	Probability	Likely knowledge of the truth.
3	Ranking	Multiple alternatives, some preferable, could be true.
4a	Multiplicity	Multiple, unrankable alternatives could be true.
4b	Ignorance	The truth is totally unknown.

Table 1.1: Levels of uncertainty

In this thesis, I focus exclusively on uncertainty levels 4a and 4b, which constitute deep uncertainty (Walker et al., 2013). While the different levels can be applied to a variety of system elements such as parameters, structure, outcomes, or weights (Marchau et al., 2019), in this thesis, I apply them to resilience metrics. In this context, the levels may be interpreted as conditions of many plausible metrics for quantifying a system's resilience being available, and no metrics being available, respectively. Accordingly, a two-pronged approach is necessary - one line of research investigating how a multiplicity of resilience metrics may affect decision making, and another line examining what to do if no resilience metrics are known (or agreed upon) at all.

The first three content chapters of this thesis deal with level 4a uncertainty, or a multiplicity of plausible resilience metrics. Chapter 2, *Resilience Metrics for Socio-Ecological and Socio-Technical Systems: A Scoping Review*, is a systematic scoping review of the peer-reviewed literature on quantifying the resilience of socio-technical and -ecological systems. We use a reproducible methodology to identify a wide range of resilience metrics, and attempt to classify them. We also investigate what kinds of systems they were applied to, and which kinds of disturbances these systems experienced. This chapter sets the foundation for the following chapters, in which different resilience metrics are compared and evaluated. In Chapter 3, *Resilient Life: An Exploration of Perturbed Autopoietic Patterns in Conway's Game of Life*, we apply a variety of disturbances to a catalogue of cellular automata, and quantify the responses using multiple metrics. This chapter demonstrates that even for simple systems, quantifying resilience is sensitive to the experienced disturbance and metric, although some general patterns can be identified. In Chapter 4, *Robust Resilience: Optimization with Ensembles of Metrics May Improve Resilience to Novel Shocks*, we study the responses of a simple resource-consumer model to a variety of disturbances, and show that using an ensemble of resilience metrics in the optimization process may improve the system's resilience to disturbance it was never optimized for.

The second line of research, dealing with level 4b uncertainty, investigates what to do if there is no knowledge or agreement about applicable resilience metrics. In the absence of any quantitative metrics, only qualitative methods for exploring resilience remain. One such approach is using scenario-based methods to identify decision-relevant dynamics and vulnerabilities, which may inform the future selection of resilience metrics. In the fourth and fifth chapters, I present two methods for generating sets of scenarios. Chapter 5, *Behavior-based Scenario Discovery Using Time Series Clustering*, describes a novel method for exploring and summarizing the possible future behavior over time of a complex system, using data from an integrated energy-security simulation model. This chapter shows that, while a model's dynamics may be highly diverse, regularities can be identified and linked to underlying drivers and uncertainties. Chapter 6, *Scenario Search: Finding Diverse, Plausible and Comprehensive Scenario Sets for Complex Systems*, presents a novel method for finding small scenario sets which optimally describe a simulation model's plausible behavior range, and compares it against three previously described methods.

Finally, in the *General Discussion*, I tie together common themes from the five content chapters, and discuss resulting insights. Based on these insights, I make recommendations for practitioners, and highlight potential future research directions for quantifying resilience under deep uncertainty.

Chapter 2

Resilience Metrics for Socio-Ecological and Socio-Technical Systems: A Scoping Review

Submitted as: Steinmann, P., Tobi, H., and van Voorn, G.A.K. Resilience Metrics for Socio-Ecological and Socio-Technical Systems: A Scoping Review.

2.1 Abstract

An increased interest in the resilience of complex socio-ecological and -technical systems has led to a variety of metrics, quantifications and indicators being proposed. An overview of these metrics and their underlying concepts would support identifying useful metrics for applications in science and engineering. This study undertakes a scoping review of resilience metrics for systems straddling the societal, ecological, and technical domains to determine how resilience has been measured, the conceptual differences between the proposed approaches, and how they align with the domains of their case studies. We find that a wide variety of resilience metrics has been proposed in the literature. Conceptually, 10 different quantification approaches were identified. Four different disturbance types were observed, including sudden, continuous, multiple, and abruptly ending disturbances. Surprisingly, there is no strong pattern regarding socio-ecological systems being studied using the "ecological resilience" concept, and socio-technical systems being studied using the "engineering resilience" concept. As a result, we recommend that researchers use multiple resilience metrics in the same study, ideally following different conceptual approaches, and compare the resulting insights. Furthermore, the used metrics should be mathematically defined, the included variables explained, and the chosen functional form justified.

2.2 Introduction

2.2.1 Background

Humanity is dependent on a variety of socio-ecological and socio-technical systems (SES and STS, respectively) to supply critical resources. Examples include agricultural food production systems, energy conversion and distribution infrastructure, and transportation networks. These systems are becoming more interconnected and -dependent (Helbing, 2013), rendering them more susceptible to disruptions as failures in systems can affect nominally separate systems (Filatova et al., 2016), or cascade across organizational levels to affect much smaller or larger parts (Iwanaga et al., 2022).

To ensure the continued functioning of SES and STS in the face of disturbances such as droughts, pandemics, or climate change, resilience has been identified as a desirable property (Arrow et al., 1995; Folke et al., 2004; Rapport, 1989; Walker et al., 2004). While the term has been interpreted in a variety of ways across time and disciplines (Brand & Jax, 2007; Nilsson & Grelsson, 1995), the shared underlying concept is that a system, after experiencing a disturbance, should be able to recover to some acceptable performance level or configuration within a useful time frame. The recovery may be achieved through a variety of mechanisms, including redundancy, buffers, evolution, or learning (Biggs et al., 2012; Desjardins et al., 2015).

The nature of the system's recovery is dependent on its specific dynamics. Broadly speaking, there are two possible options: either the system reaches the performance level it was functioning at prior to the disturbance, or it reaches some other acceptable performance level. Mathematically, we can describe these two options as the system having either a single or multiple basins of attraction. These basins may also be referred to as attractors or steady states. Holling (1996) describes the two alternatives as "engineering resilience" and "ecological resilience", respectively, and these terms are widely used in the resilience literature to describe the two possible dynamics. Figure 2.1, originally by Liao (2012), illustrates the engineering and ecological resilience concepts.

A key issue in the study of systems resilience is the translation of the concept of resilience into a measurable system property (Egli et al., 2019). For example, Holling (1996) proposed measuring return speed to equilibrium for "engineering resilience", and absorption capacity before shifting performance levels for "ecological resilience". A wide variety of other resilience metrics have been proposed in the literature, such as flow magnitudes (Ulanowicz et al., 2009) or differences between pre- and post-disturbance spatial patterns (Cika et al., 2020). The diversity of approaches has challenged our ability to effectively measure resilience (Klein et al., 2003). For existing overviews of various resilience measurement and assessment approaches, we refer to Quinlan et al. (2016), Hosseini et al. (2016) and Sun et al. (2020).

However, there has been little work on classifying resilience metrics to gain a more systematic understanding of different conceptual measurement approaches. Addressing this first knowledge gap would be useful for the gover-

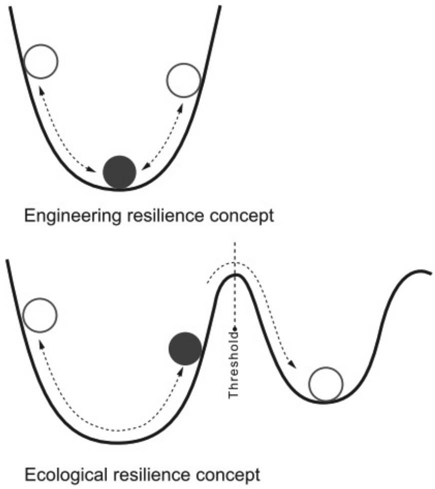


Figure 2.1: Engineering and ecological resilience concepts illustrated as cup-and-ball systems. Original figure by Liao (2012).

nance of modern SES and STS, as both scientists and policymakers are increasingly recognizing that these systems serve a variety of parties with diverse and potentially opposing needs (Gotts et al., 2019; Lempert et al., 2003). In such situations, having a clear understanding of what different parties might see as the "right" way to quantify their desired objective may facilitate successful governance by revealing aligned and opposing perspectives, asymmetries, and compromise potential (Gold et al., 2019). One important aspect that parties may disagree on is the need for resilience of the system towards different disturbances. Disturbances may differ in origin, timing and frequency, length, and more. For instance, Collins et al. (2011) summarize the possible disturbances into two categories: long-term sustained disturbances, and short-term pulse disturbances. The resilience against these different disturbances need not be the same, as resilience mechanisms differ. Increasing resilience against flooding likely entails building (higher) dikes, while this is not helpful against an increasing frequency of droughts. Depending on the resilience objective, actors may have different ideas about what to focus on, and hence what to measure.

A second gap in the literature is the correspondence between socio-ecological systems and "ecological resilience", and vice versa, socio-technical systems and "engineering resilience". As described earlier, the former are types of systems, the latter behavior patterns of such systems. While a close correspondence is suggested by their naming, it is unclear whether the pairings are observable in the published literature – that is, whether the resilience of socio-ecological systems is studied in the "ecological" sense of multiple basins of attraction, and the resilience of socio-technical systems in the "engineering" sense of a single basin of attraction.

In this article, we conduct a systematic scoping review of resilience metrics described in the peer-reviewed literature on socio-ecological and -technical systems, in order to address the two knowledge gaps identified above. In this first section, we introduce the topic and rationale, and outline our goals. In the second section, we protocol our search process, inclusion and exclusion criteria, data extraction, and synthesis methods. In the third section, we describe and synthesize our results. In the final section, we discuss our findings and their implications for future practice and research.

2.2.2 Objectives

The presented research is a systematic scoping review aimed at answering the following research questions:

1. Which metrics have been proposed to quantify the resilience of socio-ecological and -technical systems?
2. How do these metrics differ conceptually?
3. What types of disturbances have been used to study the resilience of socio-ecological and -technical systems?
4. How strictly are the concepts of engineering and ecological resilience applied to socio-ecological and socio-technical systems, respectively?

In answering the stated research questions, we make the following contributions to the methodological resilience literature:

- We conduct a systematic and reproducible scoping review of resilience metrics for socio-ecological and socio-technical systems.
- We summarize a number of conceptual approaches to quantifying resilience, and highlight which approaches were not represented, indicating potential research gaps.
- We describe two classes of system disturbances that are documented in case studies, but do not readily fit into known classifications of disturbances.
- We show how commonly socio-ecological systems are studied from an ecological resilience perspective, and correspondingly, how commonly socio-technical systems are studied from an engineering resilience perspective.

2.3 Methods

2.3.1 Protocol and Registration

The protocol for this scoping review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) scheme (Peters et al., 2020; Tricco et al., 2018), which is based on the

PRISMA scheme (Liberati et al., 2009) for systematic reviews. The protocol was not registered. Details are available from the corresponding author.

2.3.2 Eligibility Criteria

To be included in the study, papers needed to study the resilience of a system from the socio-ecological or -technical domains, and include some method for quantifying this resilience. Papers were included if they were peer-reviewed, published in English, and mentioned resilience and either socio-ecological or socio-technical systems in their title, abstract, or keywords. Papers were excluded if they discussed individual-level resilience in a psychological, clinical, or psychiatric context, or if they related to physical resilience in materials science, electrical engineering, or computer science. We also excluded non-primary sources such as reviews.

2.3.3 Information Sources

We identified potentially relevant sources through a search on Web of Science, performed on 2020-07-21, across four major databases (Science Citation Index, Social Science Citation Index, Arts & Humanities Citation Index, and Emerging Sources Citation Index).

2.3.4 Search

The final search query for Web of Science is presented in Table 2.1 as corresponding elements of the primary research question and query.

2.3.5 Selection of Sources of Evidence

To develop and refine the inclusion and exclusion criteria in the screening phase, random subsets (first round: N=15, second round: N=15, third round: N=30) from the cleaned query results were independently screened by all three authors. We discussed disagreements on selection and exclusion, and iteratively revised and clarified the criteria where necessary. For the third and final round, we reached consensus on 28 out of 30 (93%) abstracts regarding in-/exclusion.

2.3.6 Data Extraction

We jointly developed a data extraction form, iteratively refining it by independently applying it to selected articles and comparing the results. The final data extraction was performed by the lead author. Where data items were unclear, conclusions were drawn based on the metric used, as we considered this the focal point of our review.

Table 2.1: Web of Science search query elements.

Element of research question	**Element of query**
resilience	TS=(resilien*)
metric	AND TS=(metric* OR quantif* OR indicator* OR measure*)
socio-ecological or socio-technical	AND WC=(Agricultural Economics & Policy OR Agricultural Engineering OR Agriculture, Multidisciplinary OR Agronomy OR Engineering, Civil OR Management OR Engineering, Environmental OR Engineering, Industrial OR Area Studies OR Engineering, Multidisciplinary OR Materials Science, Textiles OR Mathematical & Computational Biology OR Environmental Sciences OR Environmental Studies OR Mathematics, Applied OR Mathematics, Interdisciplinary Applications OR Biodiversity Conservation OR Public Administration OR Public, Environmental & Occupational Health OR Fisheries OR Regional & Urban Planning OR Forestry OR Multidisciplinary Sciences OR Geosciences, Multidisciplinary OR Social Sciences, Mathematical Methods OR Green & Sustainable Science & Technology OR Health Policy & Services OR Statistics & Probability OR Computer Science, Interdisciplinary Applications OR History & Philosophy of Science OR Computer Science, Software Engineering OR Computer Science, Theory & Methods OR Operations Research & Management Science OR Transportation OR Transportation Science & Technology OR Demography OR Urban Studies OR Development Studies OR Ecology OR Water Resources OR Economics OR Limnology
system	AND TS=("system" OR ecosystem OR systems)

2.3.7 Data Items

For each article in our review, we extracted the following data items:

- System type: is the system socio-ecological or socio-technical?
- Disturbance: what disturbance does the system experience?
- Basins of attraction: does the system have one or multiple basins of attraction?
- Resilience metric: what metric for resilience is used?

2.3.8 Synthesis of Results

For our first and second research questions, we identified the underlying conceptual system properties the different metrics considered, and used these to classify the metrics. For our third research question, we identified and classified the disturbance(s) considered in each paper. For our fourth research question, we established for each paper whether it was more ecologically or technically inclined, and compared this with the number of basins of attraction the studied system could reach. The last two questions required some interpretation of both the metric and case study. Where necessary, this interpretation was discussed and agreed upon by all authors.

2.4 Results

2.4.1 Sources of Evidence

In total, 6743 abstracts were retrieved from Web of Science. Of these, 88 were excluded for missing metadata. From the remaining 6385 articles, a subset of 551 was selected for abstract-based screening. We generated this subset by first stratifying the query results in 5-year intervals, and then selecting 107 random papers from each stratum. For strata with less than 107 papers total (1990-1994, 1995-1999), we included all papers. For the 2020-2024 stratum, we selected 21 papers (20% of 107) to account for the one year of the stratum which was ongoing at the time the final search was performed.

Of the screened abstracts, 471 were excluded for either being about resilience in a different context (healthcare, materials science, or electrical engineering), due to being a non-primary source (e.g. a review), for not being about some form of socio-ecological or -technical system (e.g. a purely ecological study of fish populations in alpine lakes), or for not focusing on resilience (e.g. a paper motivating why persistence is distinct from resilience).

Of the remaining 80 articles, one was not accessible to the authors. The remaining 79 were read in full. Of those, 32 were excluded for not explicitly stating a resilience metric, and 6 were excluded for not focusing on resilience. The remaining 41 articles were included in the presented analysis. This entire workflow is visualized in Figure 2.2.

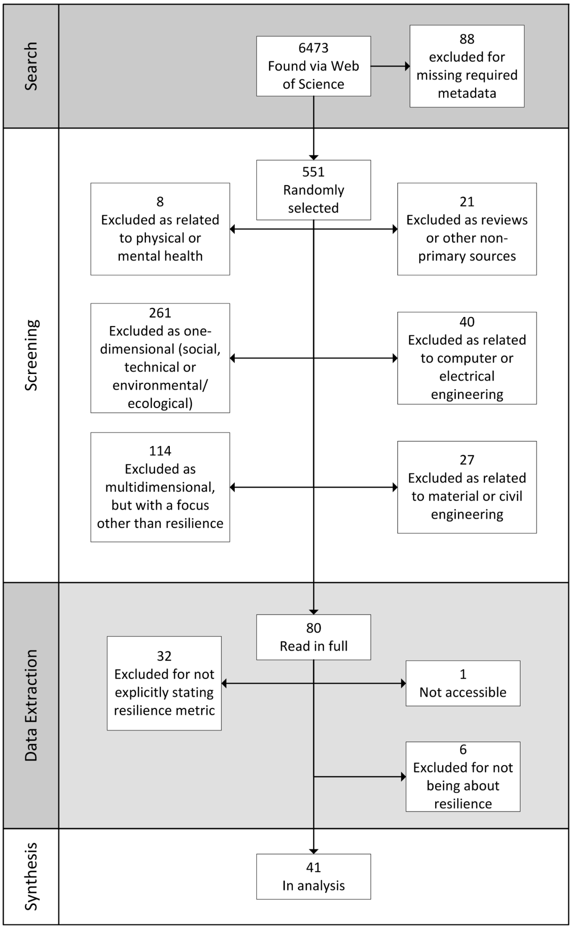


Figure 2.2: PRISMA-ScR workflow.

2.4.2 Characteristics of Sources of Evidence

In Table 2.2, we give a demographic overview of the sources considered in our scoping review. Please note that many papers are interdisciplinary and therefore fit multiple research areas (based on Web of Science's categorization).

Table 2.2: Demographics of screened, read, and included papers.

		Screened (N=551)	Read (N=80)	In synthesis (N=41)
Publication year	1990-1994	31	5	3
	1995-1999	71	15	7
	2000-2004	107	19	11
	2005-2009	107	13	7
	2010-2014	107	13	4
	2015-2019	107	9	4
	Jan 2020-July 2020	21	6	5
Research area	Arts & Humanities	0	0	0
	Life Sci & Biomed	593	67	22
	Physical Sciences	154	34	19
	Social Sciences	95	13	4
	Technology	242	43	31

2.4.3 Results of Individual Sources of Evidence

A list of the included individual sources of evidence may be found in Appendix A, and the results of the data extraction in Appendix B.

2.4.4 Synthesis of Results

Resilience Metrics

We identified 46 resilience metrics in the 41 reviewed papers. Of these, 34 were defined as mathematical functions, and 12 described verbally. Among the mathematically defined metrics, we observed a number of functional forms, including fractions, limits, sums, piecewise definitions, and probabilities. Additionally, in one verbally described metric, a trigonometric function is mentioned. Various degrees of mathematical complexity are apparent, from fractions with two variables to piecewise definitions with a dozen variables. A variety of "corrective" elements, used to coerce the output of a function to some desired range or direction, can be observed. Examples include subtraction (e.g. metric #20A), inversion (e.g. metric #4), and piecewise definition (e.g. metric #16). Despite these coercions, we observe both minimization and maximization criteria among the metrics, i.e. some metrics represent "higher resilience" as values closer to 0 (e.g. metric #17), and some as values as large as possible, potentially with an upper bound (e.g. metric #6). The reasoning behind the specific functional forms is rarely explained.

A number of letters and symbols, such as R, P, and γ, appear in multiple metrics, but with different meanings. For example, the letter R is used, in upper- or lowercase, to describe the time elapsed from the beginning to the end of a disturbance (metric #9), the incurred loss of performance (metric #18), a performance standard (ibid), the time index during a simulation run (metric #22), the resilience of an entire system (e.g. metric #25), the resilience of an individual node in a system (metric #26B), and as a coefficient of determination (metric #31A). There is no commonality in the notation. In other words, for every metric, care must be taken to identify every variable's exact meaning in the context of that given paper.

Conceptual Approaches to Quantifying Resilience

When moving beyond the mathematical implementations towards the conceptual approaches used to capture the system's resilience, we observe that some concepts appear in multiple metrics. Examples include the return time of the system to a previous performance level after a disturbance (e.g. metrics #2, #3, and #5), the total performance loss incurred due to the disturbance(s) over time (e.g. metrics #13, #25, and #29), and the largest momentary performance loss due to the disturbance(s) (e.g. metrics #9, #18, and #15). This implies that there are some common ideas about which attributes of a system and its behavior describe its resilience. In this section, we attempt to summarize these conceptual approaches.

Out of a total of 46 metrics, we identified 37 metrics with a single basin of attraction, and 9 metrics which considered multiple basins of attraction. Furthermore, we identified 27 metrics which we consider generic in that they could easily be applied to other systems and disturbances (much like Holling's return speed and absorption capacity described earlier). As a counter-example, consider metric #10, which uses the fecundity and mortality probabilities of different species in a trophic network to quantify the long-run resilience of the entire network - applying such a metric to an urban water supply network would be difficult to justify. Generic metrics are especially interesting because they facilitate cross-comparison, making resilience analyses more informative (Quinlan et al., 2016). We therefore limit the following analysis to generally applicable metrics for systems with a single basin of attraction, although we do distinguish between single- and multi-disturbance metrics, as we feel they represent distinct schools of thought on quantifying resilience.

Among the generic single-disturbance single-basin metrics, we identified six conceptual approaches that could easily be generalized, visualized in Figure 2.3. These are:

1. Return time to previous performance level (three metrics: #2, #3, #30A)

2. Total performance loss (six metrics: #13, #25, #29, #36, #38, #39)

3. Combination of maximum performance loss and recovery time (two metrics: #18, #30B)

4. Combination of relative performance loss and return time (one metric: #9)
5. Combination of return time to previous performance level with oscillations, and amplitude of performance (one metric: #15)
6. Return time to previous performance level with oscillations (one metric: #17)

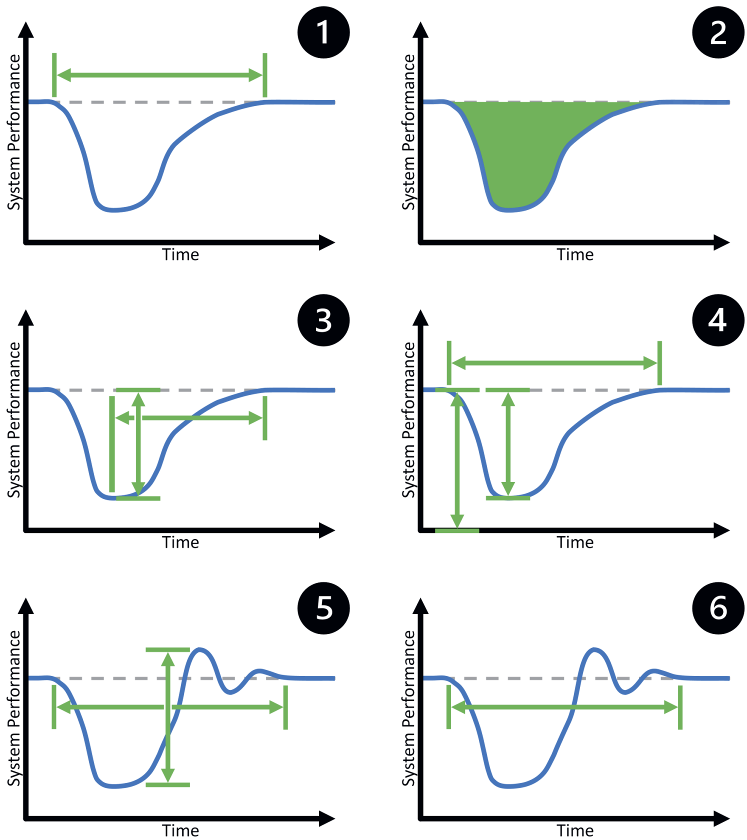


Figure 2.3: Conceptual approaches to quantifying resilience for single basin of attraction, single disturbance systems.

Among the multi-disturbance single-basin metrics, we identified four conceptual approaches, visualized in Figure 2.4. These are:

7. Total time of insufficient performance (8 metrics: #5, #7, #11, #16, #22, #23A, #28, #35)
8. Total performance loss (one metric: #34)
9. Longest period of insufficient performance (three metrics: #1, #19, #23B)

10. Total time spent outside performance range (one metric: #6)

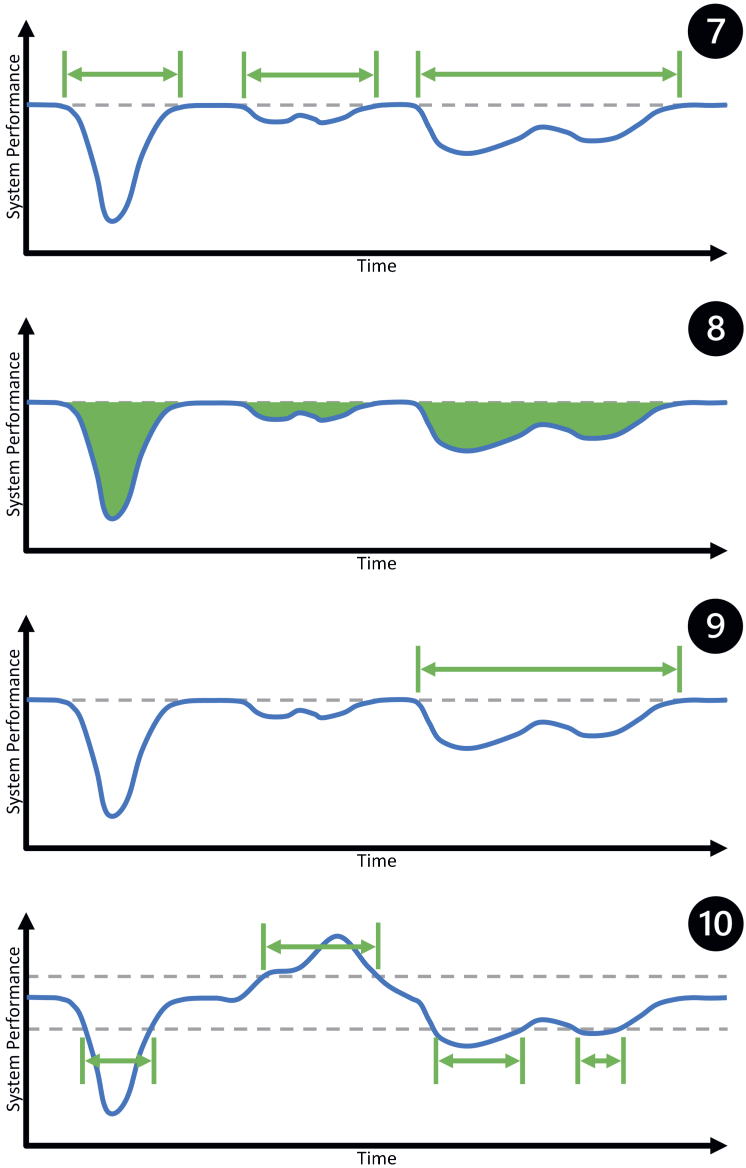


Figure 2.4: Conceptual approaches to quantifying resilience for single basin of attraction, multiple disturbance systems.

System Disturbances

In the 41 papers, we identified four distinct types of disturbances. These include 15 cases (17 metrics) of a sudden disturbance, such as an earthquake, 7 cases (8 metrics) of a continuous disturbance, such as drought, 15 cases (17 metrics) where multiple disturbances were observed, such as a multi-year period of repeated flooding, two cases (two metrics) where a continuous disturbance abruptly ends, such as a ban on fishing after a period of intense fishery activity, and two cases (two metrics) where the disturbance was not specified. Please note that in all papers that included multiple metrics, both metrics were applied to the same case study and disturbance.

Alignment between SES and Ecological Resilience, and STS and Engineering Resilience

Among the 46 metrics included in this review, we identified 37 as having a single basin of attraction, thus subscribing to the "engineering resilience" concept described by Holling (1996). A further eight metrics accommodated multiple basins of attraction, in line with what Holling called "ecological resilience". The case studies these metrics belonged to comprised 15 studies of socio-ecological systems, such as fisheries or managed forests, and 31 studies of socio-technical systems, such as water reservoirs or logistics networks.

When cross-tabulating the number of basins of attraction of each metric with the nature of the case study the metric was applied to (see Table 2.3), we note that there is no obvious pairing of socio-ecological systems with "ecological resilience", and socio-technical systems with "engineering resilience". In fact, both socio-ecological and -technical systems are studied using single and multiple basins of attraction. Socio-ecological systems do have a higher likelihood of being modelled with multiple basins of attraction (5/15 (33%) vs. 4/31 (13%)). Considering that there are many more socio-technical case studies, the metrics with multiple basins of attraction are quite evenly split between socio-ecological and -technical systems (5/9 (55%) vs 4/9 (45%)).

Table 2.3: Comparison of case study and metric types.

	Single basins of attraction	Multiple basins of attraction	Total
Socio-technical case study	27	4	31
Socio-ecological case study	10	5	15
Total	37	9	46

2.5 Discussion

2.5.1 Summary of Evidence

Through this scoping review, we found 41 articles containing 46 resilience metrics for socio-ecological or -technical systems. These sources, identified through an iteratively refined and rigorous search process, span three decades of research, and a wide variety of research areas.

While all metrics purport to quantify the same conceptual idea – the response of a system to disturbance – they do this in a variety of ways. While most papers include a mathematical definition of the employed metric(s), some papers rely exclusively on a verbal definition. Among the mathematically defined metrics, a diverse range of elements, including piecewise definitions, trigonometry, inversions, and limits can be observed. Some of these are used to coerce the metric to a particular numerical range or direction. However, these coercions are often nonlinear, potentially biasing the resulting analysis without stakeholders realizing it (Jain, 2009). Furthermore, we notice that the metrics are used as both maximization and minimization criteria. Authors rarely discuss why a particular functional form was chosen, and whether alternatives were explored, although a small number of papers do include multiple resilience metrics applied to the same case study. Finally, notation is inconsistent, with common letters such as P or R being used to represent a variety of different elements.

When looking past the mathematical implementation at the underlying conceptual ideas, we observe 10 distinct concepts, including six concepts for systems with a single basin of attraction experiencing a single disturbance, and four concepts for systems with a single basin of attraction experiencing multiple disturbances. We did not study the underlying concepts behind metrics for systems with multiple basins of attraction, as the metrics were too few, and too diverse, to categorize. This is surprising because the existence of multiple basins of attraction is a well-known and typical characteristic of socio-ecological systems (Gunderson, 2010; Ludwig et al., 1978).

The 41 papers included in this review use four distinct types of disturbances in their case studies. These include sudden disturbances, continuous disturbances, repeated or multiple disturbances, and suddenly ending disturbances. For two papers, the disturbance type could not be identified. We observe that two of these types, the repeated/multiple disturbances and the suddenly ending disturbances, do not fit into the categorizations published by Lake (2000), who distinguishes short-term pulses, long-term constant presses, and long-term increasing ramps, or Collins et al. (2011), who distinguish sustained press and short-term pulse disturbances.

2.5.2 Limitations

There is a tremendous body of literature on resilience. It is therefore almost certain that we have missed some approaches to quantifying resilience. We limited ourselves to peer-reviewed literature, excluding a wide array of grey literature.

While this increases the credibility of our source material, it also increases the likelihood that we missed some unique approach to quantifying resilience. We also only included literature that explicitly uses the word "resilience", although Grimm and Wissel (1997) highlight that a diverse range of terminology is used in studies of ecological stability. Furthermore, we conducted our search using a single scientific database, which is, like every other database, known to be incomplete.

Finally, this scoping review was an enormous undertaking, and our results are thus only up to date as of July 2020. We nevertheless believe our results are informative.

2.5.3 Recommendations

Based on the analysis and discussion presented above, we make the following two recommendations for researchers working with resilience metrics.

Firstly, as there is such a conceptual and mathematical diversity of methods for quantifying resilience, we recommend that researchers use multiple conceptually distinct metrics to quantify resilience, and compare the resulting insights. This will improve the robustness of the analysis by reducing the risk of blind spots introduced by a metric with a narrow focus. For example, a metric using only the return time to a previous performance standard will be oblivious to how the system recovers its performance, while a metric measuring just the total performance loss will not capture how long it took to regain the original performance level. Using these metrics in concert could thus lead to a more holistic understanding of the system's resilience. Additionally, using multiple metrics in an analysis also offers an exciting opportunity to engage with stakeholders about their perception(s) of resilience, and which metric(s) are best able to capture their desired outcomes. This transparency will help move resilience assessments away from using metrics that are available/known, and towards metrics that are useful and fit for purpose (Ivory & Stevenson, 2019).

Secondly, we recommend that researchers be explicit about which resilience metric(s) they use, both conceptually and mathematically. The mathematical definition of the metric should be given as a formula, the composition of which should be justified. In addition, the variables should be explained, including units, and ranges should be given. This will greatly increase the reproducibility and reusability of the conducted research, both being serious concerns in modern scientific research (Baker, 2016).

2.5.4 Future Research

Based on our scoping review, we identify three promising directions of future research. Firstly, it may be useful to expand the search query, taking into account the most recent published literature, and potential (near) synonyms of resilience such as fragility (Nilsson & Grelsson, 1995) or danger (Bergström et al., 2015). This may identify further conceptual approaches to quantifying resilience beyond what we have presented here. A substantial gap in this regard is the lack of resilience metrics for systems with multiple basins of attraction.

Alternatively, it may be worthwhile to directly create resilience metrics filling in the gaps between the concepts identified here. For example, for systems with a single basin of attraction and multiple disturbances, we observed metrics which measured the total time at an insufficient performance level (concept #7), the longest time at an insufficient performance level (concept #9), and the total loss of performance (concept #8). It stands to reason that the largest single loss of performance could therefore also be a potentially insightful approach to quantifying resilience. In this vein, novel resilience metrics could be created and evaluated against existing ones. Over time, this might lead to a compositional taxonomy of resilience metrics, comparable to work done for robustness metrics (McPhail et al., 2018).

Finally, the differences between alternative metrics could be studied by applying multiple metrics to a single case study, as was already done in a small number of papers included in this review. By testing conceptually different metrics on one system, we might be able to identify under which conditions certain metrics are preferable, or at least more conservative/optimistic. Furthermore, an exploratory approach to quantifying resilience – applying many metrics and making a holistic assessment across all the resulting data – could be a useful approach for overcoming the challenge of selecting a single metric for complex concepts such as resilience.

2.6 Conclusion

We conducted a systematic scoping review on how the resilience of socio-ecological and socio-technical systems is quantified in the relevant literature. We identify four main conclusions from this work. Firstly, a variety of resilience metrics is used. These metrics often draw on similar system properties, such as return time to equilibrium or magnitude of disturbance, but weigh these properties against each other using different functional forms, which are rarely justified. Secondly, there are some common conceptual ideas behind the different metrics, especially for systems with a single basin of attraction. Thirdly, a small number of different types of disturbances can be identified across the various case studies. However, two of these types do not fit into previously published categorizations of disturbances. Finally, we observed that the concepts of "ecological resilience" (multiple basins of attraction) and "engineering resilience" (single basin of attraction) do not seem to affect how resilience metrics are chosen for specific case studies, despite being deeply entrenched in the resilience literature. Many resilience studies of socio-ecological systems use single basin of attraction metrics, while some socio-technical systems were assessed using multiple basin of attraction metrics.

Our conclusions suggest the following two main consequences. Firstly, we recommend that researchers studying the resilience of socio-technical or -ecological systems use multiple resilience metrics, and compare the resulting insights. The different conceptual categories presented in our work provides an starting point for choosing these metrics. While this may seem to create additional work for researchers, it will also make the results more analytically robust, and provides an opportunity to engage in a dialogue with problem owners about their perception(s) of resilience. Secondly, we recommend that researchers, having compared multiple resilience metrics, explicitly justify why they chose a specific (type of) metric, and what the consequences of this choice are. By documenting this choice and the details of the used metric, both the reproducibility and reusability of the research may be improved.

Chapter 3

Resilient Life: An Exploration of Perturbed Autopoietic Patterns in Conway's Game of Life

Published as: Cika, A., Cohen, E., Kruszewski, G., Seet, L., Steinmann, P., and Yin, W. (2020). Resilient Life: An Exploration of Perturbed Autopoietic Patterns in Conway's Game of Life. *Artificial Life Conference Proceedings 32*, 656-664.

3.1 Abstract

Complex systems can exhibit autopoiesis–a remarkable capability to reproduce or restore themselves to maintain existence and functionality. We explore the resilience of autopoietic patterns–their ability to recover from shocks or perturbations–in a simplified form in Conway's Game of Life. We subject a large number of autopoietic patterns in the Game of Life to various perturbations, and record their responses using multiple resilience metrics. Our results show that while resilience is rare, we are able to identify structural features improving patterns' resilience. We also draw several parallels between the resilience of patterns in the Game of Life to real-world complex systems. Our work may be useful both for improved searching for resilient patterns in the Game of Life, and for exploring resilience in complex systems.

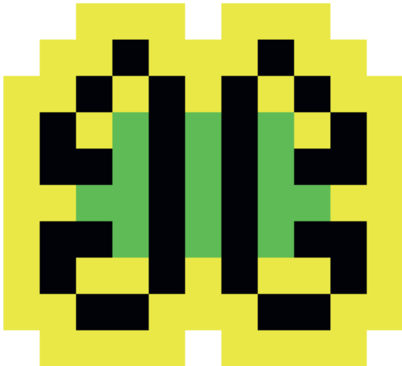


Figure 3.1: Effect of additive perturbations to still life *Inflected Clips* (apgcode xs32_4a9b8b96z259d1d96), live cells in black. Adding a live cell at green locations yields the original pattern, making the still life resilient to these perturbations. It is not resilient to perturbations at the yellow locations, where adding a live cell does not yield the original pattern.

3.2 Introduction

Living systems have been characterized as self-constructing networks of processes that maintain their own boundaries, a concept known as autopoiesis (Maturana & Varela, 1980; Varela et al., 1974). This capacity allows a system, such as the living cell, to persist throughout time, even if all of its individual components have been replaced. Nonetheless, an autopoietic system exposed to external perturbations would not survive for long without *resilience*, the ability to recover from or adapt to shocks or disruptions (Folke et al., 2004).

Here, following the characterization by Beer (2015) of emergent entities in Conway's Game of Life (Gardner, 1970) as autopoietic systems, we present an exploration of their resilience to perturbations. Conway's Game of Life (GoL) is a cellular automaton which exhibits complex emergent structures, serving as a computational environment to study the behaviour of simple systems when subjected to shocks. In particular, we focus on still lifes, which Beer (2015) identifies as autopoietic structures, together with oscillators and gliders. To study their resilience, we propose a variety of metrics and perturbations to assess whether they are able to maintain their form in response to the shocks, and identify underlying trends in their resilience.

Conway's Game of Life is, in appearance, quite brittle. Minor perturbations can cause huge collapses. Understanding whether resilience to specific types of perturbations can emerge even in these difficult conditions could shed light on whether resilience is a universal property of computational systems.

3.3 Literature Review

3.3.1 Conway's Game of Life

Conway's Game of Life (Gardner, 1970) is an example of a cellular automaton. These are computational models composed of a grid of cells that can each occupy a finite or infinite number of states. In each generation, cells change or maintain their states depending upon (i) the states of their neighbors and (ii) a set of rules. Among cellular automata, Conway's Game of Life has attracted special attention for various reasons. First, despite being based on a mere handful of states and rules, intricate structures and behaviors emerge. These structures appear capable of endless variation and evolution. In this regard, the simulation is remarkably lifelike, and has attracted substantial attention from those working at the intersection of theoretical computer science and complex systems (Adamatzky, 1998; Gotts, 2003; Lindgren & Nordahl, 1990; Margolus, 1984; Morita et al., 2002).

Conway's Game of Life is Turing-complete, meaning that it can be used to perform any computation - another possible hallmark of life (Cook, 2004; Mitchell, 2005; Wolfram, 2002). This may explain why it exhibits a wide range of interesting patterns, potentially including some that "detect" external perturbations and "repair" themselves (Goucher, 2015).

3.3.1.1 Rules of the Game of Life

Cellular automata are defined by the sets of possible states for cells to occupy, the rules which govern evolution of cell states, and the cells' initial states. The Game of Life has a two-dimensional cell arrangement, with only two cell states–"dead" and "alive"–and four rules:

1. If a dead cell has three live neighbors, then it switches states from dead to live.
2. If a live cell has fewer than two neighbors, then it switches states from live to dead.
3. If a live cell has more than three live neighbors, then it switches states from live to dead.
4. If a live cell has two or three live neighbors, then it remains alive.

The GoL uses the Moore cell neighborhood definition (Packard & Wolfram, 1985), in which each cell has eight neighbors. Surprisingly, the four simple rules described above enable Conway's GoL to compute any computable function, as shown by Conway, Berlekamp, and Guy (Austin et al., 1982). An alternative two-rule set also fully defines the GoL (Gotts, 2003), but is not commonly used.

3.3.1.2 Still Lifes

Within Conway's GoL, patterns with an evolutionary cycle of period one are called still lifes. These patterns' state configurations remain constant across time periods when left on their own, such that all dead cells remain dead, and all living cells remain alive. For example, the "block" pattern is a still life composed of 4 living cells disposed in a square arrangement. Applying GoL rules to compute the next generation yields the exact same pattern because all cells have exactly three neighbours, and thus they remain alive on the next generation, while no dead cell in its boundaries becomes alive because they all have no more than two living neighbours. Still lifes can be considered the residual of an experiment.

3.3.2 Autopoiesis

A system's viability is contingent upon exhibiting autopoietic behavior (Beer, 2015; Maturana & Varela, 1980). First introduced by Chilean biologists Maturana and Varela (1980) to distinguish living from non-living systems, the term autopoeisis is derived from the Greek words for self (αὐτο- (auto-)) and production (ποίησις (poiesis)) (Mingers, 1991). Such autopoietic systems are comprised of collections of productive units which:

> "(i) continuously regenerate and realize the network that produces them, and (ii) constitute the system as a distinguishable unity in the domain in which they exist" (Varela (1997, p. 75), as cited in Beer (2019)).

Previous work on autopoeisis in Conway's GoL has characterized the set of processes that regenerate specific still lifes, oscillators, and gliders, in the absence of perturbations (Beer, 2015). As shown, such systems "realize" their structure in perpetuity when left autonomous and in isolation. Recently, Beer (2019) examined the responses of glider patterns to environmental perturbations. Our analysis, instead, focuses on a large class of patterns, namely, still lifes, characterizing the resilience of each pattern in this class with respect to specific sets of perturbations.

3.3.3 Resilience

Real-world systems constantly experience shocks, perturbations and disruptions. The ability of a system to endure such shocks, while maintaining its form and function, is called resilience (Folke et al., 2004):

> *Resilience is the capacity of a system to absorb disturbance and reorganize while undergoing change so as to still retain essentially the same function, structure, identity, and feedbacks.*

One can view resilience as the natural behavior of a system operating in a state space with attractors (Holling, 1996). The system is initialized somewhere in the state space, and naturally gravitates towards an attractor, such as a fixed point or limit cycle (Strogatz et al., 1994). Natural shocks and perturbations push the system away from the attractor. If the system remains in the original attractor's basin of attraction, the system will return there–otherwise it will find a new attractor. These two responses are commonly referred to as engineering and ecological resilience (Holling, 1996).

The resilience of systems is generally understood in relation to specific system functions (Albert et al., 2000; Holling, 1973). As such, any definition of resilience or robustness must specify both the relevant perturbation and the feature that may persist despite said perturbation (Jen, 2003). Thus, systems do not necessarily possess a general resilience, but instead may possess many specific resiliences to different shocks, disruptions, or other perturbations.

3.4 Method

To measure resilience in Conway's GoL, we focused on a specific class of autopoietic patterns. In particular, we restricted our exploration to still lifes. In the absence of perturbations, these patterns are structurally invariant across time, making them the simplest class of such autopoietic patterns. While still lifes can be construed as fixed points in the dynamics of Conway's GoL, it is unclear whether they are stable or unstable. We apply resilience theory on single and multiple basins of attraction (Holling, 1973) as an analogy to explore the behavior of these still lifes when exposed to perturbations.

3.4.1 Quantifying resilience

Studying the resilience of still lifes required us to make two methodological decisions, as follows.

Perturbations For our study, we first identified two distinct types of perturbations to test. The first perturbation adds at least one "live" cell at a position in the Moore neighbourhood of the pattern under examination, while the second perturbation removes (subtracts) at least one "live" cell from the interior of the pattern. As such, we say a structural pattern exhibits *additive resilience* when it returns to the previous life pattern after experiencing the first type of perturbation. If a pattern returns to its previous form after experiencing the second perturbation type, we say it exhibits *subtractive resilience*. We tested the addition of either one (**add one**) or two (**add two**) consecutive living cells to the pattern perimeter, and the subtraction of one (**sub one**) or two (**sub two**) consecutive living cells from within the pattern.

Similarity metrics We next outlined how to measure recovery. While ideally a pattern will always recover its exact original structure after experiencing a shock, it may also need to adapt its structure to return to a stable state. Thus, rather than only considering exact equality to measure resilience, we chose to use multiple similarity measures in order to assess the extent to which the original structure recovered its form. Furthermore, it can take multiple steps before a pattern converges back to a new form, if ever. However, we empirically found that the number of steps required for our metrics to converge would always be well below 100 iterations (more details in the Results Section). Thus, we quantify resilience ρ as the average of the similarities between the shapes resulting from simulating GoL for 100 iterations on each perturbed shape and the original one. More formally, for a given still life s, let $\mathcal{P}(s)$ be the set of all shapes resulting from applying a given type of perturbation $\mathcal{P}$ (for example, $\mathcal{P}$ = *add one*), σ be a similarity metric, and $\mathcal{E}_k$ be the result of applying Conway's GoL rules for $k = 100$ iterations, then resilience is quantified as:

$$\rho = \frac{1}{|\mathcal{P}(s)|} \sum_{s' \in \mathcal{P}(s)} \sigma(s, \mathcal{E}_{100}(s'))$$

We further identified three variants of similarity to consider over binary vectors s_1 and s_2 representing the original and resulting patterns. Pairs of shapes were converted into same-dimensional binary vectors by taking the bounding box of the largest shape and interpreting living and dead cells as either 1 or 0, respectively.

Equality

$$\sigma(s_1, s_2) = \begin{cases} 1 & \text{if } s_1 = s_2 \\ 0 & \text{otherwise} \end{cases}$$

Inclusion

$$\sigma(s_1, s_2) = \begin{cases} 1 & \text{if } s_{1j} \leq s_{2j} \forall j \\ 0 & \text{otherwise} \end{cases}$$

Cosine

$$\sigma(s_1, s_2) = \frac{s_1 \cdot s_2}{||s_1|| ||s_2||}$$

The equality metric indicates whether the still life pattern remains identical in structure post-perturbation. The inclusion metric indicates whether the pattern absorbs, or rather includes, the perturbation into the resulting stable pattern. Lastly, our cosine metric allows us to measure the amount of overlap between the initial static pattern pre-perturbation and the ensuing pattern post-perturbation. Note that because all similarity metrics are bounded between 0 and 1, our resilience measure also lies within the same bounds.

Technical implementation Still life patterns used for our experiments came from the publicly available Catagolue census (Goucher, 2015) having population size between 4 and 42 living cells (for a total of 159 100 patterns). We also used the lifelib library (Goucher, 2017) to simulate Conway's GoL. The full set of experiments took about a day on a single-threaded Python implementation[1].

[1]We make the code available at https://gitlab.com/germank/resilient-life

3.4.2 Structural properties

We also sought to understand what structural characteristics augment a pattern's resilience to the different perturbation types. Do some structural features help protect against one type of perturbation, but not another? To answer these questions we measured five traits for each pattern: density, population size, size of the perimeter, symmetry and number of connected components.

Density The fraction of living cells within the bounding box of the pattern.

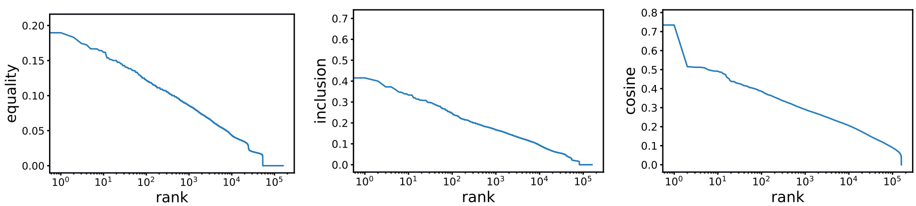
Population Size The number of living cells in the pattern.

Size of Perimeter The number of non-living cells that are at distance 1 (also in the diagonal direction) from a living cell of the pattern.

Symmetry The similarity (as given by any of the measures above) between the pattern and its rotation or reflection. Here, we also applied the same three measures of similarity defined above. For instance, whether a shape is symmetric with respect to its x-axis is computed can be measured by whether the shape and its mirror image are exactly equivalent (by applying equality), or whether they are just similar (by applying the cosine measure).

Number of connected components: The number of contiguous regions of living cells. Note that every component must be not farther than one single cell apart from another for them to be considered part of the same pattern.

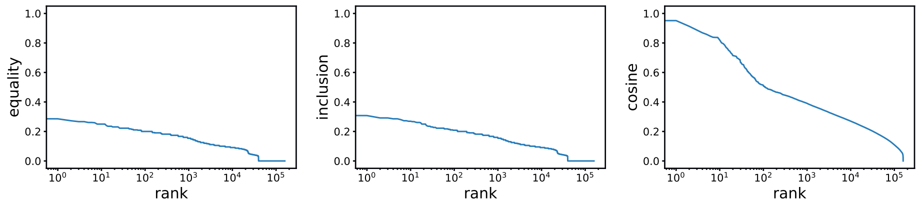
3.5 Results



(**a**) Equality (**b**) Inclusion (**c**) Cosine

Figure 3.2: Distribution of resilience results after *adding* a single cell to the pattern perimeter (*add one*) under the three considered similarity measures.

After running our experiments, we ranked each still life by the three self-similarity resilience metrics. We observed the distribution of these three metrics for single-cell perturbations (Figures 3.2 and 3.3), and found an exponential decline of resilience at a given rank, suggesting a tendency of the system towards brittleness. The results from the two-cell perturbations (not shown) confirmed this tendency. We found particularly interesting that resilience tended towards a smooth distribution for all resilience metrics, even the equality measure, which



(a) Equality **(b)** Inclusion **(c)** Cosine

Figure 3.3: Distribution of resilience results after *subtracting* a cell from the pattern (*sub one*) under the three considered similarity measures.

had the most stringent criterion. This suggests resilience, as we've defined it, is not binary, but rather a graded trait for still lifes in the GoL.

We next examined the highest ranked individual patterns for the different resilience measures within each perturbation type (Table 3.1). For ease of presentation, resilience values (on a scale of 0 to 1) are presented as percentages (on a scale of 0 to 100). Patterns that ranked highest on one resilience metric within a perturbation category, rarely ranked highest on the other metrics. For example, we found the "inflected clips" pattern (the first pattern depicted in Table 3.1) ranked first for the *add one* perturbation under the equality measure with an average resilience of 21 percent but ranked 262^{nd} under the inclusion measure and 2027^{th} under the cosine measure. Figure 3.1 highlights how perturbing cells nearest to the "inflected clips" core allowed the pattern to recover a stable structure, whereas perturbing cells at its periphery induced the pattern to collapse. The points in which the pattern is resilient to perturbations can be explained by the rules. Since the dead cells within the "inflected clips" core are surrounded by sufficiently many live neighbours, perturbations from within the core are eliminated once faced with the overpopulation rule. In other words, since living cells adjacent to this resilient core have only two neighbours, they can withstand the appearance of a third without prompting a reaction. The "tub" pattern (the second pattern depicted in Table 3.1), in contrast to the "inflected clips" pattern, exhibited no resilience under the equality measure, but obtained the highest ranking under the inclusion and cosine measures with 71 percent and 78 percent resilience scores respectively. The resilience of the "tub" comes from its ability to incorporate or to adapt to the *add one* single cell perturbations. When a cell is added to the four dead corners, the cell can "come alive" and otherwise persist without disturbing the structural integrity of the "tub" pattern. Additionally, the "tub" is resilient to eight other *add one* perturbations, whereby it can evolve into one of the former configurations with a corner filled out. Only the "block" pattern scored a rank of one across all resilience measures in any of the perturbation categories. Specifically, for the *sub one* perturbation, we found that the "block" pattern was fully resilient to removing any of its cells. Since all living cells in the pattern have three neighbours, the removal of any single live cell would automatically result in its regeneration in the subsequent time step. Finally, patterns that maximize resilience metrics for double-cell perturbations

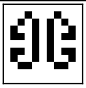
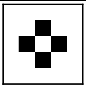
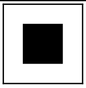
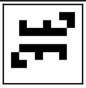
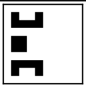
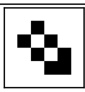
Pattern	Perturbation	Measure	Resilience	Rank
	add one	equality	21	1
		inclusion	21	262
		cosine	27	2027
	add one	equality	0	53 144
		inclusion	71	1
		cosine	78	1
	sub. one	equality	100	1
		inclusion	100	1
		cosine	100	1
	add two	equality	4	1
		inclusion	4	124
		cosine	11	20 166
	add two	equality	16	1 220
		inclusion	7	1
		cosine	0	1 274
	add two	equality	0	1 220
		inclusion	5	35
		cosine	29	1
	sub. two	equality	6	1
		inclusion	6	1
		cosine	11	86 067
	sub. two	equality	0	76
		inclusion	0	76
		cosine	67	1

Table 3.1: Patterns that are top-ranked according to at least one of the resilience measures for a given type of perturbation. All resilience values are percentages (i.e. scaled by 100). Ranks are computed out of 159 100 total patterns. In case of ties, the minimum rank is given.

showed generally much lower resilience scores, and thus, they are only resilient to few specific positions.

In general, we wanted to understand whether there were some structural features of the patterns that could predict high resilience. For this, we computed the Pearson-r correlations between the structural features described in the previous section and the resilience observations for single cell perturbations. We focused on single cell perturbations, rather than also considering two-cell perturbations because the former showed the highest amount of variance. Results are reported in Table 3.2, where each percentage value denotes the Pearson correlation between specific resilience measure and structural feature. The number of connected components, the population size, and the pattern's density come out as the most important predictors of resilience. First, only the number of connected components (*c.c.*) remained a steady positive predictor of resilience, regardless of the resilience metric or perturbation condition. Under most measures of resilience, the number of connected components was also the strongest predictor. Our "inflected clips" example in Figure 3.1 illustrated how having greater

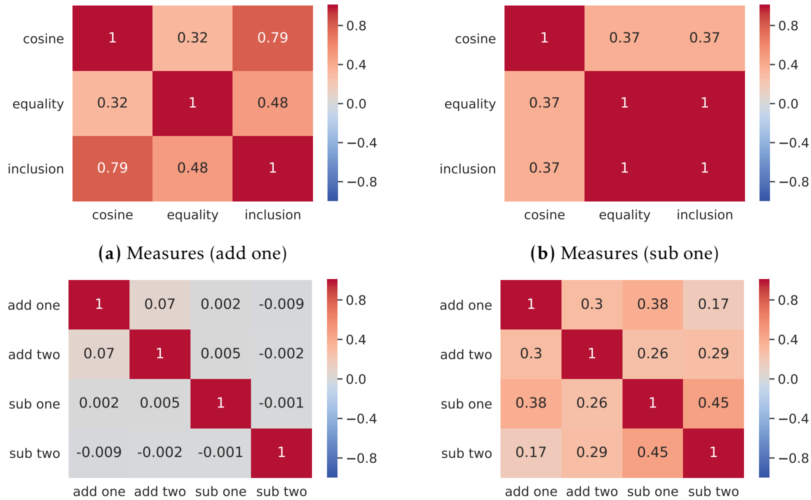
Additive Perturbation					
Equality		Inclusion		Cosine	
Feature	r	Feature	r	Feature	r
pop.	23	c.c.	9	c.c.	17
c.c.	20	dens.	−8	dens.	−10
dens.	19	sx-cos	5	pop.	−7
sx-cos	17	pop.	−4	m-y	3
sy-cos	12	sx	3	sx-cos	2

Subtractive Perturbation					
Equality		Inclusion		Cosine	
Feature	r	Feature	r	Feature	r
c.c.	35	c.c.	36	dens.	−16
pop.	18	pop.	18	c.c.	15
dens.	16	dens.	16	pop.	−14
cw-cos	9	cw-cos	9	cw-cos	−10
ccw-cos	9	ccw-cos	9	ccw	−10

Table 3.2: Pearson-r correlations (in percentage) between resilience for single-cell perturbations and structural features for the top-5 correlated or anti-correlated features. Key: *pop*=population; *c.c.*=number of connected components; *dens.*=density; *sx/sy*=symmetry on x/y axis; *sx/sy-cos*=symmetry on x/y axis as given by the cosine between the shape and its mirror image; *cw/ccw*=invariance to clockwise or counter clockwise rotation; *cw/ccw-cos*=invariance to clockwise or counter clockwise rotation measured as given by the cosine between the shape and its rotation.

numbers of connected components may facilitate additive resilience, where disconnected components may create an internal core that is protected from perturbations. The resilience observed in the subtractive case, by contrast, may be a byproduct of the fact that many composite still lifes are often composed of blocks among other shapes. That said, the strength and direction of the relationship between the other predictors and resilience depended on the type of perturbation and resilience metric considered. For example, while high density (*dens.*) positively predicted resilience under the equality metric when performing an additive perturbation, it negatively predicted resilience under the inclusion and cosine metrics. Additionally, under the subtractive perturbations, high density (*dens.*) positively predicted resilience under both the equality and inclusion metrics, but negatively predicted resilience under the cosine metric. Nevertheless, it is important to bear in mind that these are linear correlations, and may not be telling the full story. For example, when analyzing the effect of population size in the additive perturbation under equality we observed that the maximum resilience for each population size tends to grow (with some roughness) until size 32 where it maxes out, but then starts to slowly decline.

Next we analyzed how related our own definitions of resilience were with each other. For this, we computed the correlations between the resilience values obtained from different types of perturbations and different types of measures (Figure 3.4). We found that the resilience of a pattern to a given type of perturbation does not transfer to other perturbation types when we restrict to measuring



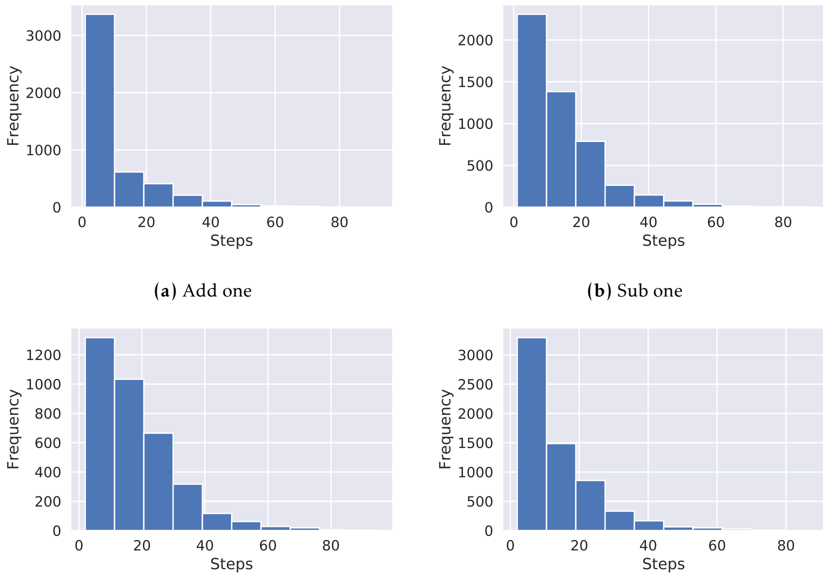
(a) Measures (add one) **(b)** Measures (sub one)

(c) Perturbations (under equality) **(d)** Perturbations (under cosine)

Figure 3.4: Pearson-r correlation coefficients between different types of perturbations and measures.

resilience through equality (Figure 3.4c). This supports the idea that systems have specific resiliences to unique disruptions, rather than a general resilience (Carpenter et al., 2001). However, when we adopted a relaxed notion of recovery as given by the cosine measure (Figure 3.4d), we saw a weak, yet significant, correlation between different types of perturbations. Thus, under this particular definition of recovery, we can see a somewhat more generalized notion of resilience. Conversely, the correlations between our measures also change as a function of the perturbation that we consider. If we restrict to the "add one" perturbation (Figure 3.4a), the inclusion and the cosine measures tend to highlight the resilience of the same patterns while equality behaves in a more idiosyncratic way. On the other hand, when we focus on the "subtract one" perturbation (Figure 3.4b) equality and inclusion behave almost in the same way, and differently from cosine.

Last, but not least, we studied the length of the transients to see how many steps these resilient patterns take to recover from a perturbation (up to a maximum of 100). We observed that for both the inclusion and the equality measures, the transient lengths were no longer than 2 steps. For this reason, we restricted the following analysis to just the cosine measure, which displayed more variation. Results are displayed in Figure 3.5. For small perturbations (Figures 3.5a and 3.5b), most transient lengths are concentrated at the lower end of the spectrum. Nonetheless, there is a graded distribution with a decreasing number of patterns having longer transient lengths. Interestingly, when a larger perturbation is applied to the system, the transient lengths increase, as seen in Figures 3.5c and 3.5d, while still maintaining a graded distribution.



(a) Add one

(b) Sub one

(c) Add two

(d) Sub two

Figure 3.5: Distribution of length of transients when using the cosine measure.

3.6 Discussion

3.6.1 Additive vs. subtractive resilience

We found virtually no still lifes in the GoL which were resilient to both additive and subtractive perturbations. Our work reinforces previous research examining resilience of real-world systems wherein resilience must be considered in specific contexts (Carpenter et al., 2001), and with specific actors and stakeholders in mind (Jones, 2018). That a system exhibits specific resiliences (vs. a general resilience) even in abstract environments such as Conway's GoL suggests resilience may be a fundamental property of complex systems. Identifying such a property has far-reaching consequences for the study of real-world systems, as it would invalidate a large body of work quantifying "the" resilience of such systems.

3.6.2 General trends in still life resilience

The Game of Life is highly symmetrical by nature, as all rules function equally in all four cardinal directions. However, symmetry, as measured by the similarity between a shape and its mirror image (both along the x axis and the y axis) did not necessarily yield strong predictions of resilience. Although, there did appear to be a weakly positive relationship between symmetry (as measured by the cosine) and resilience under the equality metric for both perturbation types. This suggests that symmetry is not an crucial factor in a still life's resilience.

Additionally, a fundamental constraint of the Game of Life is that still life patterns have a maximum density of 0.5 within their boundaries (Capcarrère & Sipper, 2001; Chu & Stuckey, 2012; Chu et al., 2009; Elkies, 1999). This represents a tipping point in the population dynamics of the still life, and follows quite naturally from the rules–as any density above 0.5 creates a cascade of cell deaths, ultimately wiping out all cells in the pattern (Griffeath & Moore, 2003). Nevertheless, we found still lifes with densities approaching 0.5 were more resilient to perturbations than those with sparser densities. It appears that still lifes existing close to the threshold of their ability to maintain structure–this might be considered their autopoietic limit–are more resilient. This observation was mentioned by Holling (1996), who described how humans operate at body temperatures close to lethal levels–allowing them to mobilize more energy for expanding their living environments or responding to threats. Here we again see an interesting parallel between the abstract dynamics of the Game of Life, and real-life complex systems.

We note here that our density measure is not fully consistent with previous work on the density of still lifes, having used an alternative method for perimeter construction (Elkies, 1999). However, we find support for our method in current autopoiesis discussions regarding what constitutes living system spaces (Harvey, 2019; Villalobos & Razeto-Barry, 2019). As such, we believe the underlying trends are strong enough to suggest a general pattern regarding density, stability and resilience in autopoietic systems.

3.6.3 Conclusions and Future Work

We explored the degree to which still lifes–fixed-points in Conway's GoL–are resilient to one- and two-celled perturbations. We found certain still lifes served as stronger system *attractors* (Holling, 1996), in that they exhibited higher propensities to return to stable states post-perturbation. Ultimately, we observed a graded, exponentially decaying, distribution of resilience among patterns. This suggests that we may find patterns that are increasingly more resilient to the studied perturbations when examining (exponentially) larger samples. Furthermore, this graded surface may be just what an evolutionary process needs to discover ever more resilient shapes, even though more work characterizing the smoothness of this surface would be needed to support that idea.

While this exponentially decaying trend of resilience in patterns was generally stable across different conditions, the particular definition of resilience given by the used perturbation and the way to measure recovery, was prone to highlight different patterns. This finding supports the idea that no universal notion of resilience exists, but rather, resilience is by nature context conditional.

Our analysis also evidenced some structural features of resilient patterns. Population, the number of connected components, and density served as the strongest general important predictors of resilience; however, for only the number of connected components was the direction of the relationship consistent. Investigations using these structural insights could help lead to the discovery of more resilient patterns in the future are left for future work. We conjecture that there may exist other not yet discovered patterns which maximize our resilience metrics with respect to the studied perturbations.

While we restricted our analysis to still lifes in this study, future work will extend our exploration to other types of autopoietic patterns in the GoL such as oscillators and gliders (Beer, 2015). Furthermore, the perturbations that we explored were two simple ones; however, they may not arise naturally in the dynamics of the GoL. Thus, a natural extension to our current analysis is to test more "naturalistic" perturbations, such as the impact of a glider hitting a still life. Indeed, some patterns with the name of "eaters" (Griffeath & Moore, 2003) have already been shown to be resilient to these types of perturbations. Thus, a more systematic analysis, such as the one performed in our study could help verify whether our resilience conclusions also hold in more complex scenarios.

Finally, exploring non-resilience is important for its own sake. For instance, through our inclusion and cosine measures of resilience, we found perturbing a still life can result in it re-stabilizing as a transformed still life pattern. Future work will explore how different perturbations may transform still lifes into oscillators, gliders, neither, or some combinations thereof.

To conclude, we have presented a systematic analysis of the resilience of still lifes in Conway's Game of Life. We have observed that while it is a rare property, increasingly higher levels of resilience are observed. As argued above, this graded landscape may provide evolution with just the right stepping stones that it needs in order to discover ever more resilient forms of organization. We remain cautiously optimistic that, even if rare in a universe which is, in appearance, as brittle as Conway's Game of Life, resilience can be found.

Chapter 4

Robust Resilience: Optimization with Ensembles of Metrics May Improve Resilience to Novel Shocks

In preparation as: Steinmann, P., van Voorn, G.A.K., and Molenaar, J. Robust Resilience: Optimization with Ensembles of Metrics May Improve Resilience to Novel Shocks.

4.1 Abstract

Resilience is the ability of systems to withstand or recover from disturbances. A key challenge in assessing a system's resilience is quantifying it, enabling the evaluation of potential resilience-increasing interventions. A wide variety of metrics have been proposed for quantifying resilience using various conceptual approaches. This introduces an uncertainty - which metric to choose? - into resilience analysis. We show that by using an ensemble of resilience metrics and many-objective optimization, a system can be optimized to be resilient to a given disturbance according to multiple independent metrics. We also observe that such a system may have the added benefit of being more resilient to novel disturbances it was never optimized for. Using ensembles of resilience metrics may therefore enable the design of systems resilient to a range of disturbances, including unforeseen ones.

4.2 Introduction

Modern society is dependent on a wide range of socio-technical and -ecological systems to ensure its continued functioning. Water supply, energy infrastructure, and communications networks are just some examples of such systems. To ensure that these systems continue functioning in the face of disturbances such as earthquakes, heatwaves, or cyber attacks, understanding and improving the so-called resilience - the ability to withstand or recover from a disturbance - of these systems has become a focal point of public and private decision making. A key step in studying the resilience of such systems is translating the concept of resilience into a measurable value. However, it is not immediately obvious how to perform this quantification of resilience - paradoxically, not because there are no established metrics for resilience, but because there are so many. Dozens, if not hundreds of resilience metrics have been described (for examples, see reviews by Hosseini et al. (2016), Quinlan et al. (2016), and Sun et al. (2020), as well as Chapter 2.

The American National Academy of Sciences formally defines resilience as “the ability to prepare and plan for, absorb, recover from, and more successfully adapt to adverse events” (Committee on Increasing National Resilience to Hazards and Disasters et al., 2012). This definition mashes together three distinct schools of thought on resilience with their own scientific heritages. The first is commonly referred to as engineering resilience, although it has also been applied to non-technical systems (e.g. Pimm, 1984). Systems which are resilient in the engineering sense quickly return to their previous performance level after experiencing a disturbance (Holling, 1973, 1996). The second conceptual approach to resilience is ecological resilience, and describes those systems as resilient which can absorb a disturbance without switching to a less desirable state (Holling, 1996). The third approach, socio-ecological resilience, incorporates adaptive mechanisms such as learning and evolution, and by means of an adaptive cycle (Holling & Gunderson, 2002), foresees systems not just recovering from disturbances, but improving their resilience and performance in the process.

The concept of resilience relates closely to a number of other concepts applied to complex systems, chiefly stability, sustainability, vulnerability, and robustness. Depending on the author (e.g. Brand and Jax, 2007; Holling, 1973; Kelly and Harwell, 1990; Nilsson and Grelsson, 1995) these concepts may be identical, related, subordinated in some constellation, or even independent. As a consequence, the term resilience has become “almost meaningless” (Klein et al., 2003) beyond a general understanding that it describes a system’s dynamic response to disturbances (Hufschmidt, 2011; Zhou et al., 2010). This focus on system dynamics may partly be the reason for the ongoing difficulties in defining and delineating resilience, as the complexity and context of a system may change over time (certainly when it is recovering from a disturbance), making it difficult to establish any kind of consensus on what is even being studied (Bohensky, 2008; Marshall et al., 2007; Nelson et al., 2007). This is compounded by the fact that such systems often have contested or unclear ownership (Gotts et al., 2019), making it unclear who even has the authority to answer such questions (Norris et al., 2008; Opdyke et al., 2017).

Since no unified definition of resilience seems imminent, the quantification of resilience has received substantial scientific attention. Some authors (e.g. Carpenter et al., 2005; Cutter et al., 2008) have called for standard resilience metrics that can be consistently applied to many systems, improving comparability and transparency. Holling (1996) saw the concepts of engineering and ecological resilience as inextricably linked to their respective metrics. However, other researchers (e.g. Simonovic et al., 1992) have opined that every study requires a unique formulation of resilience and its accompanying metric(s), in light of the complexity and lacking definitional clarity. The latter approach was later formalized as the notion that resilience is specific to a particular stakeholder, scope, disturbance, and time frame (Carpenter et al., 2001; Cutter, 2016; Meerow & Newell, 2019). Favoring the specificity of resilience, a wide variety of metrics have been created to quantify resilience in different contexts. These are based on concepts as diverse as system-scale behavior over time (Holling, 1973), flow magnitudes between entities in the system (Ulanowicz et al., 2009), and pre- and post-disturbance spatial patterns (Cika et al., 2020).

The plethora of available resilience metrics may seem like a boon to scientists studying resilience. However, precisely because at some point a choice must be made between the available options (Beccari, 2016), and there is (partly because of the definitional fuzziness of resilience itself) little epistemic scaffolding to climb on, the selection of a resilience metric represents an uncertainty in the analysis process. As Pimm (1984) and Grimm et al. (1992) showed, this uncertainty is impactful - depending on the choice of resilience metric, completely different conclusions may be reached. Furthermore, the availability of many different metrics means that the likelihood of a metric being used incorrectly or misleadingly increases (Jain, 2009).

A number of resilience scholars have proposed that using multiple metrics may be a solution to the problem of metric selection. Dore and Webb (2003) suggested that there is no systematic way of imposing a unidimensional metric on complex ecosystems, highlighting the incompleteness of a single metric. Mumby et al. (2014) emphasized that "broad, relatively uncorrelated categories of 'resilience attributes'" should be used in place of single metrics. Duveneck and Scheller (2016) similarly advised that multiple measures should be included when assessing the resilience of complex systems. However, none of these authors actually implemented a multi-metric approach to quantifying resilience. How could such an ensemble approach to quantifying resilience with multiple metrics be implemented in practice?

In this paper, we evaluate optimization of a complex system's resilience using multiple resilience metrics, and contrast it with single-metric resilience optimization. We demonstrate this using a computational model of a complex system experiencing three different disturbances, and applying five conceptually distinct resilience metrics. In a first step, we investigate how the different metrics score the system's resilience for the different disturbances, and how the scores relate across the disturbances. In a second step, we identify the system parameter combinations that make the system most resilient - i.e. achieve the highest scores - according to a given metric and disturbance. We also identify the system parameter settings which make the system most resilient to each individual disturbance across all resilience metrics simultaneously, using many-objective optimization. We then compare the performances of these multi-metric-optimal parameter settings if one of the *other* disturbances occurs - that is, how resilient the system is to disturbances it was never prepared for.

4.3 Methods

As described in the previous section, ensemble-based resilience quantification should incorporate a broad range of little-correlated metrics. To evaluate this concept in practice, we select a set of conceptually distinct resilience metrics, and apply them to a complex system experiencing a range of disturbances. In a previously performed systematic scoping review of resilience metrics described in Chapter 2, we identified six distinct categories of resilience metrics for complex systems experiencing a single, time-bound disturbance. We select resilience metrics representing five of these categories for this study. The sixth category is not applicable to our case study, outlined below, for technical reasons. The first selected metric is *return time*, or the time it takes the system to regain its pre-disturbance performance level. Such metrics have been proposed by many authors including Holling (1996) and Pimm (1984). With t_d representing the time of disturbance, and t_r representing the time of recovery to the system state at $t = t_d$, resilience is quantified as:

$$R_{rt} = t_r - t_d \tag{4.1}$$

As the system has stochastic aspects and therefore does not have a true steady state, we allow a ±10% margin for identifying t_r based on the pre- and post-disturbance performance levels. The main author experimentally derived this threshold based on the model's dynamics for the default parameter ranges given by ten Broeke et al. (2016).

The second metric is *performance loss*, or the total performance lost due to the disturbance (e.g.Ouyang et al. (2019)). With P_u representing the performance over time of an undisturbed system, and P_d representing the performance over time of a system experiencing a disturbance, resilience may be quantified as:

$$R_{pl} = \int_{t_d}^{t_r} P_u - \int_{t_d}^{t_r} P_d \tag{4.2}$$

The third metric was introduced by Arreguin-Sanchez et al. (1998) and incorporates both the *return time* and the maximal proportional performance change. P_{min}, P_{max}, and P_{t_d} represent the minimal and maximal performance levels during the disturbance, as well as the performance level at $t = t_d$. The term "performance level" refers to the system property whose resilience should be quantified, such as GDP, population level, or cost, and is generally specified by the analyst or stakeholder(s). Under this metric, resilience is quantified as:

$$R_{as} = \frac{\frac{P_{max}-P_{min}}{P_{t_d}}}{t_r - t_d} \tag{4.3}$$

The fourth metric used was proposed by Perez-España et al. (2001) and incorporates the *return time* as well as the maximal and minimal performance levels similar to the third metric, however, the pre-disturbance performance level is not considered:

$$R_{pe} = tan^{-1}\left(\frac{\frac{1}{P_{max}-P_{min}}}{t_r - t_d}\right) \tag{4.4}$$

The final metric was proposed by Lesnoff et al. (2012) and measures resilience using the population multiplication rate during the recovery period. With P_{t_r} as the performance level at $t = t_r$, resilience is quantified as:

$$R_{le} = \left(\frac{P_{t_d}}{P_{t_r}}\right)^{\frac{1}{t_r - t_d}} \tag{4.5}$$

To facilitate comparison, all scores generated with these metrics were re-scaled to a $[0,1]$ interval based on the highest and lowest scores for each metric across all performed simulation experiments. This is necessary as the individual metrics return scores differing by several orders of magnitude, hindering comparison and evaluation of trade-offs. We also note here that without this re-scaling, the different resilience metrics would have different dimensions, rendering comparison impossible. The re-scaled values are dimensionless. We have previously discussed the issue of unclear or mismatched resilience metric dimensions in Chapter 2.

To evaluate the usage of resilience metrics ensembles with complex systems, we draw upon a simulation model by ten Broeke et al. (2016) previously used for studying the resilience of complex, socio-ecological systems (ten Broeke et al., 2017). The model describes the interaction between a renewable resource and a set of consumer agents in a spatial environment. The model it implemented in Netlogo (Wilensky, 1999), a widespread agent-based modelling environment. The resource grows and diffuses across a two-dimensional space, which is traversed by consumer agents looking to harvest and sustain their metabolic needs, and if sufficiently sustained, reproduce. 15 different model parameters govern the behavior of the resource and consumers, whose interactions create different behavior patterns depending on the parameter settings. These patterns include oscillations, exponential growth, extinction, and steady states. For further information about the model and its usage, we refer to the original publication and subsequent work utilizing the model (ten Broeke et al., 2019; ten Broeke et al., 2016; ten Broeke et al., 2021; ten Broeke et al., 2017). As in all these previous studies, we use the population level of the consumer agents, which may be considered its performance level, as our outcome of interest.

In light of the previously discussed specificity of resilience to a stakeholder, place, time horizon, and disturbance (Cutter, 2016), we also use an ensemble of disturbances to evaluate the performance of our metric ensemble. Specifically, we choose three distinct disturbances, targeting three different parts of the modelled system in different ways, based on previous research investigating different types of disturbances resilient systems may experience presented in Chapter 2. The first disturbance *sudden-c* targets a property of the agents, their harvest efficiency, reducing it by 50% for a relatively short period of time, before restoring the efficiency back to its original value. The second disturbance *gradual-r* affects an aspect of the environment, the resource growth rate, gradually reducing the growth rate over a longer period of time to 50% of its original value before suddenly recovering. The final disturbance *continuous-agent-population* directly targets the agent population, continually removing a small percentage of the population every time step over a longer period of time. Figure 4.1 demonstrates the effects of these disturbances on the system's performance.

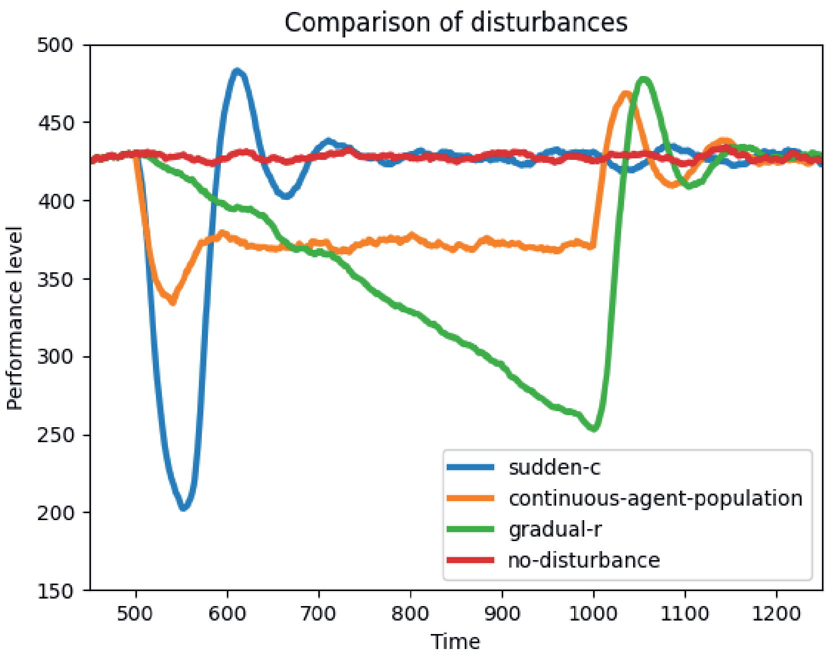


Figure 4.1: The effect of the three applied disturbances, along with the undisturbed behavior, for a randomly selected model run. The x-axis shows the simulation time steps, excluding the warm up period. The y-axis shows the total number of agents in system, which we consider its performance level - the system property for which resilience should be maximized.

To evaluate the performance of the metrics for different model parametrizations and disturbances, we perform a so-called parameter sweep, running the model many times with different input parameter combinations and model settings. The input parameter combinations were sampled using a uniform Latin Hypercube design. We identified the parameter ranges giving a low chance of population extinction using behavior-based scenario discovery (Steinmann, Auping, et al., 2020) based on the original ranges (ten Broeke et al., 2016). The used parameter ranges are listed in Table 4.1. We generated 7000 random samples of input parameter combinations. We then ran four distinct simulation experiments for each input parameter combination: no disturbance to the system, the *sudden-c* disturbance, the *gradual-r* disturbance, and the *continuous-agent-population* disturbance. As the model has stochastic aspects, we replicated each of these simulation experiments 10 times, for a total of 280 000 simulation runs. In every case the model was given a warm up period of 500 time steps, after which the disturbance was applied, and the simulation was ended after 2000 time steps. These values were experimentally derived based on the desire for reaching a quasi-steady state after initialization and the subsequent disturbance, under a variety of parameter settings. The simulation experiments were post-processed in two steps. Firstly, those experiments in which any of the disturbed runs did not return to an equilibrium within the simulation time horizon were discarded. Secondly, all experiments were re-scaled to a common population level of 100 to ensure comparability, based on each input parameter combination's mean population level for an undisturbed model run.

Table 4.1: Sampled parameter ranges. These ranges were chosen specifically to rapidly move the system to a quasi-steady state, and therefore differ from the original, broader parameter ranges.

Parameter	Range
c	[0.3, 0.7]
D	[0.05, 0.2]
E_b	[8, 10]
E_h	[0, 0.2]
E_m	[0, 0.15]
E_{move}	[0, 0.75]
K	[1.75, 3.75]
r	[0.4, 0.5]
R_0	[0, 1]
R_{max}	[1, 2]
R_{unc}	[0.4, 1]
v_b	[0, 20]
v_d	[0, 20]
z	[0, 0.5]

We analyze our results in two ways. Firstly, for each metric, we evaluate whether achieving a high resilience score when experiencing a given disturbance is suggestive of also performing well on another disturbance. We do this through visual inspection of pairwise grid plots. Secondly, we identify the Pareto-optimal fronts across all five metrics for each disturbance - that is, the input parameter combinations that have the best overall performance (calculated as the mean) across all metrics. This means that a model run scoring highly on one metric and poorly on a second metric is equally good as a model run scoring poorly on the first metric and highly on the second metric, but a model run scoring worse on both metrics will be dominated and considered inferior. We then evaluate the performance of these multi-metric-optimal combinations against the the performance of parameter combinations optimal for a single metric when experiencing the other disturbances.

4.4 Results

4.4.1 Disturbance responses per metric

In Figure 4.2, we visualize how every simulation run included in our analysis performs for a every pair of disturbances and each metric in a set of scatter plots. In essence, this figure shows whether the response to the individual disturbances under a given metric is similar for different disturbances.

For the metric R_{rt} given in Equation 4.1, it appears that there is a high degree of similarity between the disturbance responses. In other words, system parameter settings which have a high score on this metric (i.e. a low return time) for some disturbance are likely to also have a low return time on one or more other disturbances. There are very few outliers across all three disturbances, although two interesting patterns can be observed - an outlier cluster for the *sudden-c* disturbance, and two to three distinctive striations for each subplot.

The metric R_{pl} given in Equation 4.2 metric shows varying similarity levels, with the strongest similarity apparent in the first subplot with *sudden-c* versus *gradual-r* disturbances. For these two disturbances, system parameter combinations performing well on one are likely to also perform well on the other. The other two subplots show that this metric is relatively weakly sensitive to the *continuous-agent-population* shock. In the last subplot, no similarity is immediately visible, indicating that no inference can be made from the response to one disturbance for the other one. More outliers are visible.

Interestingly, the metric R_{as} given in Equation 4.3 shows both similarities and strong differences for the different pairs of disturbances. While *sudden-c* and *gradual-r* seem similar, the other two combinations appear distinct. This means that if a system performs well when experiencing one of the two disturbances, it will likely perform poorly when experiencing the other. There are not many outliers.

The metric R_{pe} given in Equation 4.4 shows little to no similarity for all three combinations of disturbances. It is therefore difficult to generalize the system's performance from one disturbance to another. A large number of outliers is readily apparent.

Finally, the metric R_{le} given in Equation 4.5 also shows varying degrees of similarity, with a strong similarity for the first pair of disturbances, and no similarity for the other two pairs. There are few outliers.

When comparing between the metrics, we observe that the metrics R_{rt} and R_{pl} have broadly similar patterns across all three disturbance pairs. Similarly, the metrics R_{as} and R_{le} also have comparable patterns across all three disturbance pairs. However, these two pairs of metrics are distinct from one another. The metric R_{pe} is unique in that its correlations between the disturbance pairs are not similar to those of any other metric.

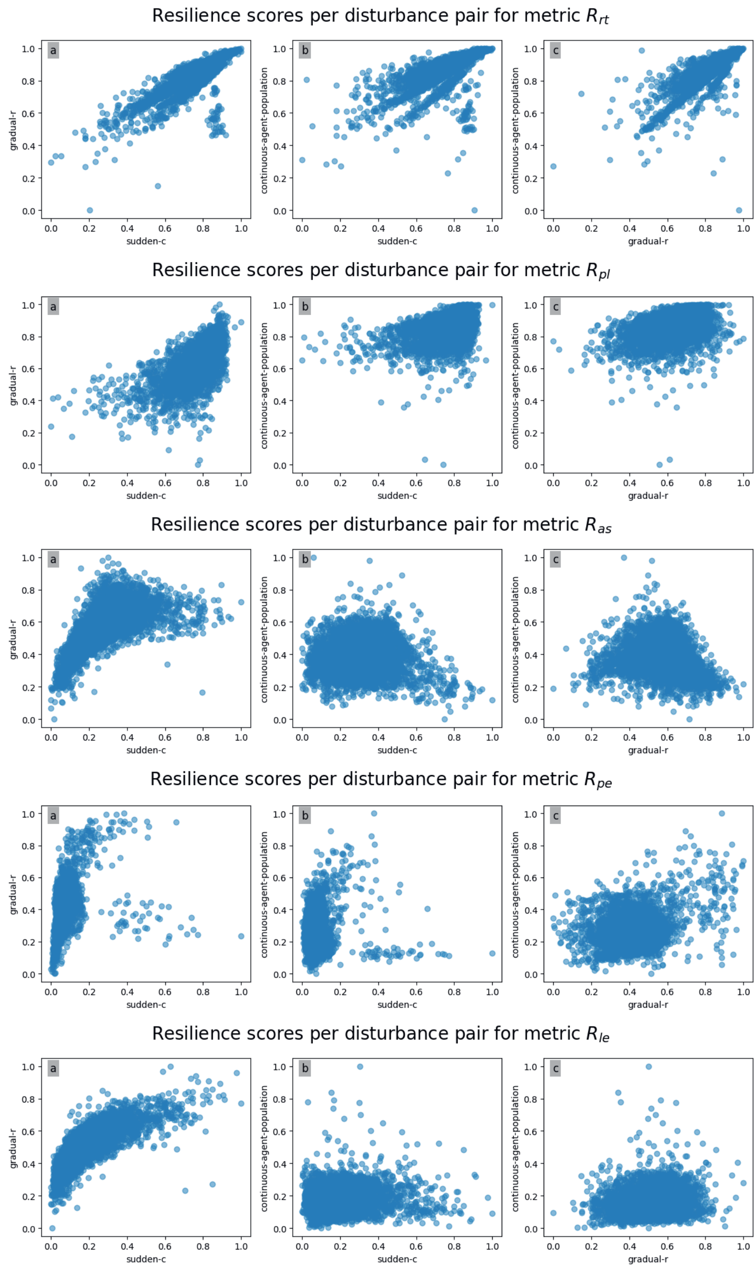


Figure 4.2: Similarities between responses to different disturbances per evaluated resilience metric. For each resilience metric, every pairing of disturbances is plotted pairwise. The data points represent the individual simulation runs, plotted using their scores per (rescaled) resilience metric.

4.4.2 Resilience to novel disturbances

Figure 4.3 compares the performance of the 20 highest-scoring system parameter settings for each metric and disturbance combination, and the 20 best-performing Pareto-optimal parameter combinations across all used resilience metrics. We selected the 20 best parameter settings to represent the fact that, when optimizing a complex system, there are often multiple configurations which perform roughly equally well, and could all be chosen with equal justification. In line with Jain (2009), we use the mean to summarize the multidimensional Pareto-optimal input parameter combinations.

For the responses to the disturbance *sudden-c*, when experiencing the disturbance the systems were optimized for, we observe that the input parameter combinations score quite highly on all metrics, as would be expected (see subplot (a) of the top plot row). The Pareto-optimal parameter combinations score somewhat lower, although still comparable with the systems optimized for the metric of R_{pe}. However, if a different disturbance occurs - that is, either *gradual-r* or *continuous-agent-population* - the parameter settings which would be optimal for a *sudden-c* disturbance suddenly perform worse. Their scores are often significantly worse than the Pareto-optimal parameter settings' scores, which are quite high across the board. The sole exception is found for the parameter combinations optimized for *return time*, which also have a high score when experiencing disturbances they were not optimized for. This is congruent with the observations made above regarding similarities between disturbance scores per metric.

This picture does not repeat itself for the *gradual-r* disturbance. For every metric, a set of high-performing system input parameter combinations can be found (center part of the middle subplot). If a different disturbance than *gradual-r* occurs, the settings optimized for the metrics R_{as}, R_{pe} and R_{le} perform worse than the Pareto-optimal input parameter combinations, on the whole. Interestingly, the systems optimized for *return time* and *performance loss* against the *gradual-r* disturbance also perform quite well when experiencing one of the other two disturbances.

For the *continuous-agent-population* disturbance, we observe that the highest-scoring input parameter combinations have quite a large distribution even for the disturbance they were optimized for (rightmost part of the third subplot). As with the previous disturbance, the parameter settings optimized using the three more intricate metrics (R_{as}, R_{pe}, R_{le}) perform substantially worse than the Pareto-optimal parameter combinations when experiencing a not-optimized-for disturbance. However, the settings optimal under the *return time* and *performance loss* metrics also perform reasonably well for the other two disturbances.

4.5 Discussion

The main objective of this paper was to evaluate how an ensemble of resilience metrics might be used to optimize a complex system's resilience, and what the benefits of this approach might be. We conducted this evaluation by using five different resilience metrics to quantify the response of a simulation model

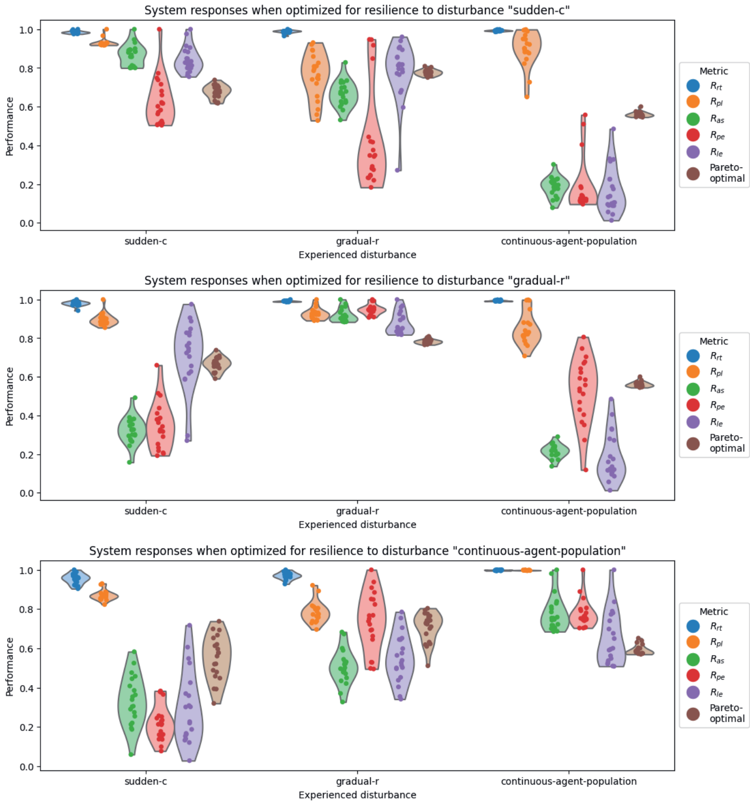


Figure 4.3: Performances of system parameter settings optimized for a given disturbance when experiencing a different disturbance. For each combination of disturbance and metric, the 20 highest-scoring input parameter combinations are shown as individual points, with an underlying violin plot (symmetrical kernel density estimate) to summarize the distribution.

of resource-consumer dynamics experiencing three different disturbances. We found that, depending on the metric, the scores for the various pairs of disturbances can be correlated in very different manners. There is also not a consistent pattern between the different metrics. Some metrics show clear similarities in how their resilience scores for the different disturbances are correlated. However, the "agreement" between metrics is not readily apparent from their functional forms. For example, metrics R_{rt} and R_{pl}, or return time and performance loss, show similar correlation patterns across all three pairs of disturbances, but have completely different functional forms and constituent elements. We also identified input parameter combinations for the simulation model which scored highly according to all five resilience metrics for a given disturbance. Intriguingly, we observed that these parameter combinations also gave the system good response capacity to *other* disturbances.

Based on these results, we posit that exploratory ensemble approaches may be useful not only for parametric and structural uncertainties, but also for metric uncertainty. To our knowledge, this has not been taken up in the literature yet. This study therefore represents one of the first works to study how an ensemble of metrics purporting to measure the same system property may be used to support decision making. This is perhaps because metrics are often regarded as a higher-level decision making process by problem stakeholders, and are thus often exogenous to the analysis (Sharma & Chen, 2020). Based on our results, endogenizing this uncertainty into the analysis may yield benefits.

Through our ensemble-based approach, we have shown that resilience metrics which are conceptually distinct may be correlated in practice. The similarity of nominally independent resilience metrics has also been observed by Kristensen et al. (2003), who referred to such metrics as "concordant". The fact that such similarities can only be observed once the metrics have been implemented and evaluated in a simulation model highlights the value of the exploratory approach. Thus, when Mumby et al. (2014) call for using a diverse range of uncorrelated metrics, we support this call - but caution that it may not be possible to identify the degree of similarity between metrics until they are applied. It is noteworthy that there were no strong patterns regarding the similarities - some metrics were correlated for one pair of disturbances, and uncorrelated for another. This implies that it is not straightforward to translate resilience metrics from one context to another, an argument against the usage of standardized metrics.

We have also shown that by reasoning across the ensemble of resilience metrics, we can identify system input parameter combinations which perform well across five metrics for a given disturbance. While this may seem trivial, we believe it shows that the exploratory approach can be a real benefit to resilience studies, especially for systems with multiple stakeholders where reaching consensus on "the" resilience metric will be difficult or impossible. However, these Pareto-optimal system parameter combinations do have a lower performance than those optimized for a single metric. In other words, there is no free lunch in resilience, as improved resilience likely comes at the cost of outright performance (Karakoc & Konar, 2021; Roege et al., 2014). Identifying which of these to optimize for depends on thorough elicitation of stakeholder needs and desires (Mumby et al., 2014).

We found that systems whose parameters have been optimized with an ensemble of resilience metrics may be resilient to disturbances they were never optimized for. In other words, it may be possible to make a system resilient to a disturbance it has not been prepared for, or that the analyst has even foreseen. Munn (1992) emphasized that resilience should be about improving system preparedness for a range of possible futures - naturally including different disturbances. Ultimately, this would lead to identifying the studied system's safe operating space (Rockström et al., 2009) or bubble of stability (Roux et al., 1999) - the domain within it can continue existing and functioning. This also implies that the bespoke approach to quantification proposed by Cutter (2016) and others may be expandable to a new kind of resilience which can respond adequately to many different disturbances - something we might refer to as general or robust resilience. This could improve both the cost-effectiveness and success rate of resilience-enhancing interventions in systems.

We note here that throughout this research, we have only studied one system with a single steady state, and metrics which are amenable to such dynamics. This is a strong limitation of our study, as natural systems are never truly stable (Holling, 1973; Resh et al., 1988). While we attempted to account for this by using a stochastic model and a relaxed interpretation of the concept of equilibrium, this nevertheless represents a strong simplification, especially since we discarded simulation experiments which did not recover to the pre-disturbance steady state within the simulation time horizon. With that said, the majority of resilience metrics used in practice assume a steady state (see Chapter 2 for details), and there is as yet little agreement on how to measure adaptive, non-equilibrium system resilience (Fernandez & Ahmed, 2019). This has previously been highlighted as a valuable future research direction (Egli et al., 2019).

4.6 Conclusions

Conceptually distinct resilience metrics may return similar results based on the disturbances and system they are applied to. This implies that the independence of resilience metrics can only be identified *ex post* using exploratory analysis. Regarding the choice of appropriate resilience metrics for systems that might experience different disturbances, we propose the following dichotomy. If a single metric can or must be used, we recommend using a simple metric such as return time R_{rt}, as the systems performing optimally under that metric for a given disturbance will also respond well to other disturbances. However, if agreement cannot be reached on a single metric due to lack of knowledge or stakeholder disagreement, then multi-metric optimization may be an effective tool for finding consensus solutions that perform reasonably well across all metrics. The resulting system parameter settings may have the (rather valuable) added advantage of performing well when experiencing other disturbances, which is not necessarily true for their constituent metrics. This implies that it may be possible to design systems which are resilient to a variety of disturbances - in other words, their resilience is robust to the uncertainty of which disturbance will occur.

In future research, we hope to study the exploratory approach to understanding resilience metric similarity further, for a variety of systems and disturbances. Furthermore, it may be useful to identify the underlying mechanisms of a system's resilience by linking the input parameter combinations giving it high resilience with an analysis of its structure, in order to better identify relevant resilience metrics.

Chapter 5

Behavior-based Scenario Discovery Using Time Series Clustering

Published as: Steinmann, P., Auping, W.L.A., and Kwakkel, J.H. (2020). Behavior-based scenario discovery using time series clustering. *Technological Forecasting and Social Change 156*, 120052.

5.1 Abstract

Scenario Discovery is a widely used method in model-based decision support for identifying common input space properties across ensembles of exploratory model runs. For model runs with behavior over time, these properties are identified by reducing each run to a single value, which obscures potentially decision-relevant dynamics. We address the problem of considering dynamics in Scenario Discovery by applying time series clustering to the ensemble of model runs, and then finding the common input properties for each cluster. This separates the input space into multiple scenarios, each corresponding to a distinct model dynamic. Policy interventions can be targeted at different scenarios by analyzing overlap of these subspaces. Our work expands Scenario Discovery by improving consideration of system behavior over time, which is highly relevant for the management of complex nonlinear systems such as ecosystems or technical infrastructure.

5.2 Introduction

Many significant societal challenges are wicked problems (Rittel & Webber, 1973). A significant hurdle in understanding these challenges is the deep (Lempert et al., 2003) or Knightian (Knight, 1921) uncertainty surrounding them. That is, the various parties involved do not know or cannot agree on the key mechanisms of the system, associated probability distributions, and which outcomes are of interest (Lempert et al., 2003). This is especially relevant for socio-environmental and -technical systems which provide essential ecosystem services, but face increasingly uncertain and volatile futures (Helbing, 2013).

Because of the intrinsic complexity of many wicked problems, model-based analyses are useful for supporting decision making and planning (Holtz et al., 2015; Kwakkel, Walker, et al., 2016). Human cognition struggles in the face of complexity due to the misapprehension of feedbacks, accumulation, time delays, and emergence (Sterman, 1994). Simulation models can be used to augment human reasoning by assessing the consequences of multiple interacting (non-linear) processes (Sterman, 2002). However, this leaves unresolved the problem of uncertainty. Bankes (1993) suggested to use computational experimentation across many alternative realizations of the various irreducible uncertain factors to systematically map out the consequences of the uncertainties. Scenario Discovery (Bryant & Lempert, 2010; Kwakkel & Jaxa-Rozen, 2016) was suggested as a way of analyzing the results of these computational experiments in order to extract decision-relevant information from them.

Scenario Discovery is by now an established model-based technique for scenario development (Gerst et al., 2013; Halim et al., 2016; Hamarat et al., 2013; Kwakkel et al., 2013; Lamontagne et al., 2018; Lempert & Groves, 2010; McJeon et al., 2011; Moallemi et al., 2017; Parker et al., 2015; Rozenberg et al., 2014). It also forms the analytical heart of various approaches for model-based decision support under deep uncertainty (Helgeson, 2018; Herman et al., 2015; Kasprzyk et al., 2013; Kwakkel & Haasnoot, 2019; Lempert et al., 2006).

The term “scenario” has various definitions and meanings (Spaniol & Rowland, 2019). Scenario Discovery traces its roots to the Intuitive Logics (Bradfield et al., 2005) school of scenario planning (Bryant & Lempert, 2010). Analogous to Intuitive Logics, Scenario Discovery starts with the identification of uncertain factors. Next, Intuitive Logics groups these factors into clusters or mega trends and identifies the most uncertain, most impactful groups which will make up the scenario logic. For an example of how the narrative-first approach of Intuitive Logics can be used to reduce uncertainty, see Willis et al. (2018). In contrast, Scenario Discovery uses computational experimentation with models to explore the implications of all the uncertain factors jointly, then the outputs of the experiments are classified as being of interest or not, before trying to find orthogonal subspaces in the uncertainty space which are predictive of the experiments of interests. In Scenario Discovery, this subspace is called a ‘scenario’, although sometimes the individual experiments are also confusingly called scenarios. Abstractly put, however, this is only a difference in where and how the dimensionality of the uncertainty space is reduced into a smaller interpretable set. Scenario

Discovery thus reduces a problem's complexity by projecting the entire uncertainty space into a few salient dimensions, thus rendering cognitive benefits for decision makers (Bryant & Lempert, 2010) and enabling clearer analysis of decision options and trade-offs (Helgeson, 2018). This has also been described as model salience (van Voorn et al., 2016).

Central to Scenario Discovery is the classification of experiments as being of interest (or "decision-relevant") or not (Dalal et al., 2013; Lempert et al., 2008). In virtually all applications of Scenario Discovery, this is done by comparing an outcome of interest for each experiment with an external threshold. Experiments are considered of interest if the outcome meets (or fails to meet, as may be the case) this threshold. That is, the dynamics of the outcome of interest over time are ignored, with some even arguing that they are not relevant (Davis et al., 2007). Evaluating system states against such a static criterion is easy to conceptualize and communicate to stakeholders.

However, a problem's dynamics may be crucial for management and policy (Gotts et al., 2019; ten Broeke et al., 2017). That is, decision makers might very well care about the dynamics over time of the outcome of interest. Analytically, dynamics are also relevant. Often, in particular in case of non-linear models, different temporal dynamics originate from different regions of the model input space. Lumping these different dynamics together through a static criterion obfuscates the different origins of the types of dynamics. When using Scenario Discovery for designing strategies, as is common in Robust Decision Making, different temporal dynamics plausibly constitute different vulnerabilities. An inability to separate these vulnerabilities because of the use of a static criterion can result in an inability to succinctly describe the subspace(s) from which the experiments of interest originate. This will in turn hamper an analyst's ability to design robust strategies.

In short, aggregate statistics of time series may be misleading (Anscombe, 1973), or such time series may be equifinal but dynamically distinct (Von Bertalanffy, 1968), impeding the discovery of decision-relevant scenarios from them at a single point in time. This paper addresses the problem of accounting for dynamics over time within Scenario Discovery.

We present *behavior-based Scenario Discovery* as a method to address the problem of accounting for dynamics over time in Scenario Discovery. Rather than identifying a single subset of decision-relevant simulation experiments by evaluating an outcome of interest at a particular point in time against an external criterion, we apply time series clustering to the entire ensemble of experiment outputs, partitioning it into multiple behaviorally distinct subsets. Next, we can identify the subspaces from which these behaviorally distinct subsets originate. Each of these represents a distinct alternative future (or scenario), and can then be used to support scenario-based decision making (Bradfield et al., 2005; Kunc & O'Brien, 2017; Kwakkel et al., 2013; Lamontagne et al., 2018; Willis et al., 2018), or for computational decision support under deep uncertainty (Herman et al., 2015; Kasprzyk et al., 2013; Kwakkel & Haasnoot, 2019; Lempert et al., 2006). This approach builds on previous efforts to combine clustering and Scenario Discovery (Kwakkel et al., 2013). We extend this work with a more effective clustering algorithm, and consideration of the full output ensemble.

In Section 5.3, we review the Scenario Discovery literature in more detail. In Section 5.4, we propose an integration of time series clustering into Scenario Discovery to account for dynamics over time. In Section 5.5, we illustrate behavior-based Scenario Discovery with a case study on the impact of the shale gas revolution on future oil prices. In Section 5.6, we discuss the method's contributions to both Scenario Discovery in particular and model-based decision making in general, and give an outlook on future research. In section 5.7, we summarize our main conclusions and contributions.

5.3 Literature Review

Groves and Lempert (2007) and Lempert et al. (2006) are the first reported applications of what is now called Scenario Discovery. Lempert et al. (2008) introduced the name Scenario Discovery, while exploring the potential of various rule induction algorithms for analysing large ensembles of computational experiments with simulation models generated using Exploratory Modelling (Bankes, 1993; Kwakkel & Pruyt, 2013). Scenario Discovery as it is being used today was put forward by Bryant and Lempert (2010).

In Table 5.1, we summarise the general steps of Scenario Discovery. Scenario Discovery starts with large scale computational experimentation with one or more simulation models. These experiments are designed to systematically cover the space spanned by the many uncertain factors associated with a model. Next, the outputs of the model for each experiment are classified as being of interest or not. Third, a rule induction algorithm, typically the Patient Rule Induction Method (PRIM) (Friedman & Fisher, 1999), is used to identify orthogonal subspaces within the uncertainty space that have a high concentration of experiments of interest. Finally, the intervals for the various uncertain factors which jointly characterize the orthogonal subspace are interpreted and communicated, possibly in the form of narratives (see e.g., Greeven et al., 2016), but sometimes also directly with evident success (Gong et al., 2017). More in depth descriptions can be found in Bryant and Lempert (2010), and Kwakkel and Jaxa-Rozen (2016).

Scenario Discovery has been used in the context of climate adaptation (e.g. Kwakkel et al., 2015; Lempert & Groves, 2010), climate mitigation (e.g. Gerst et al., 2013; Greeven et al., 2016; Hamarat et al., 2013; Lamontagne et al., 2018; McJeon et al., 2011; Moallemi et al., 2017; Rozenberg et al., 2014), water resources management (e.g. Matrosov et al., 2013; Watson & Kasprzyk, 2017), transport and logistics (e.g. Halim et al., 2016), material scarcity (e.g. Kwakkel et al., 2013), and national security (e.g. Pruyt & Kwakkel, 2014). Scenario Discovery is also at the heart of various robust decision making approaches (Kasprzyk et al., 2013; Lempert et al., 2006), and adaptation pathways (Haasnoot et al., 2013; Kwakkel, Haasnoot, & Walker, 2016).

Scenario Discovery conventionally uses a binary classification to identify model outputs as being decision-relevant or not (Bryant & Lempert, 2010). In such cases, only a single scenario region of the input parameter space can be iden-

Table 5.1: Steps for conventional Scenario Discovery

Step	Action	Comments
1. Generation	Conduct simulation experiments on a system model by sampling from the input parameter space, passing these inputs to the model and recording the generated outputs.	The simulation model acts as a black-box generator function (Lempert et al., 2006), making this step agnostic to the underlying simulation paradigm (differential equation system, agent-based model, discrete event simulation, etc.) Input sampling should be uniform to ensure unbiased and complete coverage of the parameter space. Sampling density can vary, and should be chosen carefully based on computing infrastructure and model complexity (Davis et al., 2007; Moallemi et al., 2018; Pruyt & Islam, 2015).
2. Identification	Reduce each model output to a single value. Evaluate this value against an external criterion. Model outputs fulfilling the criterion are identified as being "of interest".	"Model output" refers to all data captured for a single simulation experiment, which may include multiple variables. Analysts must then identify which variable is especially descriptive of the system's overall condition, and how to reduce it to a single value. Depending on the context, this might be an end value, mean, amplitude, or other statistical simplification. The external criterion is a proxy or direct specification of stakeholder goals. It usually takes the form of a threshold value (Greeven et al., 2016), but could also be a range (Guivarch et al., 2016).
3. Rule Induction	Find the subspace (or region) in the input parameter space where the inputs generating the outputs of interest lie, by inducing its parameter rules.	Various rule induction algorithms exist Dalal et al., 2013; Guivarch et al., 2016; Kwakkel, 2019; Kwakkel and Jaxa-Rozen, 2016; Lempert et al., 2008. They generally function by drawing a bounding box around the entire input space, and then iteratively restricting the size of this box along one or more axes of the space. These restrictions are referred to as rules, since they prescribe which values a variable can take. The restriction process is governed by three key attributes of the induced subspace - coverage, density, and interpretability (Kwakkel, 2019; Lempert et al., 2006). Coverage refers to the ratio of decision-relevant over total inputs contained in the subspace, and should be maximized. Density captures the ratio of decision-relevant over decision-irrelevant inputs in the subspace, and should also be maximized. Interpretability describes the number of parameter space axes along which the box is restricted in size, and should be minimized. Rule induction is generally conducted once, as there is only a single subset of decision-relevant model outputs. However, these outputs may stem from multiple regions of the input space, requiring multiple passes (Bryant & Lempert, 2010; Guivarch et al., 2016). Repeated rule induction for the same subset of inputs can also improve the quality of the box describing their subspace (Kwakkel & Cunningham, 2016).

tified. This approach can limit the usefulness of Scenario Discovery, as scenario-based planning benefits from having multiple comparable and internally consistent scenarios with which to evaluate possible future developments and interventions (Schoemaker, 1993; Willis et al., 2018).

The Scenario Discovery literature contains only a small number of studies that have identified multiple distinct scenarios using Scenario Discovery. Both Rozenberg et al. (2014) and Guivarch et al. (2016) do not explicitly specify outcomes of interest, instead choosing to split the model output space into multiple regions representing different Shared Socioeconomic Pathways (O'Neill et al., 2014). For the model outputs in each region, Scenario Discovery is performed in turn to identify the underlying drivers. In a study on economic growth, energy consumption, and carbon emissions, Gerst et al. (2013) partitions the input parameter space by clustering model outputs along multiple decision-relevant output dimensions, and then using another rule induction algorithm, Classification and Regression Trees (CART) (Breiman et al., 1984; Lempert et al., 2008), to identify the input parameters most useful for distinguishing clusters. While analyzing conditions under which the European Emission Trading System would fail to meet its stated objectives, Hamarat et al. (2013) identify three different subspaces using Scenario Discovery. Based on an in depth explanation of why under each subspace, the emission trading system would fail, they conclude that these subspaces represent clearly distinct vulnerability scenarios. It is noteworthy that none of these aforementioned studies analyze relations between the identified scenarios, such as overlap or adjacency. In this work, we show that analysis of scenario overlap is both possible and can yield novel and policy-relevant insights.

5.4 Behavior-Based Scenario Discovery

Behavior-based Scenario Discovery is a variation of conventional Scenario Discovery, with the goal of deriving decision-relevant future scenarios from model behaviors. Behavior-based Scenario Discovery primarily entails a modification of step 2 (see Table 5.1). Instead of imposing a binary classification on a set of computational experiments based on model outputs, behavior-based Scenario Discovery applies time series clustering to the model outputs. This enables the separation of distinct model behaviors. In step 3, rule induction can be applied to each cluster of behavior which is deemed to be of interest given the purpose of the study.

Time series clustering aims at grouping a set of time series into two or more subsets with high similarity within each subset, and low similarity across the subsets (Cryer & Chan, 2008; Shumway & Stoffer, 2017). Liao (2005) identifies three main approaches to quantifying time series similarity: feature-based, data-based, and model-based methods. Feature-based methods identify salient properties of time series, such as amplitude or number of peaks (see e.g. Paparrizos & Gravano, 2015). Similarity between time series is then a function of these features. Data-based methods match individual data points between time series, and determine the points' similarity using various distance definitions (see e.g., Berndt & Clifford, 1994; Keogh & Ratanamahatana, 2005). Model-based methods replace each time series with a mathematical model of itself, such as a linear regression or Markov Chain Monte Carlo distribution, and then determine similarity based on the parameters of those models (see e.g. Corduas & Piccolo, 2008).

Based on the determined similarities, the set of time series can be grouped into a number of internally similar clusters. Clustering methods can be hard, soft, or hierarchical (Shumway & Stoffer, 2017). Hard clustering means every time series is included in exactly one cluster, while in soft clustering, time series are members of multiple clusters, to varying degrees (Cryer & Chan, 2008). Hierarchical clustering also assigns every time series to multiple clusters, but at different levels of aggregation (see e.g. Gerst et al. (2013) and Rodrigues et al. (2008)).

The starting point of behavior-based Scenario Discovery is a verified and validated system model, together with an input parameter space capturing the model's aleatory and epistemic uncertainties (Hoffman & Hammonds, 1994). The first step is identical to that of conventional Scenario Discovery - a number of simulation experiments are performed on a system model, randomly sampling from the input parameter space to generate outputs. In the second step, time-series clustering is applied to find common macro-level behaviors in the ensemble of output time series. In step 3, rule induction is performed for each time series cluster in turn.

The primary benefits of this approach over conventional Scenario Discovery are two-fold:

1. The induced regions in the input parameter space are associated with model behaviors over time, rather than a static snapshot of the model's state.

2. Every output in the ensemble is included in a cluster and mapped to a region in the input parameter space, rather than just a small subset.

In turn, we suggest that the induced regions in the input parameter space, since they are associated with a given type of dynamics over time, will be more clearly separable compared to the existing practice of using a static threshold based approach for identification of experiments which are of interest.

5.5 Case study: Behavior-based Scenario Discovery for future oil price dynamics

In this Section, we demonstrate behavior-based Scenario Discovery through a case study. To showcase its benefits, it is necessary to analyse a complex system model with both a significant number of uncertain factors, and highly nonlinear dynamics. Therefore, we use a study on climate mitigation policies, developed by Auping et al. (2016) using a System Dynamics approach (Forrester, 1961). The model includes both uncertain input parameters as well as uncertainty about model structure in order to capture the ambiguity of the global energy market, and, as will be shown, has highly nonlinear dynamics.

5.5.1 Generation

We re-used an existing data set created by Auping et al. (2016). The data includes 2000 simulation experiments generated using the Exploratory Modelling workbench (Kwakkel, 2017) under uniform Latin Hypercube sampling (McKay et al., 1979). Figure 5.1 illustrates the diversity of model behaviors that result from different combinations of input parameters. The wide variety of possible behaviors and the non-linearity of the model is apparent.

5.5.2 Identification of dynamics

We focused our analysis on one particularly interesting variable in the model outputs - future oil price (*Oil Price* $[\frac{\$}{BBTU}]$). The ensemble of outputs for this variable, shown in Figure 5.1, features a variety of dynamics, including oscillations and extreme outliers.

We applied time series clustering with typical settings to the ensemble of model outputs. We chose Complexity-Invariant Distance (CID) (Batista et al., 2014) as our similarity metric. In previous work (Steinmann, 2018), we compared a range of time series clustering metrics. We found that CID-based time series clustering was best able to identify tipping points for two simple nonlinear systems, one featuring a saddle-node bifurcation (Ludwig et al., 1978), the other a Hopf bifurcation (Strogatz, 2018). Given that CID performed best on these two

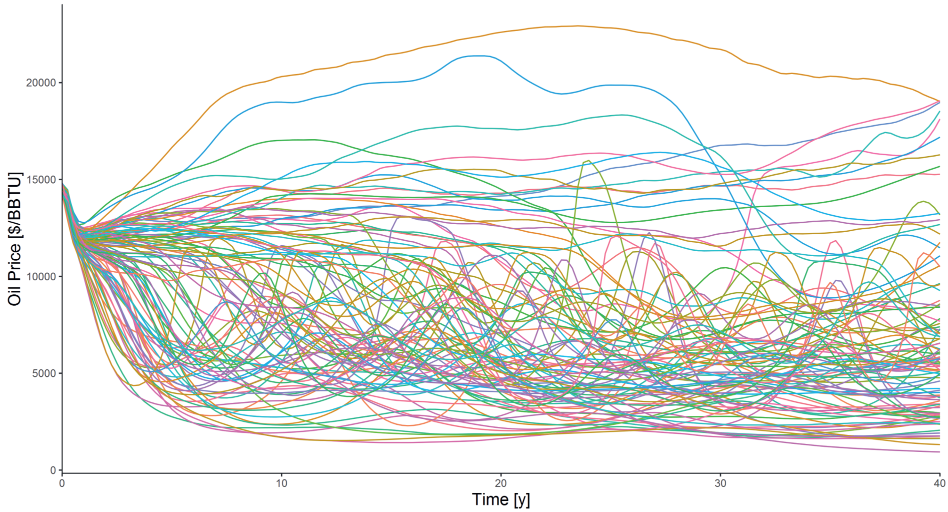


Figure 5.1: Data by Auping et al. (2016) shows a variety of nonlinear dynamics. 100 (of 2000) randomly selected outputs for the outcome of interest, each representing a possible alternative future oil price.

test problems, where we know analytically the correct answer, we expect that CID is likely to perform well on more complex, non-linear dynamical systems as well.

CID is based on Dynamic Time Warping (Berndt & Clifford, 1994), but introduces a correction factor to account for the time series complexity. For every pair of time series Q, C, the complexity-invariant distance between them is calculated as the summed Euclidean distance ED between their points, multiplied by a correction factor CF:

$$CID(Q,C) = ED(Q,C) * CF(Q,C)$$

The correction factor is the quotient of the time series' minimal and maximal complexity estimates:

$$CF(Q,C) = \frac{max(CE(Q), CE(C))}{min(CE(Q), CE(C))}$$

To estimate a time series' complexity, Batista et al. (2014) consider its path length:

$$CE(Q) = \sqrt{\sum_{i=1}^{n-1} (q_i - q_{i+1})^2}$$

Based on internal validation (Arbelaitz et al., 2013) of the solutions for different cluster counts, we selected a cluster count of $k = 6$. Hard clustering (i.e. each time series is assigned to a single cluster) was performed using the TSclust package (Montero & Vilar, 2014), which implements CID clustering for the R computing environment (R Core Team, 2018). More recently, CID-based time series clustering has also been implemented in the Python-based EMA workbench (Kwakkel, 2017). Figure 5.2 shows the six identified clusters, each representing a set of distinct future dynamics. Based on visual inspection it is apparent that CID-based time series clustering does a reasonably good job of separating different dynamics. For example, if we compare clusters #1 and #2 we see clearly different dynamics. Cluster #1 contains runs that show an early drop and then stay low, with some runs showing small oscillations. In contrast, cluster #2 contains runs where the price drops slightly, but then quickly recovers and stay high for the remainder of the simulation. For some of the other clusters, the differences are less striking. For example, both clusters #4 and #5 seem to contain many oscillating dynamics in the middle of the output range.

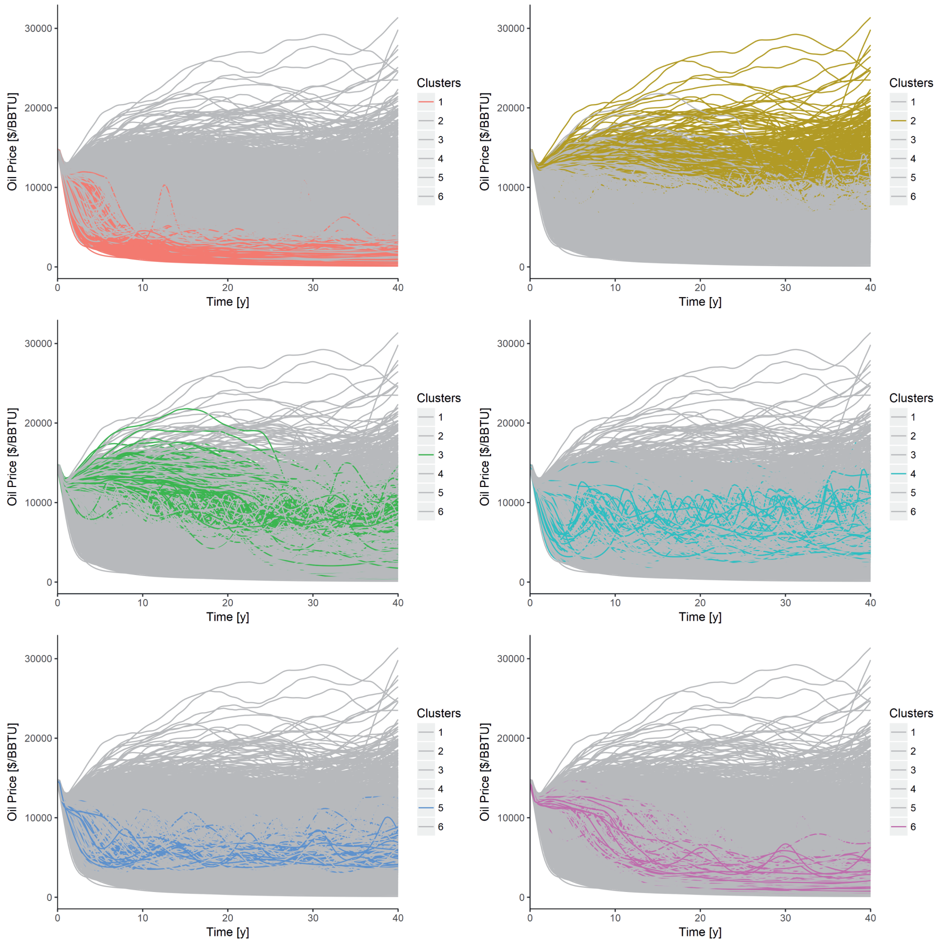


Figure 5.2: Six subsets of model dynamics, clustered using Complexity-Invariant Distance.

5.5.3 Rule Induction

We performed rule induction for each cluster using the Patient Rule Induction Method (PRIM) (Friedman & Fisher, 1999) as implemented in the Exploratory Modelling workbench (Kwakkel, 2017). In Table 5.2, we present the regions of the input parameter space for each cluster. PRIM identifies a region by restricting input parameter values with rules. We give the rules for each cluster along the five most commonly predictive dimensions, along with the coverage and density of each cluster's orthogonal box (see Table 5.1 for a formal definition of these concepts).

Table 5.2: Induced rules for most commonly predictive input parameters, per cluster

Cluster	**Rules**					**Box Statistics**	
	Switch price/supply dominance *{1, 2}*	Initial unit costs oil *[1000, 8000]*	Effect of supply shortage on GDP growth *[-0.3, 0]*	Switch legal emissions cap *{0, 1, 2, 3}*	Average throughput time stocks *[0.05, 0.2]*	Coverage *[0, 1]*	Density *[0, 1]*
1	1	[1002, 4945]	[-0.299, -0.088]	{0, 1, 2, 3}	[0.057, 0.2]	0.658	0.649
2	2	[1000, 8000]	[-0.27, 0]	0	[0.05, 0.2]	0.613	0.807
3	2	[1000, 8000]	[-0.22, 0]	{1, 2, 3}	[0.05, 0.2]	0.594	0.449
4	1	[3368, 7997]	[-0.3, -9.6e-05]	{0, 1, 2, 3}	[0.05, 0.11]	0.615	0.548
5	1	[2293, 7753]	[-0.299, -0.015]	{0, 1, 2, 3}	[0.084, 0.2]	0.724	0.766
6	2	[1002, 5819]	[-0.282, -0.028]	{1, 2, 3}	[0.05, 0.2]	0.655	0.555

For each of the six clusters, at least one of the five most predictive dimensions was not predictive at all, indicating that the parameters are only predictive for certain behavior modes, and only over a part of their uncertainty range. This underlines the inherent complexity and richness of the underlying model, and has both positive and negative aspects. While it may make it easier to target policy interventions on particular behaviors, it also may be harder to identify globally representative signposts and triggers for adaptive policy design. As an example, oil price will remain constant or increase (cluster #2) without a global emissions cap of some kind, due to unbridled demand and consumption. This indicates that an emissions cap may be an effective way to reduce oil prices long-term. However, there are also some cases in which prices stabilize without a cap, or exhibit undesirable behavior despite the presence of a cap. Similarly, cluster #5, which exhibits rapidly fluctuating oil prices, has a low average throughput time for stocks in the model. This mirrors the bullwhip effect sometimes found in stock-flow models or games such as the Beer Game (Meadows, 2007), and would indicate that enforcing slower throughput (e.g. by restricting trading) could reduce price fluctuations. However, there are also model runs with little price fluctuation and low throughput time, as well as large fluctuation despite high throughput time.

The induced rules for the *Switch legal emissions cap* variable also show how Scenario Discovery may be useful for model validation - while the categorical input is specified with four levels $\{0, 1, 2, 3\}$, only two can reliably be distinguished: 0 and $\{1, 2, 3\}$. This indicates that this part of the model structure may be unnecessarily specific - or conversely, that the emissions cap levels are not graduated well enough.

5.5.4 Analysis of Scenario Regions

One advantage of behavior-based Scenario Discovery is that policy interventions can be targeted at specific types of dynamics in light of the identified conditions under which they occur (i.e, the identified region in the input parameter space). To ensure that such interventions only affect the intended dynamics, the respective input parameter regions of each scenario must be well separated from other scenarios, with little to no overlap. Based on the explicit boundaries for each subspace region (or "box") induced by PRIM, we analysed the overlaps between them. As discussed in Section 5.3, this analysis, although potentially insightful from a multi-scenario perspective, had never previously been performed.

In Table 5.3, we give both inherent and relational properties for each induced box. Specifically, we list how many time series were attributed to the corresponding cluster, and how many of their inputs ended up inside the corresponding induced input region (true positives). We also show how many members of other clusters were included in the region (false positives).

Table 5.3: Members of input parameter regions

Region	Total in targeted cluster	Total in induced region	Included members of cluster 1	Included members of cluster 2	Included members of cluster 3	Included members of cluster 4	Included members of cluster 5	Included members of cluster 6
1	386	391	254	1	1	41	69	25
2	287	218	1	176	37	3	0	1
3	426	564	16	61	253	17	13	204
4	234	263	36	11	21	144	49	2
5	308	291	164	16	65	73	223	38
6	359	423	34	14	106	15	19	235

It is apparent that while each region predominantly includes the members of its targeted cluster, there is a significant number of false positives. We believe this is due to the orthogonality of the input space regions induced by PRIM, which has been criticized before (Auping, 2018; Quinn et al., 2017).

For policy design, overlap between the regions is a crucial point of analysis. To ensure an intervention targeting a specific scenario does not have unintended consequences for other input space regions, the regions must be clearly separated in the multidimensional input parameter space. In Table 5.4, we quantify the overlap between the six scenario regions by determining how many experiments they share.

Table 5.4: Shared experiments (overlap) for induced scenario regions

Boxes	1	2	3	4	5	6
1	391	0	0	65	208	0
2		218	0	0	0	0
3			564	0	0	324
4				263	100	0
5					291	0
6						423

It is apparent that the regions are separable - only 4 of 15 pairs overlap. The overlaps for box pairs #1 and #5, as well as #3 and #6, are noteworthy. The clustering solution in Figure 5.2 shows that the pairs have similar behavior, and Table 5.2 indicates they share structural uncertainty parameters. It is therefore not surprising that they overlap to some degree. However, it may still be interesting to distinguish them for scenario purposes, especially #3 and #6, as the latter shows a more severe oil price drop than the former, which could have significant policy implications (Smith, 2004).

5.6 Discussion

5.6.1 Case study results

In our case study, we chose to use Complex Invariant Distance in combination with hard clustering as the method for performing time series clustering. The central claims of this paper do not rest on this particular choice. It is conceivable that for other models, e.g. models characterized by more discrete temporal variations, other time series clustering algorithms are more suitable. Still, as evidenced by Figure 5.2 and the more in depth analysis of the separability of the clusters in the model input space, CID did perform quite well. As such, we suggest that CID is a promising first choice method for time series clustering in the context of behavior-based Scenario Discovery.

We combined CID with agglomerative clustering, which results in a hard assignment of individual time series to a specific cluster. The analysis of the overlap between the induced regions in Table 5.4 as well as the visual inspection suggests that a soft clustering might have been more appropriate. Only cluster 2 is perfectly separable in the model input space. While the other clusters are harder to separate both in the input space as well as having some apparent overlap based on visual inspection. It is unclear however how to embed a soft cluster assignment within Scenario Discovery, given its reliance on rule induction, which requires a binary classification.

5.6.2 Contribution to Scenario Discovery

Scenario Discovery is a computational approach for identifying consistent, decision-relevant futures (Bryant & Lempert, 2010). For the management of nonlinear systems, understanding the dynamics and behaviors of these futures is crucial (Sterman, 2000). However, conventional Scenario Discovery analyzes model outputs at a single point in time, ignoring these dynamics. With behavior-based Scenario Discovery, we present a method of deriving decision-relevant scenarios from these dynamics. Our approach builds on previous work regarding static vs. dynamic (or "transient") scenarios (Haasnoot et al., 2015), identifying common patterns in complex model outputs (Kwakkel et al., 2013), and inducing multiple decision-relevant input space regions (Gerst et al., 2013; Guivarch et al., 2016; Hamarat et al., 2013; Rozenberg et al., 2014). We combine these lines of research into a novel enhancement of Scenario Discovery, capable of identifying and characterizing behaviorally distinct futures. We see our work as a contribution to the ongoing discussion and development of Scenario Discovery in this journal (Bryant & Lempert, 2010; Kwakkel & Cunningham, 2016; Kwakkel et al., 2013; Kwakkel & Pruyt, 2013; Walker et al., 2010) and elsewhere.

5.6.3 Contribution to model-based decision support

The presented approach for applying Scenario Discovery to ensembles of dynamic model runs is of particular interest in the context of an ongoing discussion on the use of Scenario Discovery for developing global change scenarios. The existing SSP/RCP framework is based on a story line and simulate approach (Garb et al., 2008). First, internally consistent narratives are developed by groups of experts. Next, these narratives are translated into sets of inputs for specific integrated assessment models for quantitatively exploring specific Shared Socio Economic Pathways (O'Neill et al., 2014), or in combination with Reference Concentration Pathways.

Rozenberg et al. (2014) argued that the *a priori* specification of challenges to mitigation and adaptation is problematic, instead favouring *a posteriori* identification assisted by Scenario Discovery. One advantage of such a simulate and storyline approach (Greeven et al., 2016) is that it offer better guarantees of properly bounding the space of possible futures. Guivarch et al. (2016) methodologically developed this idea further. Lamontagne et al. (2018) demonstrated the feasibility of this approach by systematically sampling all possible combinations of drivers from the SSP dimensions, evaluating them using the integrated assessment model GCAM. They empirically confirmed the theoretical point raised by Rozenberg et al. (2014) - similarly looking challenges to adaptation and mitigation can arise from quite disparate combinations of driving forces. The approach presented in this paper offers a further methodological extension allowing for the explicit consideration of temporal dynamics.

5.6.4 Future Research

We see three main avenues for future work on behavior-based Scenario Discovery. Firstly, we hope to see our analytical method tested with other dynamic models to validate its general applicability. In particular, it may be interesting to consider stochastic models, such as agent-based or discrete event models, as it remains to been how sensitive our proposed approach is to noisy time series outputs. To this end, we have made our code publicly available both on GitHub (https://github.com/steipatr/BBSD-Public), and in the Exploratory Modelling Workbench (Kwakkel, 2017).

Secondly, the method presented in this paper uses univariate clustering to group the model outputs. However, multivariate clustering—grouping simulation experiments based on the similarities of multiple time-dependent model variables at once—may more reliably distinguish model dynamics by encompassing a wider range of system performance markers. It might also help integrate differing stakeholder opinions about which system variables should form the basis for deliberation—a core challenge in decision-making under deep uncertainty (Kwakkel & Haasnoot, 2019; Lempert et al., 2003; Walker et al., 2010). A similar extension from univariate to multivariate Scenario Discovery, motivated by the same considerations, has already been presented by Gerst et al. (2013) for analysis at a single point in time.

Thirdly, we found that scenario regions overlap in some cases. We believe this is due to inherent orthogonality limitations of the used rule induction algorithm, which have been discussed elsewhere (Auping, 2018; Quinn et al., 2017). Developing non-orthogonal rule induction methods, possibly based on genetic algorithms (Kwakkel, 2019), may reduce region overlap and improve separability, allowing more precise identification of the root causes of specific model behaviors.

5.7 Conclusions

We have presented behavior-based Scenario Discovery. Where the established practice of Scenario Discovery virtually always focuses on the outcomes of a model at a particular point in time, behavior-based Scenario Discovery enables the analysis of dynamics over time instead. This is relevant for several reasons. In the context of designing robust strategies through Robust Decision Making, different undesirable dynamics over time constitute different kinds of vulnerabilities. By analyzing the model outcomes at a particular point in time, analysts will be unable to differentiate these dynamically distinct vulnerabilities. Second, different types of dynamics over time often originate from distinct regions of the uncertainty space. Explicitly considering dynamics over time within Scenario Discovery will improve the partitioning of the uncertainty space that can be produced by the existing rule induction algorithms used within Scenario Discovery, such as PRIM.

Behavior-based Scenario Discovery replaces the existing threshold-based binary classification of computational experiments with a time series clustering step. Next, rule induction can be performed for each cluster of computational experiments. As demonstrated in our case study, behaviorally distinct future scenarios can be identified by clustering an ensemble of time series model outputs, and characterized by inducing and analyzing the subspaces in the uncertainty space from which they originate. These scenarios based on dynamics can then be used for a variety of model-based decision support purposes, including uncertainty exploration, adaptive policy design, and strategic planning.

We focused in this paper on the clustering of time series as a precursor to rule induction within Scenario Discovery. We suggest that our arguments in favor of this multinomial approach in case of temporal dynamics apply more broadly. For example, instead of imposing a binary classification on the results of a spatially explicit model, it might be more useful to first cluster based on spatial patterns, and then use Scenario Discovery on these clusters. In particular in the context of land use change models, as used for e.g., climate adaptation, such an approach might be fruitful for identifying the drivers behind distinct land use patterns.

Chapter 6

Scenario Search: Finding Diverse, Plausible and Comprehensive Scenario Sets for Complex Systems

Under revision as: Steinmann, P., Verstegen, J., van Voorn, G.A.K., Roman, S., and Ligtenberg, A. Scenario Search: Finding Diverse, Plausible and Comprehensive Scenario Sets for Complex Systems.

6.1 Abstract

Complex systems such as cities, energy grids, or the global climate have many plausible futures. Scenarios, or structured narratives of decision-relevant futures, are a common decision support tool for making the complexity and uncertainties of complex systems humanly interpretable. However, the effectiveness of scenario-based decision support depends in part on the usefulness of the selected scenarios. Here we show an optimization-based approach for generating scenarios that are specifically designed to be diverse, plausible, and comprehensive. We establish the advantages of our method by evaluating it against three previously proposed methods: scenario matrices, generic archetypes, and clustering. Our case study is Schelling's segregation model, a tractable yet behaviorally rich simulation of a complex system. Our results show the proposed optimization-based approach can generate more diverse, plausible, and comprehensive scenarios than existing approaches. The resulting scenarios may provide a more insightful and robust basis for policy decisions, especially for complex systems with emergent behavior, or where substantial uncertainties are present.

6.2 Introduction

Scenarios, or structured sets of plausible future narratives driven by external forces (Spaniol & Rowland, 2019), are commonly used for decision support in and across the social, technical, and environmental domains. As compelling and easy-to-grasp representations of how the future might develop, they have captured decision-makers' attention, and the public's imagination, in contexts including climate change (Nakićenović et al., 2000), pandemics (Skegg et al., 2021), and sea level rise (Wolters et al., 2018). They can be used for a variety of purposes, including presenting contrasting futures, identifying key uncertainties in systems, and evaluating policy alternatives (Bell, 2003), and are especially suited to long-term decision-making contexts (Pot et al., 2022).

A number of methods have been proposed for generating sets of scenarios which are useful for decision support. These methods generally rely on iterative interactions between scenario analysts, stakeholders, and domain experts to qualitatively identify performance indicators, causal relations, and external drivers of change. From these elements, scenarios can then be generated. However, such expert-driven approaches fail to identify policy-relevant scenarios in complex and deeply uncertain decision-making contexts, both because the range of possible outcomes is not knowable *a priori*, and because the most relevant scenarios might emerge from unexpected combinations of external forces (Dolan et al., 2021; Lamontagne et al., 2018). As McPhail et al. (2020) showed, the selection of scenarios for decision support can have substantial impact on the quantitative outcomes of the subsequent decision. Thus, we identify a knowledge gap regarding how to generate useful scenario sets when complexities and uncertainties are present.

In this paper, we address the highlighted research gap by introducing a new method for generating scenario sets for complex systems based on simulation-based optimization. We compare our method to three existing scenario generation approaches, and show that it performs best overall across three distinct criteria. Concurrently, we highlight several shortcomings in existing scenario generation methods. Finally, we discuss some implications for scenario-based planning in particular, and decision support in general.

6.3 Background

The Anthropocene is characterized by a wide variety of interdependent socio-technical-environmental systems such as energy infrastructure, financial markets, or agro-industrial production. These globally networked systems are both vulnerable and difficult to control, as disruptions can unexpectedly propagate to other domains (Helbing, 2013), cascade across levels of hierarchy (Iwanaga et al., 2022), and self-reinforce (Siegenfeld & Bar-Yam, 2020).

The challenges in design and governance of such systems are compounded by a lack of consensus on the relevant external drivers, internal causal relations, and outcomes of interest underlying a decision-making context. These *deep uncertain-*

ties (Lempert et al., 2003) amplify the difficulties of successful governance, especially in situations where ownership and control are contested between multiple actors (Gotts et al., 2019). The resulting gridlock may have critical consequences, as the *wickedness* of the decision problem affords little time for hesitation, and no possibility for a do-over (Rittel & Webber, 1973).

In order to make both the complexity and uncertainty inherent in these systems' governance comprehensible to decision-makers, a variety of decision support methods have emerged. A unifying theme across these methods is the usage of scenarios (Bell, 2003) - combinations of external drivers and resulting system narratives or outcomes. These narratives are internally consistent, plausible in the context of the studied system, and commonly appear in sets, allowing comparison between alternative futures.

A well-designed set of scenario summarizes the system's complexity, and the decision problem's uncertainties, by reducing the entirety of the future behavior space to a handful of comprehensible examples. Decision-makers can then focus on a few relevant alternatives, rather than worry about every permutation of plausible behavior. At the same time, careful selection of the included scenarios can challenge preconceived notions of the system's expected future by purposefully excluding "business as usual" futures (Voros, 2017) in favor of those requiring not only timely preparation and adaptation (Haasnoot et al., 2013), but also negotiation of distributive justice among current and future stakeholders (Jafino et al., 2021).

6.4 Theory

Sets of scenarios illustrate meaningfully different ways the future might plausibly develop. For such a set to be useful for a given decision-making context, the scenarios included in the set should be diverse, plausible, and comprehensive, as argued in the following section.

Diverse scenarios are meaningfully different alternatives to one another (Spaniol & Rowland, 2019), that is, they describe clearly distinguishable different future trajectories. Meaning stands in relation to the specific decision problem the analyst or stakeholder faces, and is derived from the legitimacy (Oreskes et al., 1994) or validity (ten Broeke & Tobi, 2021) of the conducted analysis - establishing that the proposed insights are useful to its audience. As Dolan et al. (2021) and Lamontagne et al. (2018) have highlighted, the meaningful or decision-relevant scenarios for complex systems are difficult to identify *a priori* - that is, without evaluating the behavior resulting from a system's causal relations.

At the same time, the presented alternative futures must be plausible, or within the scope of what could physically occur within the studied system - even if the probability is low. Establishing what is or is not plausible is difficult when studying complex systems, as even simple ones can exhibit any desired behavior pattern (Cook, 2004), to say nothing of the involved uncertainties (Funtowicz & Ravetz, 1993). Rittel and Webber (1973) emphasize that the futures of complex problems are not exhaustively describable, which is the limiting factor on our ability to predict their future behavior (Polhill et al., 2021). This is coupled with humans' limited capacity for "mental simulation", or the ability to reason about nonlinear interactions in complex systems (Sterman, 1994). Simulation models have become an attractive method for these input-output evaluations (de Regt & Parker, 2014), as the models can systematically explore the implications of large number of possible system configurations and assumptions (Bankes, 1993; Winsberg, 2010).

Finally, the considered scenarios should give a comprehensive overview of the system's plausible future trajectories. Otherwise, blind spots will be introduced into the decision process, with potentially disastrous results. The concept of a "futures cone", a layered arrangement containing the possible, preferable, predicted and/or projected futures extending forward in time, has been discussed by a number of authors including Voros (2017) and Maier et al. (2016). When making decisions under deep uncertainty, it may be appropriate to reason across the widest and most comprehensive range of this cone, encompassing all plausible outcomes (Derbyshire, 2020, 2022; Zatarain Salazar et al., 2022). This ensures that the resulting analysis is robust to whichever future eventually ends up materializing (Lempert et al., 2006; Rosenhead et al., 1972), up to and including the possibility of *black swans* - extreme (or even existential) risks with unknowable likelihood (Taleb, 2007).

A growing body of research on simulation-based scenario development is exploring how models can be used to improve scenario-based planning, broadly along two lines of research. The first line, which might be termed *behavior search*, focuses on how the plausible behavior space of models can be efficiently and comprehensively explored (Chérel et al., 2015; Davis et al., 2007; Islam & Pruyt, 2016; Pruyt & Islam, 2015). The second line, which is often referred to as *scenario discovery*, explores how specific model outcomes of interest can be related to regions of input space, for either one (Bryant & Lempert, 2010; Edali & Yücel, 2019; Kwakkel, 2019; Kwakkel et al., 2013; Stonedahl & Wilensky, 2011; ten Broeke et al., 2021) or multiple outcomes of interest (Jafino & Kwakkel, 2021; Steinmann, Auping, et al., 2020; Trindade et al., 2020). In the present work, we build on ideas proposed by Verstegen et al. (2017) to create a new method which straddles the two aforementioned lines of research - combining directed search (Kwakkel & Haasnoot, 2019) with many-objective optimization (Maier et al., 2019) to generate maximally diverse, plausible, and comprehensive scenario sets for complex systems.

6.5 Methods

6.5.1 Framework

In the following, we describe our proposed method for generating diverse, plausible, and comprehensive scenario sets, which we have named *scenario search*. We frame the challenge of generating such scenario sets as an optimization problem. Optimization simply means that we optimize an object function that represents the relevant optimization criteria. In this case, the function is a simulation-based

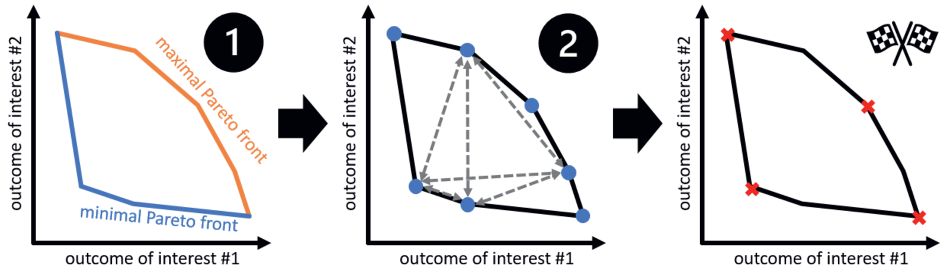


Figure 6.1: A visual representation of scenario search. In the first step, the maximal and minimal Pareto fronts across the model outcomes of interest are found through many-objective optimization. In the second step, all possible subsets of size k (here: $k = 4$) are generated from the points on the Pareto fronts, and the subset with the highest cumulative between-point distance is found through single-objective optimization. This subset then forms the final scenario set, which is maximally comprehensive, diverse, and plausible.

scenario generator, the inputs are the external drivers associated with those scenarios, and the criteria are the previously described diversity, plausibility, and comprehensiveness. We thus conceptualize a scenario as a combination of model inputs, and resulting simulation experiment outcomes or outputs.

In theory, we wish to simultaneously maximize our three criteria of diversity, plausibility, and comprehensiveness. In practice, the resulting optimization procedure would take a very long time to compute. Instead, we split the optimization into two steps - first optimizing for comprehensiveness, and then for diversity. We assume that the third criterion, plausibility, is given due to the underlying simulation model being validated and appropriate for the given decision-making context. As we outlined earlier, simulation models are assumed to already encode all plausible futures (and prohibit the unreachable ones), even though they are not known yet.

In the first step of scenario search (establishing comprehensiveness, see Figure 6.1), we define the simulation model's outputs of interest as the objectives and use many-objective optimization (Maier et al., 2019) to find the maximal and minimal Pareto fronts across combinations of outputs of interest. We then join these individual fronts together to form the Pareto hull, which encompasses all plausible outcomes of the model.

In the second step (establishing diversity), we search the model outcomes on the Pareto hull for the most diverse subset of a desired size. We measure diversity as the Euclidean distance between two points in the output space, rescaled to [0,1] along all axes. Because the Pareto hull contains only a handful of outputs, we can exhaustively calculate their distances and compare them without significant computational overhead. With this comparison, we identify the subset of most distant (i.e. diverse) model outputs. This subset then forms the scenario set, whose constituent outputs (or scenarios) are maximally diverse, plausible, and comprehensive.

The size of the final scenario set is an exogenous parameter in our method. This parameter, which we dub k as an analogy to a similar parameter in clustering, can be set based on audience and analyst desires for how many scenarios should be considered. In the presented work, we use $k = 4$ scenarios for two reasons. Pragmatically, when evaluating our method against other scenario generation methods, this is a convenient number for comparison. However, it also seems that four alternatives may be a limit of human working memory (Rouder et al., 2008), and therefore a practical upper bound for scenario-based planning with stakeholders.

6.5.2 Case study

To demonstrate and evaluate our proposed method, we draw upon a heavily studied model from the literature on complex adaptive systems, Schelling's segregation model (Schelling, 1971). This is a cellular automaton, or grid-based system in which each grid cell updates its properties based on the properties of the cells in its Moore neighborhood. In Schelling's model, two classes of cells exist. Cells seek to surround themselves with at least a certain number of neighboring cells of the same class, governed by the input *homophily*. If a cell does not have at least this many neighbors of its own class, it will relocate to a different grid location, the availability of which is controlled by the input *density*. When repeating this simple procedure for every grid cell over many time steps, macro-scale dynamics such as wastelands and neighborhoods emerge across the grid. This combination of model simplicity and behavioral richness makes Schelling's segregation model an attractive case study for us. In addition, it has a low run time, which may be desirable in model-based decision support (Helgeson et al., 2021).

In the context of our many-objective optimization, we specify the input space as [0.05,0.95] for *density* and [3,8] for *homophily* on a square lattice grid. We calculate two outputs of interest from the resulting spatial grid, *happiness* and *number of patches*. The former captures which fraction of all occupied grid cells have found at least their desired amount of same-class neighbors, and the latter describes how many neighborhoods (contiguous regions of same-class neighboring cells) have emerged. These are the two objectives which we maximize and minimize to find the Pareto hull. We choose these two outputs because they represent system state variables, or dynamic attributes of the system, which we deem of interest to decision-makers regarding segregation.

We perform the many-objective optimization using the ϵ-NSGA-II optimization algorithm (Kollat & Reed, 2006) implemented in the `Platypus` library (Hadka, D., 2015) for Python, and controlled through the Exploratory Modelling and Analysis Workbench (Kwakkel, 2017). Based on testing for convergence, we use 10 000 function evaluations (population size: 100) for the optimization with 10 replications each to account for the stochasticity in the model. All other parameters are left at `Platypus` default values. We limit the model to 100 time steps.

6.5.3 Experiment

To evaluate the effectiveness of scenario search, we compare it with three previously proposed methods for generating scenario sets with simulation models: scenario matrices, generic archetypes, and clustering.

Matrix-based scenario generation methods such as Intuitive Logics (Wright et al., 2013) generally start by identifying the most important external drivers of change. These are then clustered into a small number of axes or forces, for each of which high and low levels are determined. Across the axes, these levels form a matrix, hence the name. Each matrix cell then becomes an element of the scenario set, together with an accompanying narrative of how the world resulting from these driver levels would look. Scenario matrices thus reason from the drivers to the narratives, or, in a modelling sense, from the input to the output space. We apply this method by sampling the k corners of the model input space, representing high(est) and low(est) levels for every axis. Davis et al. (2007) previously advocated a similar approach in the context of behavior search. These corner points are then passed into the simulation model, and paired with their resulting outputs.

By contrast, scenario methods based around generic archetypes start by identifying a set of decision-relevant future narratives based on preexisting archetypes such as paradise, wastelands, or best-guess (Bezold, 2009; Dator, 2009). For each alternative narrative, the external drivers which might create that world can then identified. Thus, the reasoning is from the narratives to the drivers, or from the outputs to the inputs. We apply this method by estimating likely low and high values for every model output axis, which together form the k scenario outputs. In our case study, we selected {0.2,0.8} for *number of patches*, and {10,100} for *happiness*. We then find the input combination which generates the output closest to each desired scenario, and pair that input with the output to complete the scenario.

Finally, clustering has been explored by a number of researchers (Jafino & Kwakkel, 2021; Kwakkel et al., 2013; Rozenberg et al., 2014; Steinmann, Auping, et al., 2020) as a method of deriving scenario sets from large (computational) data sets. First, a number of simulation experiments are performed on a simulation model. Then, the resulting outputs are clustered, and a representative outcome is identified for each cluster, often using centrality or mean calculation. These representative outcomes then form the scenario set. We apply this method by conducting a uniform parameter sweep of the model, dividing the resulting outputs into k clusters using k-means (MacQueen, 1967), and identifying the cluster centroids, which may be thought of as representative of the cluster. As with the generic archetypes, we then find the best-matching input-output combinations for each cluster centroid. The parameter sweep underlying this clustering contains 3000 Latin Hypercube samples of the input space, with 10 replications each to average out the influence of the random initial patch arrangement.

For our evaluation, we draw upon the three scenario criteria introduced previously: diversity, plausibility, and comprehensiveness. As in the optimization procedure, we measure diversity as the Euclidean distance between two scenario points in the model output space. The output space is rescaled to [0,1] for all axes to give a common basis for comparison. In Equation 6.1, we give the diversity calculation for scenario points S_1, S_2 defined by Cartesian coordinates (x, y) in a two-dimensional Euclidean space.

$$D(S_1, S_2) = \sqrt{(S_{1,x} - S_{2,x})^2 + (S_{1,y} - S_{2,y})^2} \tag{6.1}$$

To determine plausibility, we measure the distance between each scenario's model output, and the closest model output generated by a parameter sweep of the model's input space. To facilitate the analysis, we construct the metric given in Equation 6.2 for the plausibility calculation for a scenario point S_1 and a set of parameter sweep outputs R. This calculation is again performed in the rescaled output space.

$$P(S_1) = \frac{1}{\min_{B \in R} D(S_1, B) + 1} \tag{6.2}$$

Finally, we measure comprehensiveness by calculating the proportion of the model's entire output range covered by the polygon spanned by the scenarios. In Equation 6.3, we give the area calculation for a scenario set S in a two-dimensional Euclidean space. Note that the scenarios must first be ordered clockwise.

$$A(S) = \frac{1}{2} \sum_{i=1}^{|S|} S_{i,x}(S_{i+1,y} - S_{i-1,y}) \tag{6.3}$$

6.6 Results

In the following section, we first describe the scenarios generated with the different methods, and then compare the scenario sets with each other using the three criteria of diversity, plausibility and comprehensiveness introduced earlier. Finally, we evaluate the overall effectiveness of the scenario generation methods by jointly considering the three criteria.

6.6.1 Scenario generation methods

Over our entire model exploration, the total number of patches ranges from 2 to roughly 150, while the happiness ranges from 0.0 to 1.0. The maximization of these two objectives results in a broadly S-shaped line with both convex and concave sections, whereas the minimization of the objectives results in a discontinuous front with an irregular shape (Figure 6.2). We note that the Pareto hull does not cover all outcomes generated by the parameter sweep, this does not have a substantial effect on the following analysis. The Pareto hull covers slightly less than half (49.9%) of the entire output space.

In the parameter sweep, six distinctive bands (grouped by the value of the *homophily* input) emerge, leaving large areas between them which are unreachable by the model. These bands are roughly aligned, but also intersect in some areas of the output space. Model outcomes are not evenly distributed, with higher densities in the corners of the output space.

The scenarios generated with the scenario matrices method appear in the four corners of the input space. In the output space, three of the four scenarios have very low happiness values, and few patches. The fourth scenario has high happiness, and also very few patches. Two of the scenarios are almost identical regarding happiness and number of patches, with correspondingly similar spatial maps. The two remaining spatial maps differ mainly in the granularity, with predominantly large and small neighborhoods, respectively.

The scenarios generated with scenario search are situated in the three corners of the Pareto hull, as well as roughly halfway along the maximization front. Their corresponding inputs roughly form a square, which is substantially smaller than the entire input space. The spatial representations of these scenarios show four distinct patterns, including dense fill with low granularity, dense fill with high granularity, sparse fill with high granularity, and sparse fill with regions of high and low granularity.

When considering the results of the generic archetypes method, we note that three of the four scenarios lie in regions of the output space which are unreachable by the underlying simulation model. Thus, there are no associated inputs or spatial representations of these three scenarios. The fourth scenario's spatial representation is characterized by a medium-density fill with high granularity, while its associated input lies roughly near the middle of the input space.

The scenarios generated with clustering are spread throughout the output space. Two of the four scenarios show a similar spatial pattern (low density and high granularity). The other two scenarios are distinct, with one showing low density and low granularity, and one medium-high density and regionally varying granularity.

6.6.2 Scenario Criteria

6.6.2.1 Diversity

The most diverse scenarios are created by scenario search (Figure 6.3), this scenario set having the largest intraset diversity, mean, and lower quartile values. Five of the intraset distances are roughly equal, with one longer outlier.

The scenarios created with a scenario matrix form two distinct and equally sized distance clusters, as three of the four scenarios are close together in the output space, and fourth is far away. Mean and lower quartile are the lowest, while the upper quartile is the highest of all four methods. Notably, one distance is close to 0, indicating these two scenarios are virtually identical regarding their outputs. Thus, distance in the input space does not translate into distance in the output space, highlighting the model's nonlinearity.

The generic archetype-based scenarios, being arranged in a square in the output space, have four identical shorter and two identical longer distances. Upper and lower quartiles are the closest together of all four methods.

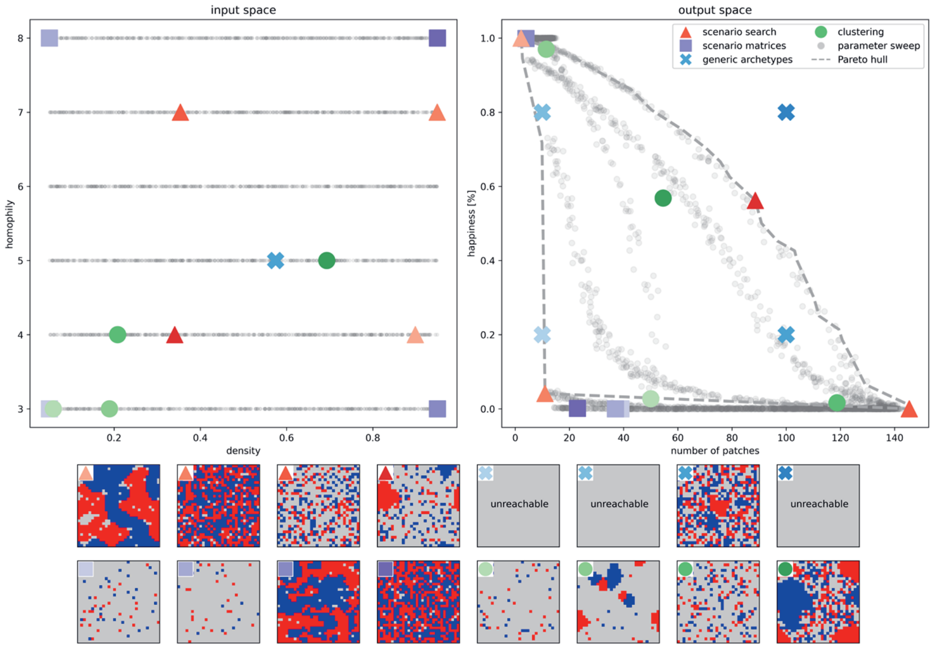


Figure 6.2: Input and output spaces of Schelling's segregation model. In each space, we show four sets of four scenarios each, one set per evaluated scenario generation method. The different methods are double-coded by color and marker shape, with the color hue distinguishing the four scenarios within each set. The markers in the in- and output spaces correspond. The underlying parameter sweep and Pareto hull are in grey. For each scenario in each set, an exemplary resulting spatial representation is shown, with the two agent classes in red and blue, and empty space in grey. Note that some markers are nearly overlapping in the output space, and that three markers are missing from the input space, as their corresponding outputs represent points which are unreachable for the model.

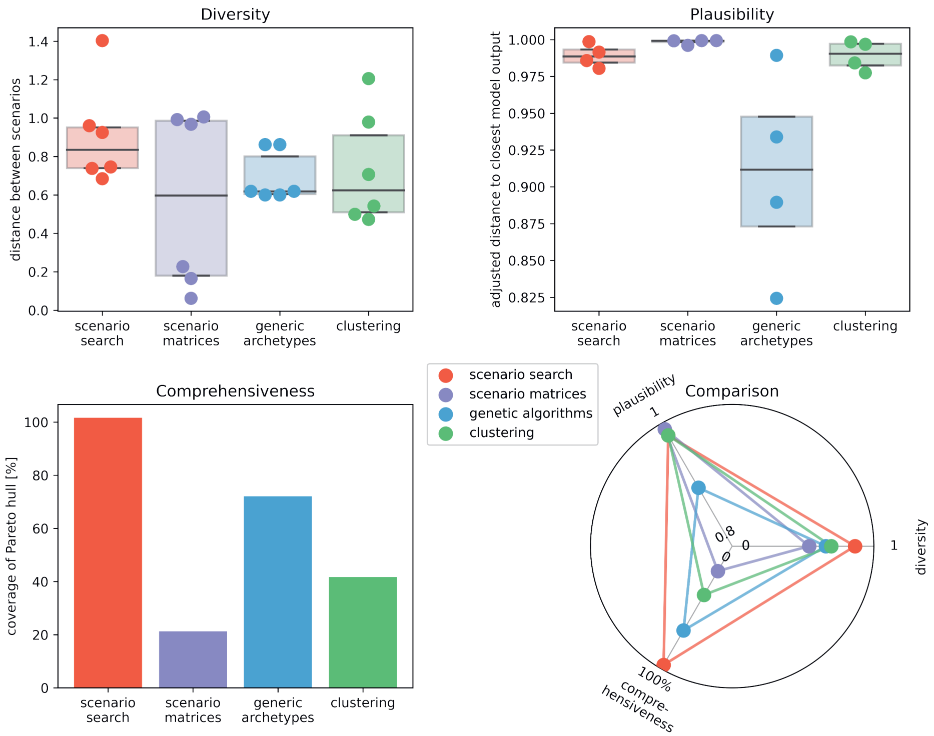


Figure 6.3: Evaluation of all four scenario generation methods against the three scenario set criteria. Where applicable, means and quartiles are represented with underlying box plots. The point markers are jittered to avoid overlap. The radar chart shows the criteria as polar axes in one figure, allowing overall comparison between the four different scenario generation methods.

The scenarios found with clustering have varying distances, with two flyers beyond the upper quartile, indicating some scenario pairs are far more diverse than others. The mean is roughly comparable to scenario matrices and generic archetypes methods, but lower than that of scenario search. This is because the representative cluster centroids by nature lie inward of the output space boundaries, and are therefore closer together.

Overall, the performance of scenario search, generic archetypes and clustering are all noticeably better than scenario matrices, with scenario search performing best.

6.6.2.2 Plausibility

The three model-based scenario generation methods (scenario search, scenario matrices, and clustering) all have comparably high plausibility scores. Furthermore, the scenarios are all within the Pareto hull of plausible model outcomes (Figure 6.2), indicating these scenarios could plausibly occur.

The generic archetypes method, which relies on *a priori* assumptions about the output space size and is therefore not strictly model-based, generates at least one impossible scenario - a hypothetical model state which is not actually reachable. Specifically, this scenario envisions a world in which both high happiness and high granularity (many patches) materialize. There are two more scenarios which, while they lie within the bounds of the Pareto hull and thus appear feasible, lie between the distinctive bands noted earlier, which the model also cannot reach.

Under the more narrow definition of plausibility mentioned above (within or near one of the bands in the output space), the clustering method also generates one scenario which is less plausible, even though the data underlying the clustering is entirely model-generated.

Overall, the three model-based methods (scenario search, scenario matrices, and clustering) perform substantially better than generic archetypes, with scenario matrices performing best by a small margin.

6.6.2.3 Comprehensiveness

Scenario search covers the output space most comprehensively. In fact, it even covers *more* of the output space than the Pareto hull, with over 105% coverage. This is because the Pareto hull is slightly concave for high patch numbers and low happiness (see Figure 6.2), which the calculated scenarios polygon does not account for.

The scenarios generated with a scenario matrix cover less than 20% of the Pareto hull, as they all have few patches (<40) and therefore miss most of the output space, which goes up to 150 patches.

The generic archetype scenarios span an area equal to almost 80% of the Pareto hull, which is the second-highest coverage. However, this is a generous calculation, since one of the scenarios included in this calculation lies outside the Pareto hull. Excluding it would reduce the coverage to around 50%.

	diversity	plausibility	comprehen-siveness	multi-plicative rank (score)	additive rank (score)
scenario search	1	3	1	1 (3)	1 (5)
scenario matrices	4	1	4	3 (16)	3 (9)
generic archetypes	3	4	2	4 (24)	3 (9)
clustering	2	2	3	2 (12)	2 (7)

Table 6.1: Rankings of the four scenario generation methods across the three scenario set criteria, with overall rankings computed using both multiplicative and additive scoring.

The clustering-generated scenarios omit the most extreme regions of the output space by necessity, and therefore span a polygon covering only roughly 40% of the Pareto hull's area. Notably, one side of this polygon is concave (decreasing its area slightly), as one scenario lies within the triangle spanned by the other three.

Overall, none of the methods apart from scenario search cover a substantial part of the entire Pareto hull. This is important because it shows that not only are many plausible futures not being considered, but that these not-considered futures are more extreme than the considered ones. In other words, the blind spots are more impactful than the "visible spots".

6.6.3 Comparison

When evaluating the four scenario generation methods across all three criteria (see radar chart in Figure 6.3), we find that scenario search scores best overall, scoring highest on diversity and comprehensiveness, and a close third on plausibility. The other three methods have varying performance across the three criteria, although they all perform poorly on at least one criterion.

By ranking the four scenario generation methods on each scenario set metric, and then combining these rankings into a global ranking, we can identify the best-performing method overall. The results are presented in Table 6.1. Using two different ranking methods, scenario search performs best overall, despite being punished for ranking a close third on plausibility. Clustering ranks second across both ranking methods, while the two model-free methods (scenario matrices and generic archetypes) perform worst.

6.7 Discussion

6.7.1 Scenario generation methods

Scenarios are widely used to support decision-making, but generating decision-relevant scenarios for complex and deeply uncertain systems is difficult. We therefore proposed a method which could computationally generate maximally diverse, plausible, and comprehensive scenario sets for such systems. We then evaluated this method against three existing scenario generation methods, and found that it performed best overall based on the three aforementioned criteria. In this section, we review our results and discuss their implications.

Overall, we find that scenario search generates the best scenario set, based on the three established criteria. Our proposed method scores best on diversity and comprehensiveness, and also performs very well on plausibility. Scenario matrices ranks third, with its most significant shortcomings being that the resulting scenarios are too similar, and that the range of plausible outcomes is poorly captured. The generic archetype-based scenarios rank last overall, failing to perform well on any criterion. Finally, clustering ranks second overall, performing reasonably well on two criteria, but failing to capture the most extreme plausible outcomes. The overall effectiveness of the two truly model-based methods (scenario search and clustering) indicates that simulation-based scenario generation may be a useful method for decision support, especially where complexity and deep uncertainty make mental simulation of the problem difficult.

Our analysis shows that at least for Schelling's segregation model, distance (which we interpret as diversity) in the input space does not translate into distance in the output space, and vice versa. The most distant input sets did not generate the most extreme outputs, and only one of the most extreme outputs lies against an edge of the input space. Supported by Lamontagne et al. (2018) and Dolan et al. (2021), we believe this generalizes to many (if not all) complex systems. By extension, existing scenario generation methods (e.g. scenario matrices or generic archetypes) may not be applicable to complex systems.

As Derbyshire (2022) argues, futures in which extreme risks or black swans materialize deserve more attention in decision-making than they currently receive. Including such extreme scenarios in scenario-based decision-making may be an effective method of doing so. However, as shown in Figure 6.2, existing scenario methods exclude the most extreme plausible scenarios, potentially blinding decision-makers to precisely those futures which require more attention. The underlying reason for this is different for every method. Matrix-based approaches cannot know *a priori* which input combinations will create extreme or otherwise decision-relevant futures, based on the system's inherent nonlinearities. Scenarios based on generic archetypes similarly presume *a priori* knowledge of the range of plausible system behaviors. Finally, clustering identifies representative scenarios by selecting the most centrally located model outputs for each cluster, and will therefore never select an edge case as a representative scenario. This further supports the notion that these methods may be insufficient for decision support where plausibility, rather than probability, is a focal point.

The goal of policy analysis is to assist decision-makers in choosing preferred courses of action, based on understanding the trade-offs between the consequences of alternative solutions (Walker, 2000). In this context, the Pareto hull, an intermediate result of our analysis, can be helpful to quantify these trade-offs (Verstegen, Jonker, et al., 2017). Furthermore, it is desirable to base such an analysis on future scenarios which could actually materialize. However, at least one, and potentially two, of the studied scenario generation methods produced scenarios which could never actually occur in the studied system. This may not only make the resulting decisions less robust and effective, but also erode trust in (computational) policy analysis as an analytical toolkit for effective decision support.

6.7.2 Limitations

There are a number of limitations to scenario search. Firstly, a simulation model must be used, which not only costs time and money to create, but may give a false sense of security about our understanding of the system's dynamics (Thompson & Smith, 2019). Secondly, the optimization requires an explicit definition of policy objectives. However, this is especially difficult under conditions of deep uncertainty (Lempert et al., 2003). Thirdly, the simulation model must contain the policy-relevant decision variables as inputs to actually generate useful insights. Finally, running the many-objective optimization is time-consuming even for simple models (Helgeson et al., 2021). In some decision-making contexts, this time may not be available, or rapidly evolving circumstances may invalidate simulation-based insights as quickly as they can be generated.

There are also methodological criticisms that can be levied against our analysis. By using Euclidean distance calculation for our optimization, we implicitly weight the two outputs of interest equally. This may not be appropriate in all situations, or there may be constraints limiting one or more outputs. Furthermore, we use a test case with a well-defined input space, which is unlikely to be the case for more complex problems. On top of that, and in line with the first limitation, our plausibility metric is based on the assumption that the simulation model is a reasonable representation of the real-world system it mimics, which may not be the case. Scenarios considered implausible by our approach, may therefore still be reachable in reality. Finally, an approximation of the Pareto hull could likely be found by drawing a convex hull around the results of a simple parameter sweep, eliminating the vast majority of function evaluations needed for the optimization. However, it is likely that this would not capture the most extreme and diverse scenarios plausible.

6.8 Conclusions

Scenario-based decision-making relies on sets of scenarios which are diverse, plausible, and comprehensive. In this paper, we presented a novel approach for generating such scenario sets which outperforms existing approaches based on a multi-criteria analysis. Our method, which we named scenario search, achieves this by applying a two-step optimization procedure to a simulation model of the studied system. Along the way, we showed that existing approaches may have significant flaws when applied to complex systems, including generating indistinguishable or nonsensical scenarios.

Based on the demonstrated effectiveness of our proposed method, and the shortcomings of existing methods, we advocate for an increased usage of simulation models when generating scenarios for decision support, especially where complex systems are concerned. At the same time, we urge that in those decision support contexts where matrix- or archetype-based scenarios are currently being used, that these scenarios be critically reviewed regarding their diversity, plausibility, and comprehensiveness. A key area where such a review might be necessary is the matrix-based set of Representative Concentration Pathways (van Vuuren et al., 2011) and Shared Socioeconomic Pathways (Riahi et al., 2017) that are widely used in climate modelling, and whose suitability and plausibility has been criticized (Pielke & Ritchie, 2021).

A key step in ensuring a sustainable, equitable and livable future for humanity is understanding and embracing the complexity and uncertainty present in the socio-technical-environmental systems surrounding us (Derbyshire, 2020). Where scenarios are used for decision-making and governance, they must also be generated in a way which acknowledges these inherent difficulties. While we have shown that existing methods may not be sufficient for this task, there is still much work to do in making more suitable concepts such as scenario search in particular, or simulation-based exploratory modelling (Bankes, 1993) in general, palatable to decision-makers. Therefore, we join Stanton and Roelich (2021) in highlighting the need for continued research on how model-based decision support can be integrated with the organisational and individual contexts of decision-making challenges.

Chapter 7

General Discussion

In this thesis, I explored methods for quantifying resilience under deep uncertainty. This final chapter contains the main findings of the individual preceding chapters, overarching themes of those chapters, implications for decision support practice, suggestions for future research, and a conclusion.

7.1 Findings of Individual Chapters

I studied the quantification of resilience under deep uncertainty along two lines. The first line of inquiry dealt with a condition I termed Multiplicity, or more formally, level 4a uncertainty. This level of uncertainty is present when there are many plausible metrics available for quantifying a system's resilience, none of which are inherently preferable. In Chapters 2, 3, and 4, we studied what types of different resilience metrics exist, how different metrics apply to a simple system, and whether ensembles of metrics are a suitable approach for overcoming the uncertainty of choosing a single resilience metric. I give a summary of these chapters below.

Chapter 2 described a systematic scoping review of resilience metrics for socio-ecological and -technical systems. We documented a variety of resilience metrics found in the peer-reviewed scientific literature, and grouped them into 10 distinct categories based on the underlying conceptual approaches to quantifying resilience, six for systems experiencing a single disturbance, and four for systems experiencing multiple sequential disturbances. We also identified four distinct categories of disturbances - sudden (e.g. earthquakes), continuous (e.g. droughts), multiple sequential (e.g. repeated droughts), and suddenly ending (e.g. bans on excessive resource extraction). Notably, the latter does not appear in at least two previously published frameworks for disturbance types (Collins et al., 2011; Lake, 2000). Finally, we found that there is little alignment between socio-technical systems being studied using "engineering resilience" metrics (i.e. within a single basin of attraction), and socio-ecological systems being studied using "ecological resilience" metrics (i.e. multiple basins of attraction), respectively. In other words, systems such as fisheries are studied as if the resource stocks can never be extinguished, and conversely, systems such as power grids are studied as if they can take on different configurations in response to a disturbance.

In Chapter 3, we studied the resilience of a set of stable cell patterns in the cellular automaton Game of Life, so-called still lifes. We did this by applying a variety of disturbances to the still lifes, and quantifying their response using different resilience metrics centered around the still lifes' pre- and post-disturbance spatial patterns. We found that still lifes which are highly resilient to a particular disturbance are often poorly resilient to other disturbances. We also observed that specific attributes of the still lifes, namely their number of connected components, their size, and their density were good predictors of their resilience. However, the strength and direction of the predictors varied depending on the chosen resilience metric, indicating that no single metric can usefully describe a still life's resilience, and that a still life's resilience is always specific to a given disturbance and metric.

In Chapter 4, we studied whether the sensitivity to the choice of resilience metric, established in the two previous chapters, could be overcome by using not a single metric to study a system's resilience, but an ensemble of metrics. This had been proposed in the scientific literature by several authors, but not applied in an optimization context yet. We applied a variety of disturbances to a simulation model of resource-consumer dynamics, and quantified the system's response using a variety of conceptually distinct resilience metrics. Both the disturbances and the metrics were based on the systematic scoping review performed in Chapter 2. Using many-objective optimization techniques, we identified system parameter settings which gave high resilience scores across all five utilized metrics for a given disturbance. We further showed that, with these multi-metric-optimal parameter settings, the system may be more resilient when experiencing other disturbances it was not optimized for, compared to parameter settings optimized with a single metric. This indicates it may be possible to prepare systems to cope with unforeseen future disturbances.

The second line of research dealt with a condition I termed Ignorance, or more formally, level 4b uncertainty, which is present when there is no knowledge whatsoever about suitable metrics for quantifying a system's resilience. Under such conditions, immediate quantification is not advisable, as choosing a metric without fully understanding its characteristic behavior might have undesired consequences for the resulting decision outcomes (Jain, 2009). In place of quantification, it may instead be appealing to qualitatively explore a system's plausible future behavior. This exploration can then be used as the basis for the analyst and/or stakeholders to identify what constitutes resilience in this system, and what appropriate metrics capturing this resilience might be - reducing the uncertainty level from 4b to 4a, and enabling quantitative methods to be used. In Chapters 5 and 6, I described two methods for exploring and summarizing a system's behavior using sets of scenarios. I recount these two chapters below.

In Chapter 5, we studied how the diverse dynamics of a complex system can be summarized using time series clustering and rule induction. We first clustered a simulation model's time series outputs, each representing a plausible future trajectory of the modelled system. We then performed rule induction for each cluster, linking it to the underlying input parameter ranges from which its constituent time series originated. Each of these clusters and associated generative input ranges can be thought of as a scenario representing some portion of the system's plausible behavior over time. We found it was possible to link clusters to highly specific parameter settings and model structure elements, exposing potential time-dependent vulnerabilities to analysts and stakeholders.

Finally, Chapter 6 describes a novel approach to exploring the behavioral diversity of a complex system's output space. We applied a two-step optimization procedure to a simulation model, with the goal of finding a small number of model outputs which summarize the model's entire range of potential outputs. These outputs can again be thought of as a scenario set, each output representing one plausible future of the system which is maximally distinct from all others. We compared this approach to three other methods for generating such scenario sets, including one method inspired by Chapter 5, and found that our proposed method performed equally or better than the existing methods on a number of criteria, creating more diverse, plausible, and comprehensive sets. These scenarios may be beneficial for showing analysts and stakeholders unexpected or extreme future states of the system.

7.2 Overarching Themes

From the different papers included in this thesis, a number of overarching themes and insights emerge. In the following, I describe these insights, framed as statements, and explain how they tie my and others' work together. I also highlight potential future research directions, and implications for model-based decision support.

Statement 1: In the presence of a multiplicity of resilience metrics, quantifying resilience using an ensemble of conceptually distinct metrics is a preferential solution to the problem of metric choice.

Selecting a metric when analysing a system's resilience is a sensitive task (Jain, 2009; Quinlan et al., 2016). Using an ensemble of conceptually distinct resilience metrics for decision making, an approach which we laid the groundwork for in Chapter 2, explored in Chapter 3, and demonstrated in Chapter 4, is a feasible solution to the problem of selecting a resilience metric, with the additional benefit of potentially making the system more resilient to novel disturbances. This addresses my first line of research: how to quantify resilience under conditions of multiplicity, or level 4a uncertainty.

In Chapter 2, we found that many alternative metrics exist for quantifying what is conceptually the same attribute of complex socio-technical and -ecological systems - their resilience. This observation is in line with the outcomes of other reviews of resilience metrics, such as those by Hosseini et al. (2016), Quinlan et al. (2016), and Sun et al. (2020). However, we believe our systematic approach and resulting conceptual classification are a useful contribution to the literature, addressing the need for conceptually distinct (or “independent”) resilient metrics highlighted in previous work (e.g. Kristensen et al. (2003) and Mcmillan et al. (2017)).

In Chapter 3, we took the idea of conceptually distinct resilience metrics from the systematic scoping review and applied it to a complex system experiencing a variety of disturbances. We found that the evaluation of the system's resilience depended heavily on the specific resilience metric and disturbance. This outcome resonates well with previous theoretical work framing resilience as specific to a particular system, place, time, disturbance, etc. (Carpenter et al. (2001), Cutter (2016), Meerow and Newell (2019)). Resilience analysis using multiple metrics simultaneously, as we did it, has also been reported by a number of authors, including Angeler and Allen (2016), Ingrisch and Bahn (2018), and Knippenberg et al. (2019). By investigating correlations across the different applied metrics and disturbances, we found that certain system parameter settings were more or less resilient across the different disturbances, implying that resilience might generalize after all.

In Chapter 4, we showed that using a set of conceptually distinct resilience metrics may enable the configuration of a system which is more resilient to disturbances it was not optimized for than a system which was optimized using a single metric. This may represent a significant step towards general resilience, or systems which can respond effectively to any disturbance they might experience, even ones they were never explicitly prepared for. Multi-metric reasoning for decision support in general has been proposed by Mitchell (2009) and Manheim (2023), among others. In the context of resilience, Mumby et al. (2014) and Duveneck and Scheller (2016) have proposed ensemble approaches to quantification. To our knowledge, we are the first to apply the ensemble approach in a resilience context using many-objective optimization. In a comparison of multi-objective decision analysis methods, Huang et al. (2011) found that, for the quality of the outcomes, the exact choice of method was secondary to the fact that multi-objective approaches were used in the first place. This is tentative support that our proposed multi-metric approach to optimizing resilience, which we applied to just one model, may generalize to other problem contexts.

Based on the outcomes of my first line of research, I recommend that resilience analysts evaluate a number of conceptually distinct metrics instead of indiscriminately selecting a single one. Chapter 2 of this thesis may be helpful in identifying candidate metrics. Furthermore, if the system's resilience is to be optimized, a many-objective optimization approach as demonstrated in Chapter 4 may improve the robustness of an optimization process, at least in terms of disturbances. However, the outcome of this is heavily dependent on the involved metrics. Future research should therefore investigate more thoroughly what correlation or similarity means in the context of (resilience) metrics, allowing a more thoroughly grounded identification of candidate metrics for an ensemble approach to quantifying resilience.

Statement 2: In the absence of any suitable resilience metrics, computational methods for scenario generation can explore and summarize a system's dynamics and the resulting vulnerabilities and risks.

Under the most severe conditions of deep uncertainty, analysts are totally ignorant of, or stakeholders in complete disagreement about, suitable metrics to quantify a system's resilience. With quantification impossible, only qualitative approaches remain, such as using scenarios to explore and summarize a system's plausible futures. Ideally, such scenario-based methods could be used to identify decision-relevant dynamics and vulnerabilities, as well as stakeholder preferences, which may then inform the future selection of resilience metrics. In Chapters 5 and 6, we developed and demonstrated novel methods for generating and identifying such scenarios. The resulting scenarios may be useful for exposing a system's characteristic behavior modes and vulnerabilities over space and time (Sterman, 2000), both of which are important for analyzing resilience (Smerlak & Vaitla, 2017; Zelnik et al., 2018). The two presented methods contribute to my second line of research: how to quantify resilience under conditions of total ignorance, or level 4b uncertainty.

Abandoning measurement may be a useful strategy under some circumstances, such as the absence of agreement or significant fear of the metric(s) being gamed (Manheim, 2023). It also prevents undesired outcomes stemming from an uninformed choice of metric (Boerlijst et al., 2013; Jain, 2009). However, it also puts useful methods of model-based decision support such as global sensitivity analysis (Saltelli & Homma, 1992; Steinmann, Wang, et al., 2020) or (many-objective) optimization (Kasprzyk et al. (2013), Chapter 4) out of reach. Thus, it may be desirable to use qualitative methods such as scenario analysis to engage in a dialogue with analysts and/or stakeholders about their perspectives on the system's resilience, and how this might be quantified. As Hitch (1955) pointed out, eliciting these objectives is the primary goal of any systems analysis. The process of building the simulation model underlying the generated scenarios may also contribute in this regard, serving as a repository of knowledge (Alexandra et al., 2023) or competing perspectives (Gotts et al., 2019) about the system, and help kick-start dialogue and joint reasoning among stakeholders about their shared system. With appropriate guidance (Moallemi et al., 2023), this may help create shared perspectives, and eventually enable a quantitative approach as demonstrated in Chapter 4, with all the benefits that brings.

Based on my second line of research, I recommend that analysts tasked with decision support for improving the resilience of a complex system use scenario methods to explore plausible future trajectories of the system in order to expose new insights about the system's behavior and the stakeholders' objectives. The method presented in Chapter 6 may be especially effective in this regard, as it outperforms common scenario generation approaches, as well as a simplified version of the clustering approach used in Chapter 5. Future research building on this work should investigate how sets of scenarios can be used to elicit stakeholder objectives and preferences specifically for resilience, for example by applying the behavior clustering presented in Chapter 5 to resilience curves (Poulin & Kane, 2021), or by applying the scenario search concept explored in Chapter 6 to ensembles of resilience metrics such as those in Chapter 4 to identify model parameter settings generating maximal disagreement between resilience metrics.

Statement 3: Resilience metrics are models.

Resilience is not an inherent property of a complex system, but an attribute that can be ascribed to (parts of) its dynamic behavior (Derissen et al., 2011; Meerow & Newell, 2019; Park et al., 2013). The exact conditions under which this attribute can be ascribed will differ based on the chosen metric. Therefore, it may be more useful to think of the various resilience metrics identified in Chapter 2 and applied in Chapters 3 and 4 not as some inviolable truth about the studied systems, but as (imperfect) models of what someone thinks resilience looks like in those systems.

A resilience metric, such as the return time of a system to its pre-disturbance performance level following a disturbance, is a simplified expression of the multidimensional concept of resilience - the ability of a system to withstand and recover from a disturbance. This relation maps well to the term “model” - a purposeful simplified representation of some other thing (Ackoff & Gharajedaghi, 1996). As Batty (2021) put it, models contain “the essence of the phenomenon under scrutiny for the particular purpose in mind”. One could therefore say that different resilience metrics are competing models of the concept “resilience”. As Thomas and Uminsky (2022) pointed out, metrics merely show what is important to the analysts who create them - it is therefore no surprise that different metrics exist, with no inherent primacy over one another. To be clear, resilience metrics-as-models are not *simulation* models like the case studies described in this thesis. However, they certainly represent the complex concept of resilience in a simplified, static form. In the framing of Thompson and Smith (2019), the concept of resilience is the Real World, and the metric is Model Land - and moving between these domains must be done carefully and explicitly.

Under this proposed framing of resilience metrics as models, many of the lessons formulated for modelling in general become immediately applicable to such metrics. As Levins (1966) pointed out, the truth in modelling sits at the intersection of independent lies - an interesting nod to the ensemble approach for resilience metrics we proposed in Chapter 4. Page (2016) and Batty (2021) similarly advocated for “many-model thinking”, which can only be interpreted as an ensemble approach. By framing metrics as models, scientists formulating resilience metrics may also be encouraged to take on some habits which the (simulation) modelling community has spent decades developing, such as documentation and reproducibility standards. This may alleviate several issues we encountered when studying resilience metrics, especially poor or lacking definitions of applied metrics (Chapter 2, Myers-Smith et al. (2012), Smaldino (2017)). Finally, by accepting that models are necessarily incomplete (Rosenblueth & Wiener, 1945) and serve more as a heuristic than a definitive representation (Oreskes et al., 1994), the urge to define or identify “the one” metric for resilience may be tempered or even extinguished.

Based on the framing resilience metrics as models outlined above, it would benefit analysts creating and using resilience metrics for decision support to recognize that their metrics are incomplete simplifications of their chosen interpretation of the concept of resilience. I recommend that they apply best practices from the simulation modelling community to their resilience metrics, such as documentation or the exploration of alternative hypotheses (read: metrics). Future research in this regard might investigate how documentation standards for simulation models, such as the ODD protocol (Grimm et al., 2010), could be conceptually translated for resilience metrics.

Statement 4: The paucity of generic multiple-attractor resilience metrics is a major obstacle to scientific progress in understanding and managing the resilience of complex systems.

Resilience metrics fall into two distinct categories - those operating within a single basin of attraction, and those operating across multiple basins of attraction. Holling (1996) referred to these two categories as "engineering resilience" and "ecological resilience", respectively. As we found out in Chapter 2, the former category is very well represented among published resilience metrics, and the latter quite poorly. This imbalance - or rather, the overall lack of multiple-attractor resilience metrics - became a genuine hurdle for our research described in Chapters 3 and 4.

In Chapter 2, we found that the majority of resilience metrics included in our synthesis, 37 out of 46 metrics (80%), operated within a single basin of attraction. This included 10 case studies of socio-ecological systems. One noteworthy example was the work of Harada et al. (1992), who studied the resilience of whale stocks to industrial whaling. In light of the extensive legal protections afforded to whales on account of their near-extinction due to whaling, modelling such a system with a single basin of attraction - that is, without the possibility of the population dying out - is at least naive, if not outright dishonest. In some sense, every system has at least two basins of attraction which are relevant for its behavior - existence and nonexistence. However, most systems we documented in our review were still studied using a single basin of attraction. In other words, their continued existence was never at threat. Of the few metrics identified through the systematic review which did allow for multiple basins of attraction, none of them were formulated generically enough that they could easily be translated to a different target system, unlike many of the single-attractor metrics. As a result, we were also not able to classify the multiple-attractor metrics.

In Chapter 3, we used only metrics with a single basin of attraction, even though many of the still lifes we examined disappeared ("died") when exposed to disturbances, in effect moving to a second basin of attraction. We also observed several still lifes which reached alternative "live" steady states, indicating that further basins of attraction existed. However, it was unclear how to quantify these moves between attractors in a principled manner. In Chapter 4, we similarly restricted ourselves to metrics with a single basin of attraction, as we wanted to ground our metric choices in the conceptual categories identified in Chapter 2. We also had to discard a number of simulation experiments in which the system did not reach its pre-disturbance performance level within the chosen time frame. Some of these experiments may have shown the system reaching a second basin of attraction, however, we were not able to quantify this with our chosen metrics. In both chapters, the lack of generic resilience metrics for systems with multiple basins of attraction hindered us from fully analyzing the studied systems.

As this statement describes an observed gap in the scientific literature, it is difficult to formulate positive recommendations for practitioners. This gap has previously been highlighted by Angeler and Allen (2016), Egli et al. (2019) and Fernandez and Ahmed (2019), among others. The future direction of research is clear: developing generic resilience metrics for systems with multiple basins of attraction will greatly enhance our ability to understand and manage the resilience of complex systems.

7.3 Conclusion

Based on the findings of this thesis, the following can be concluded: (1) if many resilience metrics are available, resilience can be quantified using an ensemble of conceptually distinct metrics, (2) if no resilience metrics are available, scenarios can be generated exploring the system's dynamics and vulnerabilities, (3) resilience metrics are models of the concept of resilience, and (4) the lack of generic resilience metrics allowing for multiple basins of attraction is hindering scientific progress in managing complex systems. The work in this thesis adds important knowledge to the fields of resilience and model-based decision support. Advances at the intersection of these fields may result in more resilient socio-technical and -ecological systems, and ultimately a planet better prepared for an uncertain and volatile future.

Bibliography

Ackoff, R. L., & Gharajedaghi, J. (1996). Reflections on systems and their models. *Systems Research*, *13*(1), 13–23.

Adamatzky, A. (1998). Universal dynamical computation in multidimensional excitable lattices. *International Journal of Theoretical Physics*, *37*(12), 3069–3108.

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378.

Alexandra, C., Daniell, K. A., Guillaume, J., Saraswat, C., & Feldman, H. R. (2023). Cyber-physical systems in water management and governance. *Current Opinion in Environmental Sustainability*, *62*, 101290.

Angeler, D. G., & Allen, C. R. (2016). Quantifying resilience. *Journal of Applied Ecology*, *53*(3), 617–624.

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, *27*(1), 17–21.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, *46*(1), 243–256.

Arreguín-Sánchez, F., & Manickchand-Heileman, S. (1998). The trophic role of lutjanid fish and impacts of their fisheries in two ecosystems in the Gulf of Mexico. *Journal of Fish Biology*, *53*(sA), 143–153.

Arrow, K., Bolin, B., Costanza, R., Dasgupta, P., Folke, C., Holling, C., Jansson, B.-O., Levin, S., Mäler, K.-G., Perrings, C., & Pimentel, D. (1995). Economic growth, carrying capacity, and the environment. *Ecological Economics*, *15*(2), 91–95.

Auping, W. L., Pruyt, E., de Jong, S., & Kwakkel, J. H. (2016). The geopolitical impact of the shale revolution: Exploring consequences on energy prices and rentier states. *Energy Policy*, *98*, 390–399.

Auping, W. (2018). *Modelling Uncertainty* (Doctoral dissertation). Delft University of Technology.

Austin, A. K., Berlekamp, E. R., Conway, J. H., & Guy, R. K. (1982). *Winning ways for your mathematical plays*. Academic Press.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454.

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

Batista, G. E. A. P. A., Keogh, E. J., Tataw, O. M., & de Souza, V. M. A. (2014). CID: An efficient complexity-invariant distance for time series. *Data Mining and Knowledge Discovery*, *28*(3), 634–669.

Batty, M. (2021). Multiple models. *Environment and Planning B: Urban Analytics and City Science*, *48*(8), 2129–2132.

Beccari, B. (2016). A Comparative Analysis of Disaster Risk, Vulnerability and Resilience Composite Indicators. *PLoS Currents*, *8*, ecurrents.dis.453df025e34b682e9737f95070f9b970.

Beer, R. D. (2015). Characterizing autopoiesis in the Game of Life. *Artificial Life*, *21*(1), 1–19.

Beer, R. D. (2019). Bittorio revisited: Structural coupling in the Game of Life. *Adaptive Behavior*.

Bell, W. (2003). *Foundations of futures studies: Human science for a new era*. Transaction Publishers.

Bergström, J., van Winsen, R., & Henriqson, E. (2015). On the rationale of resilience in the domain of safety: A literature review. *Reliability Engineering & System Safety*, *141*, 131–141.

Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *AAAI Technical Report WS-94-03*, (16), 359–370.

Bezold, C. (2009). Aspirational Futures: Understanding Threats, Opportunities and Visionary Possibilities. *Journal of Futures Studies*, *13*(4), 81–90.

Biggs, R., Schlüter, M., Biggs, D., Bohensky, E. L., BurnSilver, S., Cundill, G., Dakos, V., Daw, T. M., Evans, L. S., Kotschy, K., Leitch, A. M., Meek, C., Quinlan, A., Raudsepp-Hearne, C., Robards, M. D., Schoon, M. L., Schultz, L., & West, P. C. (2012). Toward Principles for Enhancing the Resilience of Ecosystem Services. *Annual Review of Environment and Resources*, *37*(1), 421–448.

Boerlijst, M. C., Oudman, T., & Roos, A. M. d. (2013). Catastrophic Collapse Can Occur without Early Warning: Examples of Silent Catastrophes in Structured Ecological Models. *PLOS ONE*, *8*(4), e62033.

Bohensky, E. L. (2008). Discovering Resilient Pathways for South African Water Management: Two Frameworks for a Vision. *Ecology and Society*, *13*(1).

Bradfield, R., Wright, G., Burt, G., Cairns, G., & Van Der Heijden, K. (2005). The origins and evolution of scenario techniques in long range business planning. *Futures*, *37*(8), 795–812.

Brand, F. S., & Jax, K. (2007). Focusing the Meaning(s) of Resilience: Resilience as a Descriptive Concept and a Boundary Object. *Ecology and Society*, *12*(1), art23.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.

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.

Capcarrère, M. S., & Sipper, M. (2001). Necessary conditions for density classification by cellular automata. *Physical Review E*, *64*(3), 036113.

Carpenter, S. R., Westley, F., & Turner, M. G. (2005). Surrogates for Resilience of Social–Ecological Systems. *Ecosystems*, *8*(8), 941–944.

Carpenter, S., Walker, B., Anderies, J. M., & Abel, N. (2001). From Metaphor to Measurement: Resilience of What to What? *Ecosystems*, *4*(8), 765–781.

Chérel, G., Cottineau, C., & Reuillon, R. (2015). Beyond Corroboration: Strengthening Model Validation by Looking for Unexpected Patterns. *PLOS ONE*, *10*(9), e0138212.

Chu, G., & Stuckey, P. J. (2012). A complete solution to the maximum density still life problem. *Artificial Intelligence*, *184-185*, 1–16.

Chu, G., Stuckey, P. J., & de la Banda, M. G. (2009). Using relaxations in maximum density still life. In I. P. Gent (Ed.), *Principles and practice of constraint programming* (pp. 258–273). Springer.

Cika, A., Cohen, E., Kruszewski, G., Seet, L., Steinmann, P., & Yin, W. (2020). Resilient Life: An Exploration of Perturbed Autopoietic Patterns in Conway's Game of Life. *The 2020 Conference on Artificial Life*, 656–664.

Collins, S. L., Carpenter, S. R., Swinton, S. M., Orenstein, D. E., Childers, D. L., Gragson, T. L., Grimm, N. B., Grove, J. M., Harlan, S. L., Kaye, J. P., Knapp, A. K., Kofinas, G. P., Magnuson, J. J., McDowell, W. H., Melack, J. M., Ogden, L. A., Robertson, G. P., Smith, M. D., & Whitmer, A. C. (2011). An integrated conceptual framework for long-term social–ecological research. *Frontiers in Ecology and the Environment*, *9*(6), 351–357.

Committee on Increasing National Resilience to Hazards and Disasters, Policy and Global Affairs, & Committee on Science, Engineering, and Public Policy (Eds.). (2012). *Disaster Resilience: A National Imperative*. National Academies Press.

Cook, M. (2004). Universality in elementary cellular automata. *Complex systems*, *15*(1), 1–40.

Corduas, M., & Piccolo, D. (2008). Time series clustering and classification by the autoregressive metric. *Computational statistics & data analysis*, *52*(4), 1860–1872.

Cryer, J. D., & Chan, K.-S. (2008). *Time series analysis* (2nd ed.). Springer-Verlag New York.

Cutter, S. L. (2016). Resilience to What? Resilience for Whom? *The Geographical Journal*, *182*(2), 110–113.

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. *Global Environmental Change*, *18*(4), 598–606.

Dalal, S., Han, B., Lempert, R., Jaycocks, A., & Hackbarth, A. (2013). Improving scenario discovery using orthogonal rotations. *Environmental Modelling & Software*, *48*, 49–64.

Dator, J. (2009). Alternative Futures at the Manoa School. *Journal of Futures Studies*, *14*(2), 1–18.

Davis, P. K., Bankes, S. C., & Egner, M. (2007). *Enhancing strategic planning with massive scenario generation: Theory and experiments*. RAND National Security Research Division.

Derbyshire, J. (2020). Answers to questions on uncertainty in geography: Old lessons and new scenario tools. *Environment and Planning A: Economy and Space*, *52*(4), 710–727.

Derbyshire, J. (2022). Increasing Preparedness for Extreme Events using Plausibility-Based Scenario Planning: Lessons from COVID-19. *Risk Analysis*, *42*(1), 97–104.

de Regt, H. W., & Parker, W. S. (2014). Introduction: Simulation, Visualization, and Scientific Understanding. *Perspectives on Science*, *22*(3), 311–317.

Derissen, S., Quaas, M. F., & Baumgärtner, S. (2011). The relationship between resilience and sustainability of ecological-economic systems. *Ecological Economics*, *70*(6), 1121–1128.

Desjardins, E., Barker, G., Lindo, Z., Dieleman, C., & Dussault, A. C. (2015). Promoting Resilience. *The Quarterly Review of Biology*, *90*.

Dolan, F., Lamontagne, J., Link, R., Hejazi, M., Reed, P., & Edmonds, J. (2021). Evaluating the economic impact of water scarcity in a changing world. *Nature Communications*, *12*(1), 1915.

Dore, M. H. I., & Webb, D. (2003). Valuing Biodiversity: Reality or Mirage? *Environmental Monitoring and Assessment*, *86*(1), 91–104.

Duveneck, M. J., & Scheller, R. M. (2016). Measuring and managing resistance and resilience under climate change in northern Great Lake forests (USA). *Landscape Ecology*, *31*(3), 669–686.

Edali, M., & Yücel, G. (2019). Exploring the behavior space of agent-based simulation models using random forest metamodels and sequential sampling. *Simulation Modelling Practice and Theory*, *92*, 62–81.

Egli, L., Weise, H., Radchuk, V., Seppelt, R., & Grimm, V. (2019). Exploring resilience with agent-based models: State of the art, knowledge gaps and recommendations for coping with multidimensionality. *Ecological Complexity*, *40*, 100718.

Elkies, N. D. (1999). The still-life density problem and its generalizations. *arXiv:math/9905194*.

Fernandez, G., & Ahmed, I. (2019). "Build back better" approach to disaster recovery: Research trends since 2006. *Progress in Disaster Science*, *1*, 100003.

Filatova, T., Polhill, J. G., & van Ewijk, S. (2016). Regime shifts in coupled socio-environmental systems: Review of modelling challenges and approaches. *Environmental Modelling & Software*, *75*, 333–347.

Folke, C. (2006). Resilience: The emergence of a perspective for social–ecological systems analyses. *Global Environmental Change*, *16*(3), 253–267.

Folke, C., Carpenter, S., Walker, B., Scheffer, M., Elmqvist, T., Gunderson, L., & Holling, C. (2004). Regime Shifts, Resilience, and Biodiversity in Ecosystem Management. *Annual Review of Ecology, Evolution, and Systematics*, *35*(1), 557–581.

Forrester, J. W. (1961). *Industrial dynamics*. Pegasus Communications.

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, *9*(2), 123–143.

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures*, *25*(7), 739–755.

Funtowicz, S. O., & Ravetz, J. R. (1994). Uncertainty, complexity and post-normal science. *Environmental Toxicology and Chemistry*, *13*(12), 1881–1885.

Garb, Y., Pulver, S., & VanDeveer, S. D. (2008). Scenarios in society, society in scenarios: Toward a social scientific analysis of storyline-driven environmental modeling. *Environmental Research Letters*, *3*(4), 045015.

Gardner, M. (1970). Mathematical games - the fantastic combinations of John Conway's new solitaire game "Life". *Scientific American*, *223*, 120–123.

Gerst, M., Wang, P., & Borsuk, M. (2013). Discovering plausible energy and economic futures under global change using multidimensional scenario discovery. *Environmental Modelling & Software*, *44*, 76–86.

Gold, D. F., Reed, P. M., Trindade, B. C., & Characklis, G. W. (2019). Identifying Actionable Compromises: Navigating Multi-City Robustness Conflicts to Discover Cooperative Safe Operating Spaces for Regional Water Supply Portfolios. *Water Resources Research*, *55*(11), 9024–9050.

Gong, M., Lempert, R., Parker, A., Mayer, L., Fischbach, J., Sisco, M., Mao, Z., Krantz, D., & Kunreauther, H. (2017). Testing the scenario hypothesis: An experimental comparison of scenarios and forecasts for decision support in a complex decision environment. *Environmental Modelling & Software*, *91*, 135–144.

Gotts, N. M. (2003). Self-organized construction in sparse random arrays of Conway's Game of Life. *New Constructions in Cellular Automata. Oxford University Press, New York*, 1–53.

Gotts, N. M., van Voorn, G. A., Polhill, J. G., Jong, E. d., Edmonds, B., Hofstede, G. J., & Meyer, R. (2019). Agent-based modelling of socio-ecological systems: Models, projects and ontologies. *Ecological Complexity*, *40*, 100728.

Goucher, A. P. (2015). Catagolue [Accessed: 2019-10-25].

Goucher, A. P. (2017). Lifelib [Accessed: 2019-10-22].

Greeven, S., Kraan, O., Chappin, É. J. L., & Kwakkel, J. H. (2016). The emergence of climate change mitigation action by society: An agent-based scenario discovery study. *Journal of Artificial Societies and Social Simulation*, *19*(3).

Griffeath, D., & Moore, C. (2003). *New constructions in cellular automata*. Oxford University Press.

Grimm, V., & Wissel, C. (1997). Babel, or the ecological stability discussions: An inventory and analysis of terminology and a guide for avoiding confusion. *Oecologia*, *109*(3), 323–334.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling*, *221*(23), 2760–2768.

Grimm, V., Schmidt, E., & Wissel, C. (1992). On the application of stability concepts in ecology. *Ecological Modelling*, *63*(1), 143–161.

Groves, D., & Lempert, R. (2007). A new analytic method for finding policy-relevant scenarios. *Global Environmental Change*, *17*(1), 73–85.

Guivarch, C., Rozenberg, J., & Schweizer, V. (2016). The diversity of socio-economic pathways and co2 emissions scenarios: Insights from the investigation of a scenarios database. *Environmental Modelling & Software*, *80*, 336–353.

Gunderson, L. (2010). Ecological and Human Community Resilience in Response to Natural Disasters. *Ecology and Society*, *15*(2).

Haasnoot, M., Schellekens, J., Beersma, J. J., Middelkoop, H., & Kwadijk, J. C. J. (2015). Transient scenarios for robust climate change adaptation illustrated for water management in the netherlands. *Environmental Research Letters*, *10*(10), 105008.

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, *23*(2), 485–498.

Hadka, D. (2015). Platypus.

Halim, R. A., Kwakkel, J. H., & Tavasszy, L. A. (2016). A scenario discovery study of the impact of uncertainties in the global container transport system on european ports. *Futures*, *81*, 148–160.

Hamarat, C., Kwakkel, J., & Pruyt, E. (2013). Adaptive robust design under deep uncertainty. *Technological Forecasting and Social Change*, *80*(3), 408–418.

Harada, Y., Sakuramoto, K., & Tanaka, S. (1992). On the stability of the stock-harvesting system controlled by a feedback management procedure. *Population Ecology*, *34*(1), 185–201.

Harvey, I. (2019). Habeas corpus: The ins and outs of autopoiesis. *Adaptive Behavior*.

Helbing, D. (2013). Globally networked risks and how to respond. *Nature*, *497*(7447), 51–59.

Helgeson, C. (2018). Structuring decisions under deep uncertainty. *Topoi*, *39*, 257–269.

Helgeson, C., Srikrishnan, V., Keller, K., & Tuana, N. (2021). Why Simpler Computer Simulation Models Can Be Epistemically Better for Informing Decisions. *Philosophy of Science*, *88*(2), 213–233.

Herman, J. D., Reed, P. M., Zeff, H. B., & Characklis, G. W. (2015). How should robustness be defined for water systems planning under change? *Journal of Water Resources Planning and Management*, *141*(10), 04015012.

Hitch, C. J. (1955). *On the Choice of Objectives in Systems Studies* (tech. rep.). The RAND Corporation.

Hoffman, F. O., & Hammonds, J. S. (1994). Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis*, *14*(5), 707–712.

Holling, C. S. (1973). Resilience and Stability of Ecological Systems. *Annual Review of Ecology and Systematics*, *4*(1), 1–23.

Holling, C. S. (1996). Engineering resilience versus ecological resilience. *Engineering within ecological constraints*, *31*(1996), 32.

Holling, C. S., & Gunderson, L. H. (2002). *Resilience and adaptive cycles*. Washington, D.C.: Island Press.

Holtz, G., Alkemade, F., de Haan, F., Köhler, J., Trutnevyte, E., Luthe, T., Halbe, J., Papachristos, G., Chappin, E., Kwakkel, J., & S., R. (2015). Prospects of modelling societal transitions: Position paper of an emerging community. *Environmental Innovation and Societal Transitions*, *17*, 41–58.

Hosseini, S., Barker, K., & Ramirez-Marquez, J. E. (2016). A review of definitions and measures of system resilience. *Reliability Engineering & System Safety*, *145*, 47–61.

Huang, I. B., Keisler, J., & Linkov, I. (2011). Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. *Science of The Total Environment*, *409*(19), 3578–3594.

Hufschmidt, G. (2011). A comparative analysis of several vulnerability concepts. *Natural Hazards*, *58*(2), 621–643.

Ingrisch, J., & Bahn, M. (2018). Towards a Comparable Quantification of Resilience. *Trends in Ecology & Evolution*, *33*(4), 251–259.

Intergovernmental Panel On Climate Change. (2023). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press.

Islam, T., & Pruyt, E. (2016). Scenario generation using adaptive sampling: The case of resource scarcity. *Environmental Modelling & Software*, *79*, 285–299.

Ivory, V. C., & Stevenson, J. R. (2019). From contesting to conversing about resilience: Kickstarting measurement in complex research environments. *Natural Hazards*, *97*(2), 935–947.

Iwanaga, T., Steinmann, P., Sadoddin, A., Robinson, D., Snow, V., Grimm, V., & Wang, H.-H. (2022). Perspectives on confronting issues of scale in systems modeling. *Socio-Environmental Systems Modelling*, *4*.

Jafino, B. A., & Kwakkel, J. H. (2021). A novel concurrent approach for multiclass scenario discovery using Multivariate Regression Trees: Exploring spatial inequality patterns in the Vietnam Mekong Delta under uncertainty. *Environmental Modelling & Software*, *145*, 105177.

Jafino, B. A., Kwakkel, J. H., & Taebi, B. (2021). Enabling assessment of distributive justice through models for climate change planning: A review of recent advances and a research agenda. *WIREs Climate Change*, *12*(4), e721.

Jain, S. K. (2009). Statistical performance indices for a hydropower reservoir. *Hydrology Research*, *40*(5), 454–464.

Jen, E. (2003). Stable or robust? what's the difference? *Complexity*, *8*(3), 12–18.

Jones, L. (2018). Resilience isn't the same for all: Comparing subjective and objective approaches to resilience measurement. *Wiley Interdisciplinary Reviews: Climate Change*, *10*(1), e552.

Karakoc, D. B., & Konar, M. (2021). A complex network framework for the efficiency and resilience trade-off in global food trade. *Environmental Research Letters*, *16*(10), 105003.

Kasprzyk, J., Nataraj, S., Reed, P., & Lempert, R. (2013). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, *42*, 55–71.

Kelly, J. R., & Harwell, M. A. (1990). Indicators of ecosystem recovery. *Environmental Management*, *14*(5), 527–545.

Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and information systems*, *7*(3), 358–386.

Klein, R. J. T., Nicholls, R. J., & Thomalla, F. (2003). Resilience to natural hazards: How useful is this concept? *Global Environmental Change Part B: Environmental Hazards*, *5*(1), 35–45.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Boston, New York, Houghton Mifflin Company.

Knippenberg, E., Jensen, N., & Constas, M. (2019). Quantifying household resilience with high frequency data: Temporal dynamics and methodological options. *World Development*, *121*, 1–15.

Kollat, J., & Reed, P. (2006). Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Advances in Water Resources*, *29*(6), 792–807.

Kristensen, N. P., Gabric, A., Braddock, R., & Cropp, R. (2003). Is maximizing resilience compatible with established ecological goal functions? *Ecological Modelling*, *169*(1), 61–71.

Kunc, M., & O'Brien, F. A. (2017). Exploring the development of a methodology for scenario use: Combining scenario and resource mapping approaches. *Technological Forecasting and Social Change*, *124*, 150–159.

Kwakkel, J. H., & Cunningham, S. C. (2016). Improving scenario discovery by bagging random boxes. *Technological Forecasting and Social Change*, *111*, 124–134.

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 & Software*, *96*, 239–250.

Kwakkel, J. H. (2019). A generalized many-objective optimization approach for scenario discovery. *Futures & Foresight Science*, *1*(2), e8.

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., & Haasnoot, M. (2019). Supporting DMDU: A Taxonomy of Approaches and Tools. In V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, & S. W. Popper (Eds.), *Decision Making under Deep Uncertainty* (pp. 355–374). Springer International Publishing.

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

Kwakkel, J. H., & Pruyt, E. (2013). Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. *Technological Forecasting and Social Change*, *80*(3), 419–431.

Kwakkel, J., Haasnoot, M., & Walker, W. (2015). Developing dynamic adaptive policy pathways: A computer-assisted approach for developing adaptive strategies for a deeply uncertain world. *Climatic Change*, *132*(3), 373–386.

Kwakkel, J., Haasnoot, M., & Walker, W. (2016). Comparing robust decision-making and dynamic adaptive policy pathways for model-based decision support under deep uncertainty. *Environmental Modelling & Software*, *86*, 168–183.

Kwakkel, J., Walker, W., & Haasnoot, M. (2016). Coping with the wickedness of public policy problems: Approaches for decision making under deep uncertainty. *Journal of Water Resources Planning and Management*, *142*(3).

Lake, P. S. (2000). Disturbance, patchiness, and diversity in streams. *Journal of the North American Benthological Society*, *19*(4), 573–592.

Lamontagne, J. R., Reed, P. M., Link, R., Calvin, K. V., Clarke, L. E., & Edmonds, J. A. (2018). Large Ensemble Analytic Framework for Consequence-Driven Discovery of Climate Change Scenarios. *Earth's Future*, *6*(3), 488–504.

Lempert, R., & Groves, D. (2010). Identifying and evaluating robust adaptive policy responses to climate change for water management agencies in the american west. *Technological Forecasting and Social Change*, *77*(6), 960–974.

Lempert, R. J., Bryant, B. P., & Bankes, S. C. (2008). *Comparing algorithms for scenario discovery* (Report). RAND.

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006). A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science*, *52*(4), 514–528.

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the next one hundred years: New methods for quantitative, long-term policy analysis*. RAND.

Lesnoff, M., Corniaux, C., & Hiernaux, P. (2012). Sensitivity analysis of the recovery dynamics of a cattle population following drought in the Sahel region. *Ecological Modelling*, *232*, 28–39.

Levins, R. (1966). The Strategy of Model Building in Population Biology. *American Scientist*, *54*(4), 421–431.

Liao, K.-H. (2012). A Theory on Urban Resilience to Floods–A Basis for Alternative Planning Practices. *Ecology and Society*, *17*(4), art48.

Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern Recognition*, *38*(11), 1857–1874.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *Annals of Internal Medicine*, *151*(4), W–65.

Lindgren, K., & Nordahl, M. G. (1990). Universal computation in simple one-dimensional cellular automata. *Complex Systems*, *4*(3), 299–318.

Ludwig, D., Jones, D. D., & Holling, C. S. (1978). Qualitative analysis of insect outbreak systems: The spruce budworm and forest. *Journal of Animal Ecology*, *47*(1), 315–332.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, *5.1*, 281–298.

Maier, H., Guillaume, J., van Delden, H., Riddell, G., Haasnoot, M., & Kwakkel, J. (2016). An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? *Environmental Modelling & Software*, *81*, 154–164.

Maier, H., Razavi, S., Kapelan, Z., Matott, L., Kasprzyk, J., & Tolson, B. (2019). Introductory overview: Optimization using evolutionary algorithms and other metaheuristics. *Environmental Modelling & Software*, *114*, 195–213.

Manheim, D. (2023). Building less-flawed metrics: Understanding and creating better measurement and incentive systems. *Patterns*, *4*(10).

Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., & Popper, S. W. (Eds.). (2019). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing.

Margolus, N. (1984). Physics-like models of computation. *Physica D: Nonlinear Phenomena*, *10*(1-2), 81–95.

Marshall, N. A., Fenton, D. M., Marshall, P. A., & Sutton, S. G. (2007). How Resource Dependency Can Influence Social Resilience within a Primary Resource Industry*. *Rural Sociology*, *72*(3), 359–390.

Matrosov, E., Woords, A., & Harou, J. (2013). Robust decision making and info-gap decision theory for water resource system planning. *Journal of Hydrology*, *494*(28 June 2013), 43–58.

Maturana, H. R., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. D. Reidel.

McJeon, H. C., Clarke, L., Kyle, P., Wise, M., Hackbarth, A., Bryant, B. P., & Lempert, R. J. (2011). Technology interactions among low-carbon energy technologies: What can we learn from a large number of scenarios? [Special Issue on The Economics of Technologies to Combat Global Warming]. *Energy Economics*, *33*(4), 619–631.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239–245.

Mcmillan, H., Westerberg, I., & Branger, F. (2017). Five guidelines for selecting hydrological signatures. *Hydrological Processes*, *31*(26), 4757–4761.

McPhail, C., Maier, H. R., Kwakkel, J. H., Giuliani, M., Castelletti, A., & Westra, S. (2018). Robustness Metrics: How Are They Calculated, When Should They Be Used and Why Do They Give Different Results? *Earth's Future*, *6*(2), 169–191.

McPhail, C., Maier, H. R., Westra, S., Kwakkel, J. H., & Linden, L. (2020). Impact of Scenario Selection on Robustness. *Water Resources Research*, *56*(9).

Meadows, D. (2007). A brief and incomplete history of operational gaming in system dynamics. *System Dynamics Review: The Journal of the System Dynamics Society*, *23*(2-3), 199–203.

Meerow, S., & Newell, J. P. (2019). Urban resilience for whom, what, when, where, and why? *Urban Geography*, *40*(3), 309–329.

Mingers, J. (1991). The cognitive theories of Maturana and Varela. *Systems Practice*, *4*, 319–338.

Mitchell, M. (2005, January 24). Computation in cellular automata: A selected review. In *Non-standard computation* (pp. 95–140). Wiley.

Mitchell, S. D. (2009). *Unsimple Truths*. University of Chicago Press.

Moallemi, E. A., de Haan, F., Kwakkel, J., & Aye, L. (2017). Narrative-informed exploratory analysis of energy transition pathways: A case study of india's electricity sector. *Energy Policy*, *110*, 271–287.

Moallemi, E. A., Elsawah, S., & Ryan, M. J. (2018). An agent-monitored framework for the output-oriented design of experiments in exploratory modelling. *Simulation Modelling Practice and Theory*, *89*, 48–63.

Moallemi, E. A., Zare, F., Hebinck, A., Szetey, K., Molina-Perez, E., Zyngier, R. L., Hadjikakou, M., Kwakkel, J., Haasnoot, M., Miller, K. K., Groves, D. G., Leith, P., & Bryan, B. A. (2023). Knowledge co-production for decision-making in human-natural systems under uncertainty. *Global Environmental Change*, *82*, 102727.

Montero, P., & Vilar, J. A. (2014). TSclust: An R package for time series clustering. *2014*, *62*(1), 43.

Morita, K., Tojima, Y., Imai, K., & Ogiro, T. (2002). Universal computing in reversible and number-conserving two-dimensional cellular spaces. In *Collision-based computing* (pp. 161–199). Springer.

Mumby, P. J., Wolff, N. H., Bozec, Y.-M., Chollett, I., & Halloran, P. (2014). Operationalizing the Resilience of Coral Reefs in an Era of Climate Change. *Conservation Letters*, *7*(3), 176–187.

Munn, R. E. (1992). Towards sustainable development. *Atmospheric Environment. Part A. General Topics*, *26*(15), 2725–2731.

Myers-Smith, I. H., Trefry, S. A., & Swarbrick, V. J. (2012). Resilience: Easy to use but hard to define. *Ideas in Ecology and Evolution*, *5*.

Nakićenović, N., Alcamo, J., Grubler, A., Riahi, K., Roehrl, R., Rogner, H., & Victor, N. (Eds.). (2000). *Special report on emissions scenarios: A special report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Nelson, D. R., Adger, W. N., & Brown, K. (2007). Adaptation to Environmental Change: Contributions of a Resilience Framework. *Annual Review of Environment and Resources*, *32*(1), 395–419.

Nilsson, C., & Grelsson, G. (1995). The Fragility of Ecosystems: A Review. *The Journal of Applied Ecology*, *32*(4), 677.

Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F., & Pfefferbaum, R. L. (2008). Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness. *American Journal of Community Psychology*, *41*(1), 127–150.

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., & van Vuuren, D. P. (2014). A new scenario framework for climate change research: The concept of shared socioeconomic pathways. *Climatic change*, *122*(3), 387–400.

Opdyke, A., Lepropre, F., Javernick-Will, A., & Koschmann, M. (2017). Interorganizational resource coordination in post-disaster infrastructure recovery. *Construction Management and Economics*, *35*(8-9), 514–530.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, *263*(5147), 641–646.

Ouyang, M., Liu, C., & Xu, M. (2019). Value of resilience-based solutions on critical infrastructure protection: Comparing with robustness-based solutions. *Reliability Engineering & System Safety*, *190*, 106506.

Packard, N. H., & Wolfram, S. (1985). Two-dimensional cellular automata. *Journal of Statistical Physics*, *38*, 901–946.

Page, S. E. (2016). Many Model Thinking. *2016 Winter Simulation Conference (WSC)*, 1–1.

Paparrizos, J., & Gravano, L. (2015). K-shape: Efficient and accurate clustering of time series. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1855–1870.

Park, J., Seager, T. P., Rao, P. S. C., Convertino, M., & Linkov, I. (2013). Integrating Risk and Resilience Approaches to Catastrophe Management in Engineering Systems. *Risk Analysis*, *33*(3), 356–367.

Parker, A., Srinivasan, S., Lempert, R., & Berry, S. (2015). Evaluating simulation-derived scenarios for effective decision support. *Technological Forecasting and Social Change*, *91*(February), 64–77.

Perez-España, H., & Arreguıin-Sanchez, F. (2001). An inverse relationship between stability and maturity in models of aquatic ecosystems. *Ecological Modelling*, *145*(2), 189–196.

Peters, M. D., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis*, *18*(10), 2119–2126.

Pielke, R., & Ritchie, J. (2021). Distorting the view of our climate future: The misuse and abuse of climate pathways and scenarios. *Energy Research & Social Science*, *72*, 101890.

Pimm, S. L. (1984). The complexity and stability of ecosystems. *Nature*, *307*(5949), 321–326.

Polhill, J. G., Hare, M., Bauermann, T., Anzola, D., Palmer, E., Salt, D., & Antosz, P. (2021). Using agent-based models for prediction in complex and wicked systems. *Journal of Artificial Societies and Social Simulation*, *24*(3).

Pot, W., Scherpenisse, J., & 't Hart, P. (2022). Robust governance for the long term and the heat of the moment: Temporal strategies for coping with dual crises. *Public Administration*.

Poulin, C., & Kane, M. B. (2021). Infrastructure resilience curves: Performance measures and summary metrics. *Reliability Engineering & System Safety*, *216*, 107926.

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

Pruyt, E., & Islam, T. (2015). On generating and exploring the behavior space of complex models. *System Dynamics Review*, *31*(4), 220–249.

Quinlan, A. E., Berbés-Blázquez, M., Haider, L. J., & Peterson, G. D. (2016). Measuring and assessing resilience: Broadening understanding through multiple disciplinary perspectives. *Journal of Applied Ecology*, *53*(3), 677–687.

Quinn, J. D., Reed, P. M., & Keller, K. (2017). Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points. *Environmental Modelling & Software*, *92*, 125–141.

R Core Team. (2018). R: A language for statistical computing.

Rapport, D. J. (1989). What Constitutes Ecosystem Health? *Perspectives in Biology and Medicine*, *33*(1), 120–132.

Renn, O., Laubichler, M., Lucas, K., Kröger, W., Schanze, J., Scholz, R. W., & Schweizer, P.-J. (2022). Systemic Risks from Different Perspectives. *Risk Analysis*, *42*(9), 1902–1920.

Resh, V. H., Brown, A. V., Covich, A. P., Gurtz, M. E., Li, H. W., Minshall, G. W., Reice, S. R., Sheldon, A. L., Wallace, J. B., & Wissmar, R. C. (1988). The Role of Disturbance in Stream Ecology. *Journal of the North American Benthological Society*, *7*(4), 433–455.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., ... Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., ... Foley, J. A. (2009). A safe operating space for humanity. *Nature*, *461*(7263), 472–475.

Rodrigues, P. P., Gama, J., & Pedroso, J. (2008). Hierarchical clustering of time-series data streams. *IEEE transactions on knowledge and data engineering*, *20*(5), 615–627.

Roege, P. E., Collier, Z. A., Mancillas, J., McDonagh, J. A., & Linkov, I. (2014). Metrics for energy resilience. *Energy Policy*, *72*, 249–256.

Rosenblueth, A., & Wiener, N. (1945). The Role of Models in Science. *Philosophy of Science*, *12*(4), 316–321.

Rosenhead, J., Elton, M., & Gupta, S. K. (1972). Robustness and Optimality as Criteria for Strategic Decisions. *Journal of the Operational Research Society*, *23*(4), 413–431.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975–5979.

Roux, D., Kempster, P., Kleynhans, C., Van, V., & Du, P. (1999). PROFILE: Integrating Stressor and Response Monitoring into a Resource-Based Water-Quality Assessment Framework. *Environmental Management*, *23*(1), 15–30.

Rozenberg, J., Guivarch, C., Lempert, R., & Hallegatte, S. (2014). Building ssps for climate policy analysis: A scenario elicitation methodology to map the space of possible future challenges to mitigation and adaptation. *Climatic Change*, *122*(3), 509–522.

Saltelli, A., & Homma, T. (1992). Sensitivity analysis for model output: Performance of black box techniques on three international benchmark exercises. *Computational Statistics & Data Analysis*, *13*(1), 73–94.

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, *1*(2), 143–186.

Schoemaker, P. J. (1993). Multiple scenario development: Its conceptual and behavioral foundation. *Strategic management journal*, *14*(3), 193–213.

Sharma, P., & Chen, Z. (2020). Probabilistic Resilience Measurement for Rural Electric Distribution System Affected by Hurricane Events. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *6*(2), 04020021.

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications* (4th ed.). Springer International Publishing.

Siegenfeld, A. F., & Bar-Yam, Y. (2020). An Introduction to Complex Systems Science and its Applications. *Complexity*, *2020*, 1–16.

Simonovic, S. P., Venema, H. D., & Burn, D. H. (1992). Risk-based parameter selection for short-term reservoir operation. *Journal of Hydrology*, *131*(1), 269–291.

Skegg, D., Gluckman, P., Boulton, G., Hackmann, H., Karim, S. S. A., Piot, P., & Woopen, C. (2021). Future scenarios for the COVID-19 pandemic. *The Lancet*, *397*(10276), 777–778.

Smaldino, P. E. (2017). Models are stupid, and we need more of them. In *Computational Social Psychology* (pp. 311–331).

Smerlak, M., & Vaitla, B. (2017). A non-equilibrium formulation of food security resilience. *Royal Society Open Science*, *4*(1), 160874.

Smith, A. B., & Katz, R. W. (2013). US billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases. *Natural Hazards*, *67*(2), 387–410.

Smith, B. (2004). Oil wealth and regime survival in the developing world, 1960–1999. *American Journal of Political Science*, *48*(2), 232–246.

Spaniol, M. J., & Rowland, N. J. (2019). Defining scenario. *Futures & Foresight Science*, *1*(1), e3.

Stanton, M. C. B., & Roelich, K. (2021). Decision making under deep uncertainties: A review of the applicability of methods in practice. *Technological Forecasting and Social Change*, *171*, 120939.

Steinmann, P. (2018). *Behavior-based scenario discovery* (Thesis). Delft University of Technology.

Steinmann, P., Auping, W. L., & Kwakkel, J. H. (2020). Behavior-based scenario discovery using time series clustering. *Technological Forecasting and Social Change*, *156*, 120052.

Steinmann, P., Wang, J. R., van Voorn, G. A., & Kwakkel, J. H. (2020). Don't try to predict COVID-19. If you must, use Deep Uncertainty methods. *Review of Artificial Societies and Social Simulation*.

Sterman, J. (2002). All models are wrong: Reflections on becoming a systems scientist. *System Dynamics Review*, *18*(4), 501–531.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, *10*(2-3), 291–330.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Irwin, McGraw-Hill.

Stonedahl, F., & Wilensky, U. (2011). Finding Forms of Flocking: Evolutionary Search in ABM Parameter-Spaces. In T. Bosse, A. Geller, & C. M. Jonker (Eds.), *Multi-Agent-Based Simulation XI* (pp. 61–75). Springer Berlin Heidelberg.

Strogatz, S., Friedman, M., Mallinckrodt, A. J., & McKay, S. (1994). Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering. *Computers in Physics*, *8*(5), 532.

Strogatz, S. H. (2018). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. CRC press.

Sun, W., Bocchini, P., & Davison, B. D. (2020). Resilience metrics and measurement methods for transportation infrastructure: The state of the art. *Sustainable and Resilient Infrastructure*, *5*(3), 168–199.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable* (1st ed). Random House.

ten Broeke, G. A., van Voorn, G. A. K., Ligtenberg, A., & Molenaar, J. (2019). Cooperation can improve the resilience of common-pool resource systems against over-harvesting. *Ecological Complexity*, *40*, 100742.

ten Broeke, G., & Tobi, H. (2021). Mapping validity and validation in modelling for interdisciplinary research. *Quality & Quantity*, *55*(5), 1613–1630.

ten Broeke, G., van Voorn, G., & Ligtenberg, A. (2016). Which Sensitivity Analysis Method Should I Use for My Agent-Based Model? *Journal of Artificial Societies and Social Simulation*, *19*(1), 5.

ten Broeke, G., van Voorn, G., Ligtenberg, A., & Molenaar, J. (2021). The Use of Surrogate Models to Analyse Agent-Based Models. *Journal of Artificial Societies and Social Simulation*, *24*(2), 3.

ten Broeke, G. A., van Voorn, G. A. K., Ligtenberg, A., & Molenaar, J. (2017). Resilience through adaptation. *PLOS ONE*, *12*(2), e0171833.

Thomas, R. L., & Uminsky, D. (2022). Reliance on metrics is a fundamental challenge for AI. *Patterns*, *3*(5), 100476.

Thompson, E. L., & Smith, L. A. (2019). Escape from model-land. *Economics*, *13*(1), 20190040.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*(7), 467–473.

Trindade, B. C., Gold, D. F., Reed, P. M., Zeff, H. B., & Characklis, G. W. (2020). Water pathways: An open source stochastic simulation system for integrated water supply portfolio management and infrastructure investment planning. *Environmental Modelling & Software*, *132*, 104772.

Ulanowicz, R. E., Goerner, S. J., Lietaer, B., & Gomez, R. (2009). Quantifying sustainability: Resilience, efficiency and the return of information theory. *Ecological Complexity*, *6*(1), 27–36.

van Voorn, G. A. K., Verburg, R. W., Kunseler, E. M., Vader, J., & Janssen, P. H. M. (2016). A checklist for model credibility, salience, and legitimacy to improve information transfer in environmental policy assessments. *Environmental Modelling & Software*, *83*, 224–236.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., & Rose, S. K. (2011). The representative concentration pathways: An overview. *Climatic Change*, *109*(1), 5.

Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, *5*(4), 187–196.

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition*, *34*(1), 72–87.

Verstegen, J. A., Jonker, J. G. G., Karssenberg, D., van der Hilst, F., Schmitz, O., de Jong, S. M., & Faaij, A. P. C. (2017). How a pareto frontier complements scenario projections in land use change impact assessment. *Environmental modelling & software*, *97*, 287–302.

Verstegen, J. A., van der Hilst, F., & Karssenberg, D. (2017). Locating the position of a scenario projection in solution space. *The 20th AGILE Conference on Geographic Information Science*.

Villalobos, M., & Razeto-Barry, P. (2019). Are living beings extended autopoietic systems? An embodied reply. *Adaptive Behavior*.

Von Bertalanffy, L. (1968). *General system theory*. Braziller.

Voros, J. (2017). Big History and Anticipation. In R. Poli (Ed.), *Handbook of Anticipation* (pp. 1–40). Springer International Publishing.

Walker, B., Holling, C. S., Carpenter, S. R., & Kinzig, A. P. (2004). Resilience, Adaptability and Transformability in Social-ecological Systems. *Ecology and Society*, *9*(2), art5.

Walker, W. E. (2000). Policy analysis: A systematic approach to supporting policymaking in the public sector. *Journal of Multi-Criteria Decision Analysis*, *9*(1-3), 11–27.

Walker, W. E., Lempert, R. J., & Kwakkel, J. H. (2013). Deep Uncertainty. In S. I. Gass & M. C. Fu (Eds.), *Encyclopedia of Operations Research and Management Science* (pp. 395–402). Springer US.

Walker, W. E., Marchau, V. A. W. J., & Swanson, D. (2010). Addressing deep uncertainty using adaptive policies: Introduction to section 2. *Technological Forecasting and Social Change*, *77*(6), 917–923.

Walker, W., Harremoës, P., Rotmans, J., van der Sluijs, J., van Asselt, M., Janssen, P., & Krayer von Krauss, M. (2003). Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment*, *4*(1), 5–17.

Watson, A., & Kasprzyk, J. (2017). Incorporating deeply uncertain factors into the many objective search process. *Environmental Modelling & Software*, *89*, 159–171.

Weinkle, J., Landsea, C., Collins, D., Musulin, R., Crompton, R. P., Klotzbach, P. J., & Pielke Jr, R. (2018). Normalized hurricane damage in the continental United States 1900–2017. *Nature Sustainability*, *1*(12), 808–813.

Wilensky, U. (1999). NetLogo.

Willis, G., Cave, S., & Kunc, M. (2018). Strategic workforce planning in healthcare: A multi-methodology approach. *European Journal of Operational Research*, *267*(1), 250–263.

Winsberg, E. (2010). Science in the Age of Computer Simulation. In *Science in the Age of Computer Simulation*. University of Chicago Press.

Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media.

Wolters, H., G.J. van den Born, E. Dammers, & S. Reinhard. (2018). *Deltascenario's voor de 21e eeuw, actualisering 2017* (tech. rep.). Deltares.

Wright, G., Bradfield, R., & Cairns, G. (2013). Does the intuitive logics method – and its recent enhancements – produce "effective" scenarios? *Technological Forecasting and Social Change*, *80*(4), 631–642.

Zatarain Salazar, J., Castelletti, A., & Giuliani, M. (2022). Multi-Objective Robust Planning Tools. In *Oxford Research Encyclopedia of Environmental Science*. Oxford University Press.

Zelnik, Y. R., Arnoldi, J.-F., & Loreau, M. (2018). The Impact of Spatial and Temporal Dimensions of Disturbances on Ecosystem Stability. *Frontiers in Ecology and Evolution*, *6*, 224.

Zhou, H., Wang, J., Wan, J., & Jia, H. (2010). Resilience to natural hazards: A geographic perspective. *Natural Hazards*, *53*(1), 21–41.

Appendix A

Scoping Review: Individual Sources of Evidence

Nr.	Author(s)	Publication year	Title	Journal
1	Onta, PR; Dasgupta, A; Harboe, R	1991	Multistep planning model for conjunctive use of surface-and ground-water resources	JOURNAL OF WATER RESOURCES PLANNING AND MANAGEMENT
2	Harada, Y; Sakuramoto, K; Tanaka, S	1992	On the stability of the stock-harvesting system controlled by a feedback management procedure	RESEARCHES ON POPULATION ECOLOGY
3	Mujumdar, PP; Vedula, S	1992	Performance evaluation of an irrigation system under some optimal operating policies	HYDROLOGICAL SCIENCES JOURNAL
4	Ives, AR	1995	Measuring resilience in stochastic systems	ECOLOGICAL MONOGRAPHS
5	Srinivasan, K; Philipose, M	1996	Evaluation and Selection of Hedging Policies Using Stochastic Reservoir Simulation	WATER RESOURCES MANAGEMENT
6	Loucks, DP	1997	Quantifying trends in system sustainability	HYDROLOGICAL SCIENCES JOURNAL
7	Xu, ZX; Jinno, K; Kawamura, A; Takesaki, S; Ito, K	1998	Performance Risk Analysis for Fukuoka Water Supply System	WATER RESOURCES MANAGEMENT
8	Perrings, C	1998	Resilience in the dynamics of economy-environment systems	ENVIRONMENTAL & RESOURCE ECONOMICS
9	Arreguin-Sanchez, F; Manickchand-Heileman, S	1998	The trophic role of lutjanid fish and impacts of their fisheries in two ecosystems in the Gulf of Mexico	JOURNAL OF FISH BIOLOGY
10	Batabyal, AA	1999	Species substitutability, resilience, and the optimal management of ecological-economic systems	MATHEMATICAL AND COMPUTER MODELLING
11	de Azevedo, LGT; Gates, TK; Fontane,	2000	Integration of water quantity and quality in	JOURNAL OF WATER RESOURCES

	DG; Labadie, JW; Porto, RL		strategic river basin planning	PLANNING AND MANAGEMENT
12	Perrings, C; Stern, DI	2000	Modelling loss of resilience in agroecosystems: Rangelands in Botswana	ENVIRONMENTAL & RESOURCE ECONOMICS
13	Ares, J; Bertiller, M; del Valle, H	2001	Functional and structural landscape indicators of intensification, resilience and resistance in agroecosystems in southern Argentina based on remotely sensed data	LANDSCAPE ECOLOGY
14	Maier, HR; Lence, BJ; Tolson, BA; Foschi, R	2001	First-order reliability method for estimating reliability, vulnerability, and resilience	WATER RESOURCES RESEARCH
15	Perez-Espana, H; Arreguin-Sanchez, F	2001	An inverse relationship between stability and maturity in models of aquatic ecosystems	ECOLOGICAL MODELLING
16	Merabtene, T; Kawamura, A; Jinno, K; Olsson, J	2002	Risk assessment for optimal drought management of an integrated water resources system using a genetic algorithm	HYDROLOGICAL PROCESSES
17	Kristensen, NP; Gabric, A; Braddock, R; Cropp, R	2003	Is maximizing resilience compatible with established ecological goal functions?	ECOLOGICAL MODELLING
18	Chang, SE; Shinozuka, M	2004	Measuring improvements in the disaster resilience of communities	EARTHQUAKE SPECTRA
19	El-Baroudy, I; Simonovic, SP	2004	Fuzzy criteria for the evaluation of water resource systems performance	WATER RESOURCES RESEARCH
20A, 20B	Prasad, TD; Park, NS	2004	Multiobjective genetic algorithms for design of water distribution networks	JOURNAL OF WATER RESOURCES

				PLANNING AND MANAGEMENT
21	Peterson, GD	2006	Estimating resilience across landscapes	CONSERVATION ECOLOGY
22	Mondal, MS; Wasimi, SA	2007	Evaluation of risk-related performance in water management for the Ganges Delta of Bangladesh	JOURNAL OF WATER RESOURCES PLANNING AND MANAGEMENT
23A, 23B	Jain, SK	2009	Statistical performance indices for a hydropower reservoir	HYDROLOGY RESEARCH
24	Petchey, OL; Gaston, KJ	2009	Effects on ecosystem resilience of biodiversity, extinctions, and the structure of regional species pools	THEORETICAL ECOLOGY
25	Reed, DA; Kapur, KC; Christie, RD	2009	Methodology for Assessing the Resilience of Networked Infrastructure	IEEE SYSTEMS JOURNAL
26A, 26B	Wang, DW; Ip, WH	2009	Evaluation and Analysis of Logistic Network Resilience With Application to Aircraft Servicing	IEEE SYSTEMS JOURNAL
27	Whitson, JC; Ramirez-Marquez, JE	2009	Resiliency as a component importance measure in network reliability	RELIABILITY ENGINEERING & SYSTEM SAFETY
28	Dolling, OR; Varas, EA	2010	Decision support model for operation of multi-purpose water resources systems	JOURNAL OF HYDRAULIC RESEARCH
29	Cox, A; Prager, F; Rose, A	2011	Transportation security and the role of resilience: A foundation for operational metrics	TRANSPORT POLICY
30A, 30B	Lesnoff, M; Corniaux, C; Hiernaux, P	2012	Sensitivity analysis of the recovery dynamics of a cattle population following drought in the Sahel region	ECOLOGICAL MODELLING
31A, 31B	Roe, E; Schulman, PR	2012	Toward a Comparative Framework for Measuring Resilience	JOURNAL OF COMPARATIVE POLICY ANALYSIS

			in Critical Infrastructure Systems	
32	Mumby, PJ; Wolff, NH; Bozec, YM; Chollett, I; Halloran, P	2013	Operationalizing the Resilience of Coral Reefs in an Era of Climate Change	CONSERVATION LETTERS
33	Duveneck, MJ; Scheller, RM	2016	Measuring and managing resistance and resilience under climate change in northern Great Lake forests (USA)	LANDSCAPE ECOLOGY
34	Bakhshipour, AE; Dittmer, U; Haghighi, A; Nowak, W	2019	Hybrid green-blue-gray decentralized urban drainage systems design, a simulation-optimization framework	JOURNAL OF ENVIRONMENTAL MANAGEMENT
35	Hassan, D; Burian, SJ; Bano, R; Ahmed, W; Arfan, M; Rais, MN; Rafique, A; Ansari, K	2019	An Assessment of the Pakistan Water Apportionment Accord of 1991	RESOURCES
36	Min, O; Chuang, L; Min, X	2019	Value of resilience-based solutions on critical infrastructure protection: Comparing with robustness-based solutions	RELIABILITY ENGINEERING & SYSTEM SAFETY
37	Leandro, J; Chen, KF; Wood, RR; Ludwig, R	2020	A scalable flood-resilience-index for measuring climate change adaptation: Munich city	WATER RESEARCH
38	Salomon, J; Broggi, M; Kruse, S; Weber, S; Beer, M	2020	Resilience Decision-Making for Complex Systems	JOURNAL OF RISK AND UNCERTAINTY IN ENGINEERING SYSTEMS PART B-MECHANICAL ENGINEERING
39	Sharma, P; Chen, ZQ	2020	Probabilistic Resilience Measurement for Rural Electric Distribution	JOURNAL OF RISK AND UNCERTAINTY IN ENGINEERING SYSTEMS PART A-

			System Affected by Hurricane Events	CIVIL ENGINEERING
40	Verol, AP; Lourenco, IB; Fraga, JPR; Battemarco, BP; Merlo, ML; de Magalhaes, PC; Miguez, MG	2020	River Restoration Integrated with Sustainable Urban Water Management for Resilient Cities	SUSTAINABILITY
41	Wang, Y; Taylor, JE; Garvin, MJ	2020	Measuring Resilience of Human-Spatial Systems to Disasters: Framework Combining Spatial-Network Analysis and Fisher Information	JOURNAL OF MANAGEMENT IN ENGINEERING

Appendix B

Scoping Review: Results of Individual Sources of Evidence

		Data extracted		Data inferred		
Nr.	Source	Resilience metric	Elements	System type	Basin(s) of attraction	Disturbance
1	Onta, PR; Dasgupta, A; Harboe, R (1991)	"Resiliency (RES) is considered the maximum number of consecutive periods of shortages that occur prior to recovery to an acceptable state within the planning period."	Not applicable	Socio-technical	single	multiple
2	Harada, Y; Sakuramoto, K; Tanaka, S (1992)	$\frac{-1}{\ln \lvert\lambda_D\rvert}$	λ_D: dominant eigenvalue	Socio-ecological	single	continuous
3	Mujumdar, PP; Vedula, S (1992)	$\gamma = P(X_{t+1} \in V_{t+1} \mid X_t \in U_t)$	V_t: set of satisfactory outputs in period t X_t: output in period t U_t: set of unsatisfactory outputs in period t	Socio-technical	single	multiple
4	Ives, AR (1995)	$1/(1 - \lvert\lvert\lambda_{\pm}\rvert\rvert^2)$	λ: eigenvalue	Socio-ecological	single	continuous
5	Srinivasan, K; Philipose, MC (1996)	"Resilience is computed as the ratio of the number of times the system moved from failure to success, to the total number of periods the system was in a failure state."	Not applicable	Socio-technical	single	multiple
6	Loucks, DP (1997)	$Resilience\ of\ C = \frac{number\ of\ times\ satisfactory\ C_t\ follows\ unsatisfactory\ C_t}{number\ of\ unsatisfactory\ C_t\ values}$	C:selected criterion t:simulated time steps	Socio-technical	single	multiple
7	Xu, ZX; Jinno, K; Kawamura, A; Takesaki, S; Ito, K (1998)	$\beta = \begin{cases} \frac{1}{(1/NF)\sum_{i=1}^{NF} FP_i}, & NF \neq 0 \\ 1, & NF = 0 \end{cases}$	N: length of planning period F: set of all unsatisfactory outputs FP_i: total days of ith consecutive period of water deficit NF: not explained	Socio-technical	single	multiple
8	Perrings, C (1998)	$\lim_{t\to\infty} p^t = p(0)P^\infty$	p: vector of state probabilities P^∞: matrix of limiting state transition probabilities	Socio-ecological	multiple	unclear
9	Arreguin-Sanchez, F; Manickchand-Heileman, S (1998)	$\frac{\frac{B_{max} - B_{min}}{B_{ini}}}{R}$	$B_{max} - B_{min}$: maximum proportional change in biomass B_{ini}: baseline biomass value R: time lapsed from the beginning to the end of the impact	Socio-ecological	single	continuous with abrupt end
10	Batabyal, AA (1999)	$\lim_{t\to\infty} \text{Prob(ecosystem functional at time t)} = \left\{\frac{\alpha_1(\alpha_2+\beta_2)+\alpha_2\beta_1}{(\alpha_1+\beta_1)(\alpha_2+\beta_2)}\right\}\left\{\frac{\alpha_3(\alpha_4+\beta_4)+\alpha_4\beta_3}{(\alpha_3+\beta_3)(\alpha_4+\beta_4)}\right\}$	α_i: mean of life distribution function of species i β_i: mean of death distribution function of species i	Socio-ecological	multiple	continuous
11	de Azevedo, LGT; Gates, TK; Fontane, DG; Labadie, JW; Porto, RL (2000)	$P_{res} = \begin{cases} \frac{1}{N_f N_{cf}}\ for\ N_f \geq 1 \\ 1\ for\ N_f = 0 \end{cases}$	N_f: total number of occurrences of system failures over period of observation N_{cf}: maximum number of consecutive periods of failure over period of observation	Socio-technical	single	multiple
12	Perrings, C; Stern, DI (2000)	"This use of the threshold level of $M_t/\hat{K}_{t+1}$ below which M is unchanged expresses the idea of loss of resilience. The system is less resilient the further K is from M."	K_t : current carrying capacity M_t : long-run equilibrium carrying capacity	Socio-ecological	single	continuous
13	Ares, J; Bertiller, M; del Valle, H (2001)	$r = -\sum_{i=110-260}(\text{NDVI}_{j,i} - \text{NDVI}_{est,i})$	$i = 110 - 260$: drying part of annual cycle NDVI: Normalized Difference Vegetation Index j: sampling area	Socio-ecological	single	continuous
14	Maier, HR; Lence, BJ; Tolson, BA; Foschi, RO (2001)	$\gamma = \frac{\Phi(-\beta_1, -\beta_2; \rho_{12})}{\Phi(-\beta_1)}$	β: reliability index $\Phi(-x)$: failure probability of x ρ: correlation coefficient	Socio-technical	single	multiple
15	Perez-Espana, H; Arreguin-Sanchez, F (2001)	"[...] resilience was estimated as the inverse tangent of the ratio of resistance versus the *recovery time* or the time biomass requires to reach a level close to the original value, where 'close' represents a value within ±10% of the original biomass."	Not applicable	Socio-ecological	single	continuous with abrupt end

16	Merabtene, T; Kawamura, A; Jinno, K; Olsson, J (2002)	$\text{Res} = \begin{cases} \frac{1}{\frac{1}{f}\sum_{i=1}^{f} \text{df}_i} & \text{if } f \neq 0 \\ 1 & \text{if } f = 0 \end{cases}$	f: total number of failures df_i: number of days of performance deficit during ith failure	Socio-technical	single	multiple
17	Kristensen, NP; Gabric, A; Braddock, R; Cropp, R (2003)	"Resilience is defined the negative real part of the eigenvalue closest to zero."	Not applicable	Socio-ecological	single	sudden
18	Chang, SE; Shinozuka, M (2004)	$Pr(A\|i) = Pr(r_0 < r^* \text{ and } t_1 < t^*)$	A: predefined performance standards i: magnitude of seismic event r_0: loss of performance r^*: rapidity performance standard t_1: time to full recovery t^*: rapidity performance standard	Socio-technical	single	sudden
19	El-Baroudy, I; Simonovic, SP (2004)	$RS_f = \left[\frac{\int_{t1}^{t2} t\tilde{T}(t)dt}{\int_{t1}^{t2} \tilde{T}(t)dt}\right]^{-1}$	$t1$: lower bound of the support of the system recovery time $t2$: upper bound of the support of the system recovery time $\tilde{T}(t)$: system fuzzy maximum recovery time	Socio-technical	single	multiple
20A	Prasad, TD; Park, NS (2004)	$I_r = 1 - \left(\frac{P_{int}}{P_{int}^{max}}\right)$	P_{int}: power dissipated in network P_{int}^{max}: maximum power that would be dissipated internally per design	Socio-technical	single	continuous
20B	Prasad, TD; Park, NS (2004)	$I_n = \frac{X}{X_{max}}$	X: weighted surplus power X_{max}: maximum surplus power	Socio-technical	single	continuous
21	Peterson, GD (2006)	"The behavior of a discrete state can be assessed in terms of the probabilities of leaving that state and remaining in that state. The probability that a state will persist is a measure of its resilience. If the probability that a state will persist is less than the probability that it will not, then it is vulnerable to change. By mapping these probabilities across space, the areas of vulnerability and resilience in a landscape can be estimated (Fig. 1)."	Not applicable	Socio-ecological	multiple	multiple
22	Mondal, MS; Wasimi, SA (2007)	$\gamma_{r,s}(n) = 1 - \text{Prob}\left[\left(X_{r,s+1} \cap X_{r,s+2} \cap \ldots \ldots \cap X_{r,s+n}\right) \in F \| X_{r,s} \in F\right]$	r: year s: season X: system output state n: steps ahead F: failure state	Socio-technical	single	multiple
23A	Jain, SK (2009)	$\gamma_{\text{mean}} = \frac{1}{M}\left[\sum_{j=1}^{M} d_j\right]^{-1}$	M: total number of failure events d_j: duration of failure event j	Socio-technical	single	multiple
23B	Jain, SK (2009)	$\gamma_{\max} = \left[\max\{d_j\}\right]^{-1}$	d_j: duration of failure event j	Socio-technical	single	multiple
24	Petchey, OL; Gaston, KJ (2009)	"Our measure of resilience, R_X, is therefore change in functional diversity caused by the loss (or gain) of a species, subtracted from 1."	Not applicable	Socio-ecological	multiple	sudden
25	Reed, DA; Kapur, KC; Christie, RD (2009)	$R = \frac{\int_{t_1}^{t_2} Q(t)dt}{(t_2 - t_1)}$	$Q(t)$: quality curve t_1, t_2: endpoints of the time interval under consideration	Socio-technical	single	sudden
26A	Wang, DW; Ip, WH (2009)	$r_i = \frac{\sum_{j=1}^{g} p_j q_j \min\{d_i, s_j, c_j\}}{d_i}$	g: suppliers p_j: supply reliability of node j q_j: reliability of edge j d_i: demand of node i s_j: available supply of node j c_j: flow capacity of edge j	Socio-technical	single	sudden
26B	Wang, DW; Ip, WH (2009)	$R = \sum_{i=1}^{n_1} w_i r_i$	n_1: total number of nodes w_i: weight of demand node i r_i: node resilience	Socio-technical	single	sudden

27	Whitson, JC; Ramirez-Marquez, JE (2009)	$R_{(\alpha,\beta)} = P(\varphi(x) \geq d \mid \alpha,\beta)\forall\beta$	α: external failures β: specific α-failure case scenario $\varphi(x)$: network capacity	Socio-technical	multiple	sudden
28	Dolling, OR; Varas, EA (2010)	"[...] function includes the complement of resilience which is the probability that the system does not recover from a state of failure. It is calculated as the probability that the system is in state of failure in the following period given that it is currently in a state of failure."	Not applicable	Socio-technical	single	multiple
29	Cox, A; Prager, F; Rose, A (2011)	$DSER = \frac{\%\Delta DY^m - \%\Delta DY}{\%\Delta DY^m}$	$\%\Delta DY^m$: maximum percent change in direct output $\%\Delta DY$: estimated percent change in direct output	Socio-technical	single	sudden
30A	Lesnoff, M; Corniaux, C; Hiernaux, P (2012)	"The first output considered was the recovery time T (in year)"	Not applicable	Socio-ecological	single	sudden
30B	Lesnoff, M; Corniaux, C; Hiernaux, P (2012)	$m = \left(\frac{n}{n(0)}\right)^{1/T}$	m: average annual empirical population multiplication rate n: population size T: recovery time	Socio-ecological	single	sudden
31A	Roe, E; Schulman, PR (2012)	"[...] we developed a graphic display of what we term an "edge resilience trajectory" (ERT). [...] One measure of the ERT is to track a moving range of edge R^2s across the baseline period."	Not applicable	Socio-technical	multiple	multiple
31B	Roe, E; Schulman, PR (2012)	"A second approach to an ERT is to generate a series of R^2s by adding one day at a time to an initial subperiod, such that the final R^2 becomes the baseline R^2 for the entire period."	Not applicable	Socio-technical	multiple	multiple
32	Mumby, PJ; Wolff, NH; Bozec, YM; Chollett, I; Halloran, P (2013)	"Resilience was calculated as the probability that a reef remained above the unstable equilibrium after a prescribed period of time during which external disturbance could occur."	Not applicable	Socio-ecological	multiple	continuous
33	Duveneck, MJ; Scheller, RM (2016)	$r_{jk} = \frac{\sqrt{2} - d_{jk}}{\sqrt{2}}$	d_{jk}: minimum multi-dimensional Euclidean distance between time (j) and (k)	Socio-ecological	single	sudden
34	Bakhshipour, AE; Dittmer, U; Haghighi, A; Nowak, W (2019)	$HPI = 100 \times \left(1 - \frac{V_{flooding}}{V_{runoff}}\right)$	$V_{flooding}$: total water that overflows the nodes V_{runoff}: total runoff volume	Socio-technical	single	multiple
35	Hassan, D; Burian, SJ; Bano, R; Ahmed, W; Arfan, M; Rais, MN; Rafique, A; Ansari, K (2019)	$\text{Resilience} = \text{P}\{\text{S}(\text{t}+1)\epsilon\text{NF} \mid \text{S}(\text{t})\epsilon\text{F}\}$	S(t): system state variable under consideration NF: not explained F: not explained	Socio-technical	single	multiple
36	Min, O; Chuang, L; Min, X (2019)	$R = \sum_{i=1}^{m} w_i \times P_R(t_{ci})$	w_i: weight coefficient $P_R(t_{ci})$: real functionality level at time t_{ci}	Socio-technical	single	sudden
37	Leandro, J; Chen, KF; Wood, RR; Ludwig, R (2020)	$FRI_h(t) = \begin{cases} \frac{\sum WF_y \cdot I_y(t)}{\sum WF_y}, t \in [t_s, t_e] \\ FRI_h(t-1) \cdot \left[\prod (I_x)^{WF_x}\right]^{\frac{0,001}{\sum WF_x}}, t > t_e \end{cases}$	WF_y: event phase weighting factor I_y: event phase indicators $[t_s, t_e]$: time interval of event WF_x: recovery phase weighting factor I_y: recovery phase indicators	Socio-technical	single	sudden
38	Salomon, J; Broggi, M; Kruse, S; Weber, S; Beer, M (2020)	$Res = E\left[\frac{\int_0^T Q(t)dt}{\int_0^T \mathcal{T}Q(t)dt}\right]$	$Q(t)$: system performance $\mathcal{T}Q(t)$: target system performance	Socio-technical	single	sudden
39	Sharma, P; Chen, ZQ (2020)	$R_{sys} = \frac{\int_{t_{oe}}^{t_{re}} f_{rec}(t)\text{d}t + (T_C - T_{\text{sys}})Q_{100}}{T_C \times Q_{100}}$	$f_{rec}(t)$: system recovery function T_C: control period T_{sys}: time required for the system to recover after the strike Q_{100}: performance measurement when the system is fully functional	Socio-technical	single	sudden

			t_{oe}: time of the event t_{re}: time for the entire system to fully recover			
40	Verol, AP; Lourenco, IB; Fraga, JPR; Battemarco, BP; Merlo, ML; de Magalhaes, PC; Miguez, MG (2020)	$mFResI = 1 - \frac{\left(FRI_{Project}^{Future} - FRI_{Project}^{Present}\right)}{FRI_{Doing\,nothing}^{Future}}$	$FRI_{Project}^{Future}$: Flood Risk Index considering the project in a future condition $FRI_{Project}^{Present}$: Flood Risk Index considering the project in the present condition $FRI_{Doing\,nothing}^{Future}$: Flood Risk Index considering 'doing nothing' in a future condition	Socio-technical	single	sudden
41	Wang, Y; Taylor, JE; Garvin, MJ (2020)	$FI \approx 4\sum_{i=1}^{n}\left[q_i - q_{(i+1)}\right]^2$	n: number of states q: root of the probability of an observed network metric	Socio-technical	multiple	sudden

Appendix C

Acknowledgements

I am immensely grateful to my promotor, **Jaap Molenaar** for his guidance and wisdom over the past five years. Though my work was always much more "applied" than "mathematics", I like to think we found a shared language in modelling. My gratitude also goes to my supervisor, **George van Voorn**, who bestowed infinite trust and patience upon me. He was always willing to discuss my newest idea, proofread material on short notice, and share encouraging words when they were necessary. Both gave me every opportunity and support to research what I was interested in, which is the greatest gift a young scholar can receive - second only to the number of conferences, workshops and seminars they encouraged me to attend so I could learn, meet, and grow as a scientist.

Any words I could write here would be a hollow shell of my true gratitude towards **Hilde Tobi**. There is a Yiddish word, *mentsh*, which has been loaned into English and applies (unlike the Dutch *mens*) exclusively to people of the utmost integrity and honor. When I grow up, I hope to be her calibre of *mentsh*.

James Adams, **Mikhail Sirenko**, **Jillian Student**, **Jason R. Wang** and I went through the highs and lows of early career research together. I am grateful for their enthusiastic memeing, scheming, celebrating, and commiserating, as was appropriate.

I am thankful towards my friends at the Group for Mathematical and Statistical Methods - **Alessio Albanese**, **Bader Arouisse**, **Guus ten Broeke**, **Daniela Bustos Korts**, **Henry Ehlers**, **Vincent Garin**, **Geerten Hengeveld**, **Dominique Joubert**, **Bas Jacobs**, **Emma Keijzer**, **Antoine Languillaume**, **Marco van Lenthe**, **Wenhao Li**, **Kevin Mildau**, **Emilie Millet**, **Jip Ramakers**, **Bart Jan van Rossum**, **Laura van Schijndel**, **Peter Tamas**, **Yutaka Tsutsumi**, **Hilde Vaessen**, and **Maikel Verouden** - for the shared memories and laughs. Special mentions of gratitude are reserved for the respective head and heart of our chair group, **Peter van Heijster** and **Dinie Verbeek**, for their kindness and support over the past years.

I would like to thank all my co-authors, many of whom became friends, for their bright ideas and hard work. **Ioannis N. Athanasiadis**, **Willem L. Auping**, **Arta Cika**, **Elissa Cohen**, **Hedwig Van Delden**, **Sondoss Elsawah**, **Carlo Giupponi**, **William E. Grant**, **Volker Grimm**, **Takuya Iwanaga**, **Wander Jager**, **Anthony J. Jakeman**, **Mark R. Kramer**, **Germán Kruszewski**, **Jan H. Kwakkel**, **Arend Ligtenberg**, **John C. Little**, **Ahmadreza Marandi**, **Birgit Müller**, **Elisa Perrone**, **Derek Robinson**, **Sabin Roman**, **Amir Sadoddin**, **Luther Seet**, **Val Snow**, **Jillian Student**, **Zhanli Sun**, **Judith Verstegen**, **Hsiao-Hsuan "Rose" Wang**, **Els Weinans**, **Wenqian Yin**, and **Fateme Zare** all made me a better scientist and writer, and I am proud to be listed next to them on our various publications.

My mother and father, **Katalin** and **Toni**, and my sisters, **Cristina** and **Olivia**, gave me all the support I could have wished for as I moved abroad to pursue my interests, and came home far too rarely with incomprehensible tales of my research. I am grateful for their love and patience.

Finally, my ultimate gratitude goes towards my amazing wife, **Anne**, our son, **Jasper**, and our dog, **Tycho**, who have made the Netherlands my home, and are the greatest joys in my life.

Appendix D

Author Biography

Patrick Steinmann was born in 1991 in Basel, Switzerland, and split his formative years between Switzerland and the United States of America. After his compulsory military service in the Swiss Armed Forces, he studied Mechanical Engineering at the Bern University of Applied Sciences. Driven by a desire to not spend his life at a computer, he spent a year in the Balkans on a military peacekeeping mission. It was there that he became interested in complex societal problems, and how to solve them.

To learn more about such problems and their solutions, he moved to the Netherlands to study Engineering & Policy Analysis at the Technical University of Delft. At the TU, he made his first forays into research, motivating him to apply for a PhD at Wageningen University & Research. During his five years in Wageningen, he conducted the research which is presented in this thesis, and contributed to a number of other papers.

Patrick now works at the Netherlands Organisation for Applied Scientific Research (TNO) in the Defence, Safety and Security unit. At TNO, he sits at a computer, applying model-based decision support methods in various projects including [redacted] for [redacted], automated inference of [redacted], and an analysis of [redacted] in the Netherlands.

Financial support from the Chair Group for Mathematical and Statistical Methods for printing this thesis, and from the Graduate School for Production Ecology & Resource Conservation for the cover design is gratefully acknowledged.

Cover design by Noor Kissels.
Thesis style design by Martijn Wieling, adapted by Fabian Dablander.