Delft University of Technology

Simulation-based generation and analysis of multidimensional future scenarios with time series clustering

Steinmann, Patrick; van der Zwet, Koen; Keijser, Bas

DOI
10.1002/ffo2.194

Publication date
2024

Document Version
Final published version

Published in
Futures and Foresight Science

Citation (APA)
Steinmann, P., van der Zwet, K., & Keijser, B. (2024). Simulation-based generation and analysis of multidimensional future scenarios with time series clustering. *Futures and Foresight Science*, *6*(4), Article e194. https://doi.org/10.1002/ffo2.194

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Received: 15 February 2024 | Revised: 11 July 2024 | Accepted: 22 July 2024

DOI: 10.1002/ffo2.194

ORIGINAL ARTICLE

WILEY

Simulation-based generation and analysis of multidimensional future scenarios with time series clustering

Patrick Steinmann[1,2] | Koen van der Zwet[2] | Bas Keijser[2]

[1]Delft University of Technology, Delft, The Netherlands

[2]TNO Defence, Safety & Security, The Hague, The Netherlands

Correspondence
Patrick Steinmann, Delft University of Technology, Delft, The Netherlands.
Email: p.steinmann@tudelft.nl

Funding information
Korps Nationale Politie

Abstract

Scenarios are commonly used for decision support and future exploration of complex systems. Using simulation models to generate these scenarios, called scenario discovery, has received increased attention in the literature as a principled method of capturing the uncertainty, complexity, and dynamics inherent in such problems. However, current methods of incorporating dynamics into scenario discovery are limited to a single outcome of interest. Furthermore, there is little work on the post-generation evaluation of the generated scenarios. In this work, we extend scenario discovery to multiple dynamic outcomes of interest, and present a number of visual and statistical approaches for evaluating the resulting scenario sets. These innovations make model-based scenario generation more widely applicable in decision support for complex societal problems, and open the door to multimethod scenario generation combining model-based and model-free methods such as Intuitive Logics or futures cones.

KEYWORDS

modeling and simulation, scenario, scenario discovery, system dynamics, time series clustering

1 | INTRODUCTION

Many modern societal decision problems are plagued by the presence of uncertainty, complexity, and multiple involved actors (Gotts et al., 2019; Vermeulen et al., 2013). Scenario-based planning has emerged as a popular solution to these challenges (Bradfield et al., 2005; Godet, 2000; Schoemaker, 1993). Scenarios describe the future as a systematized set of plausible narrative descriptions with underlying drivers. This enables improved understanding of key uncertainties, exploration of policy alternatives, and clarification of stakeholder objectives. Scenario-based decision support is considered especially effective for long-term decision contexts (Pot et al., 2023).

A key challenge in scenario-based decision support is how to create a set of scenarios which comprehensively summarizes the decision-relevant possible future developments of the studied problem, and whose constituent scenarios are both individually plausible and distinct from one another (Dhami et al., 2022; Lord et al., 2016). Under the banner of *scenario discovery*, a growing body of literature seeks to investigate how (simulation) models may be used to generate such scenarios, as this allows an explicit coupling of driving factors (i.e., model inputs) to resulting futures (i.e., model outputs) (Bryant & Lempert, 2010; Kwakkel & Jaxa-Rozen, 2016). A topic of special interest in this regard is how temporal dynamics may be included in such a scenario generation process, as temporal processes such as delays, feedbacks, and accumulation are notable challenges for human cognition (Sterman, 1994) and could be tackled with computational methods (Lustick & Tetlock, 2021). Behavior-based or dynamic scenario discovery based on time series clustering

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 The Author(s). *Futures & Foresight Science* published by John Wiley & Sons Ltd.

(Kwakkel et al., 2013; Steinmann et al., 2020) has been proposed as a possible solution for addressing this challenge (Kwakkel & Auping, 2021).

However, the current state of the art in behavior-based scenario discovery is deficient in (at least) two ways. First, behavior-based scenario discovery is currently only applied to a single outcome of interest, or policy objective, by which the scenarios are characterized and differentiated. This is too simplistic for application in complex, real-world use cases, which often involve conflicting and uncertain trade-offs between multiple objectives (Hakanen et al., 2023; Kasprzyk et al., 2013; Kwakkel et al., 2016). Second, there is a lack of methods for analyzing the differences between the various scenarios constituting the generated scenario set. Understanding the quantitative and qualitative differences between the scenarios in the set is crucial for informed decision making, especially regarding the boundaries between them (Bankes, 2011).

In this study, we extend behavior-based scenario discovery to multidimensional scenarios, that is, scenarios including multiple decision-relevant objectives, using a multidimensional clustering algorithm. Furthermore, we demonstrate a number of visual and quantitative methods for evaluating the resulting scenarios which have not previously been applied to scenario discovery. The goal of these contributions is to enhance the usefulness of behavior-based or dynamic scenario discovery for simulation-based decision support, especially where uncertain, complex, and multiactor problems are concerned.

2 | BACKGROUND

2.1 | Scenarios

Scenarios can be defined as a systematized set of plausible future oriented descriptions of a phenomenon that include external context and are comparatively different (Spaniol & Rowland, 2019). In the words of Lustick and Tetlock (2021), "scenarios tend to be colorful inside-view accounts of events as they could unfold if key causal drivers took on either lower or higher values." This accessibility makes them especially attractive for decision makers as on-ramps toward engaging with a problem's full complexity (Wilkinson et al., 2013). Nowadays, scenarios are widely used in decision support in a number of domains including climate change adaptation (Lee et al., 2021), national security (Veldhuis et al., 2020), business (Halim et al., 2016), and public health (Crawford & Wright, 2022).

The effectiveness of scenario-based decision support hinges on the usefulness or fitness for purpose of the underlying scenario set (Sluijs et al., 2021). A number of authors have proposed criteria for evaluating scenario sets. For example, Dhami et al. (2022) proposed completeness, context, plausibility, coherence, and order effects, while Nowack et al. (2011) put forth credibility, transferability, and legitimacy. For an extensive review of such criteria, we point the reader to the review of Amer et al. (2013).

In this study, we follow the scenario criteria put forth by Steinmann et al. (n.d.), namely diversity, plausibility, and comprehensiveness. Diversity implies that the scenarios in the set are qualitatively distinct from one another (Spaniol & Rowland, 2019) and therefore not redundant (Litchfield et al., 2011). This may allow a more comprehensive assessment of the policy interventions (Lord et al., 2016; Wilkinson & Eidinow, 2008). Plausibility refers to the notion that scenarios should represent future states of the world which could actually occur (Lord et al., 2016; Schoemaker, 1993), although no claim is made to the likelihood of this occurrence (Wiek et al., 2013). Finally, comprehensiveness captures the idea that the scenario set should give as complete an account of the system's potential future developments as possible, so as to avoid blind spots (Derbyshire, 2020; Derbyshire & Morgan, 2022). This means reasoning across the widest possible cross-section of the futures cone (Dhami et al., 2022; Gall et al., 2022). This may improve the robustness of the resulting decisions, as a wider range of possible future conditions is considered (Lempert et al., 2006; Rosenhead et al., 1972).

2.2 | Generating scenarios

Generating scenario sets which meet some desired criteria can be accomplished in a number of ways. One example is Intuitive Logics (Bradfield et al., 2005), which first identifies a broad range of factors, reduces these to the most impactful drivers, and then creates scenarios based on a high-low matrix for these drivers. Futures cones (Voros, 2003) are an alternative approach, often starting from a "business-as-usual" base case and then seeking to expand this base case in various directions to grow a cone of plausible alternative developments over time. A common thread across these methods of scenario generation is that they rely heavily on tacit domain expertise and implicit mental models of the studied problem or system to develop the decision-relevant scenarios.

However, there are two main difficulties with this type of scenario generation. First, the human brain is ill-suited to reasoning about nonlinear systems (Sterman, 1994). Furthermore, there is often substantial uncertainty surrounding these systems (Lempert et al., 2003), making precise reasoning and forecasting difficult. As Lamontagne et al. (2018) and Dolan et al. (2021) highlighted, the combination of these two factors implies that approaches based on mental models and human reasoning may not fully identify the decision-relevant scenarios. A growing body of literature from the exploratory modeling (Bankes, 1993) community has therefore investigated the applicability of scenario generation methods based on simulation models. The underlying idea is that by first representing the system in a simulation model, and only then identifying the decision-relevant scenarios based on a thorough analysis of that model, that the aforementioned pitfalls of complexity and uncertainty can be dealt with in a principled and reproducible way. Guivarch et al. (2017) referred to this as a simulate-and-story approach, as a

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

counterpoint to the more established story-and-simulate approach (e.g., Kunc, 2024).

2.3 | Model-based scenario generation: Scenario discovery

Bryant and Lempert (2010) proposed scenario discovery as a method for generating decision-relevant scenarios using simulation models. First, a number of computational experiments are performed on a simulation model, thus creating a data set of external drivers (i.e., model inputs) and resulting dynamics (i.e., model outputs). Then, some criterion for decision relevance is defined, such as a minimal performance level of a specific model output. This criterion is often based on stakeholder objectives or policy goals. It can then be used to identify all computational experiments which are decision-relevant. In the final step, a so-called rule induction algorithm can be used to identify the part of the model's input space from which these decision-relevant outputs originate. The identified input subspace, together with the criterion applied to the outputs, then forms the resulting scenario in the conventional sense—a pairing of external driving forces and resulting dynamics.

Multiple algorithms are available for performing the last step, the rule induction. Lempert et al. (2008) compared the patient rule induction method (PRIM) (Friedman & Fisher, 1999) and classification and regression trees. The former appears to have more uptake in the literature (e.g., Halim et al., 2016; Hidayatno et al., 2020; McJeon et al., 2011; Parker et al., 2015; Popper, 2019; Student et al., 2020b), and has also been extended and improved upon by a number of researchers, including Dalal et al. (2013), who proposed a Principal Components Analysis preprocessing step, Kwakkel and Jaxa-Rozen (2016), who evaluated alternative objective functions, and Kwakkel (2019), who generalized PRIM's core objectives into a many-objective optimization. In essence, PRIM is a method for finding the region (or "subspace") of data space in which a subset of that data with certain characteristics is more commonly found than elsewhere. PRIM rule induction is performed by first drawing a bounding box around all points in the space, and then iteratively reducing the size of this box along one dimension of the space. The choice of dimension to be restricted is driven by three metrics, namely coverage, density, and interpretability. Coverage describes how many of the points of interest are still included in the box. Density captures how many *other* points are still included in the box. Interpretability finally represents how many dimensions have already been restricted. Coverage and density should both be maximized, but in practice often trade off against one another. Interpretability should be minimized, as restricting the box along fewer input space dimensions produces more understandable and intuitive resulting regions. For in-depth explanations of PRIM and this box reduction process, we refer the reader to previous works on the method (Bryant & Lempert, 2010; Kwakkel & Jaxa-Rozen, 2016; Lempert et al., 2008).

2.4 | Extensions of scenario discovery

The original implementation by Bryant and Lempert (2010) focused on finding the input parameter ranges associated with a single set of outputs of interest. A number of researchers have since extended scenario discovery to multiple such sets. This is sometimes called multiclass scenario discovery, and is commonly done by grouping the model outputs in some fashion and then performing rule induction for each cluster in turn (e.g., Gerst et al., 2013; Guivarch et al., 2016; Jafino and Kwakkel, 2021; Rozenberg et al., 2014). An alternative approach is to identify alternative criteria for the decision-relevant model runs, and then perform rule induction for each of those criteria individually (e.g., Greeven et al., 2016; Student et al., 2020a). The former may improve the distinctiveness of the resulting scenarios, while the latter may make them better targeted to the decision problem at hand. In either case, the result is a set of scenarios, which aligns more closely with the conventional understanding and usage of scenarios than the original single-scenario approach by Bryant and Lempert (2010). However, the generation of multiple scenarios also requires subsequent verification that these scenarios fulfill the quality requirements defined for the decision context at hand. The evaluation of diversity, which has also been called distinctiveness (Lord et al., 2016) or separability (Jafino & Kwakkel, 2021), deserves special attention in this regard as the model-generated scenarios cannot be assumed to be diverse in a human-interpretable way. Steinmann et al. (2020) proposed using overlap between the different scenario regions to evaluate separability, while Jafino and Kwakkel (2021) incorporated separability directly into the scenario generation process by clustering in the in- and output spaces simultaneously. Finally, Gerst et al. (2013) used visual inspection to evaluate separability.

The identification of the decision-relevant model outputs (sometimes referred to as the "outputs of interest") before rule induction is a critical step in successful scenario generation with scenario discovery. The key challenge is that the model outputs, which are often multidimensional time series, must be reduced to a single binary variable. The choice of criterion by which to do this has downstream effects, and deserves substantial attention during the analytical process (Hitch, 1960). A common approach is to apply a threshold criterion to the last value in one of the model's time series outputs. However, multidimensional criteria, or criteria based on statistical properties of the time series outputs such as mean or amplitude values, can also be imagined. Steinmann et al. (2020) proposed an alternative approach using time series clustering to first find clusters of similar dynamics among one of the model's outputs, and then perform rule induction for each cluster in turn. This approach, referred to as behavior-based or dynamic scenario discovery, allows the dynamics of the system to be directly considered when creating the scenarios (Kwakkel & Auping, 2021). It was based on earlier work by Kwakkel et al. (2013), Gerst et al. (2013), and Guivarch et al. (2016), and has since successfully been applied in at least one decision support context (Kahagalage et al., 2024). An added benefit of behavior-based scenario discovery is that the dynamics in the resulting clusters can more readily be translated into

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

verbal narratives than the point values common in conventional scenario discovery, which was also demonstrated by Greeven et al. (2016). The overall result of behavior-based scenario discovery is a systematized set of model output clusters with distinct dynamics, each associated with a defined region of the input space. This is the sense in which we use the term "scenario"—a combination of a contiguous region of a model's input parameter space and a set of qualitatively similar model outputs which originate from that region. In our view, this is the usage most consistent with established definitions of the term (e.g., Spaniol & Rowland, 2019) and more conventional scenario generation methods.

3 | METHODS

In this section, we describe our methodological approach to generating multivariate scenarios with a simulation model, as well as analyzing the resulting scenarios. We closely follow the original approach for scenario discovery with clusters of time series introduced by Steinmann et al. (2020), but extended to multivariate model outputs. In this sense, we draw heavily upon both extensions to scenario discovery described previously: the consideration of multiple classes of outcomes, and the consideration of temporal dynamics. First, we describe our case study and associated simulation model, which we use to demonstrate our approach. Second, we document the conducted simulation experiments, through which we generated the data underlying the generated scenarios. Third, we describe the multivariate clustering approach used to identify the distinct dynamics associated with the scenarios. Last, we document the rule induction performed to link the dynamics to their generative parameter input ranges, and the reconstruction process used to verify the rule induction.

3.1 | Case study

The fictional decision making context which we use throughout this paper to demonstrate our methodological innovations is a protest movement with potential for violent escalation. In particular, we are interested in generating scenarios which high-level decision makers in national security and law enforcement may use to evaluate policy intervention alternatives. These alternatives might theoretically include the deployment of law enforcement personnel, monitoring of the movement, or information operations, up to the repression of the movement due to an excessive risk to public safety and security. This is a suitable case study for behavior-based scenario discovery for multiple reasons. First, the system is characterized by dynamic complexity (Sterman, 2002) and delays in policy response, making temporal analysis necessary. Second, ensuring public safety and security is an ongoing task not defined by an end state (as is common in scenario discovery, see for example, work by Bryant and Lempert, 2010; Student et al., 2020b). We assume in this decision context that it is possible to create a simulation model of the protest movement in question which is accepted and considered useful by all relevant stakeholders.

A system dynamics model of the fictional protest movement supports our case study. The purpose of our model is to explore the potential evolution of a protest movement through simulation. As such, the goal of our modeling efforts is not to calibrate our model to a specific protest movement. Rather, we include dynamics like growth, decline, and escalation of violence that characterize the behavior patterns of protest movements in general. The model is exploratory in nature, and should not be regarded as a fully validated model with which protest dynamics can be studied.

The underlying dynamics of a complex societal phenomenon are usually hard to grasp and understand. System dynamics models are able to include a multitude of assumptions on causal relationships of and between societal dynamics, and therefore these models are applicable to study a system's emergent behavior by analyzing the impact of feedback loops and nonlinear behavior (Sterman, 2000; Veldhuis et al., 2023). The core factors and causal relationships of our simulation model, visualized in Figure 1, are adapted from theory driven modeling studies focused on the dynamics of protest movements (Alsulami et al., 2022; van der Zwet et al., 2022).

The core stock-flow-feedback mechanism of our model describes the behavioral transition of people from being a potential protester to being mobilized into an active protester, and eventually fatigue or withdrawing as a retired protester. Active protesters are divided into two groups: new protesters and experienced protesters. New protesters typically have a higher exhaustion rate and lower mobilization effect on potential protesters, and therefore a lower positive impact on the longevity of the protest movement compared to experienced protesters (Alsulami et al., 2022). New protesters mature toward experienced protesters as they spend time together organizing and participating in protest activities. It can be assumed experienced protesters establish more strong relationships that yield stronger collective effects. If the core group of experienced protesters exceeds a certain threshold, the *limit core support*, the collective effect makes them less likely to retire.

The *disturbance* is the trigger for the mobilization dynamic of the model. This factor mimics the level of animosity a certain part of the population has toward a specific government or other type of relevant topic. The disturbance is modeled as a sigmoidal function that transits from a "relaxed" state to a "tense" state, in which the mobilization effects are larger and protests activities are more legitimized by the population (Gallo, 2013). The *threshold disturbance* factor determines when a disturbance overshoots the tolerance in a society and the tension is triggered. We assume that disturbances could occur with different intensities and duration. The *steepness disturbance* factor influences how quickly the transition from the relaxed state to the tense state occurs.

The output of the mobilization dynamic are the protests organized by the protesters. It is assumed that the number of protests is a function of the number of protesters. A percentage of these protests escalate into violent protest or riots depending on a number of factors: a high rate of new protesters compared to experienced

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

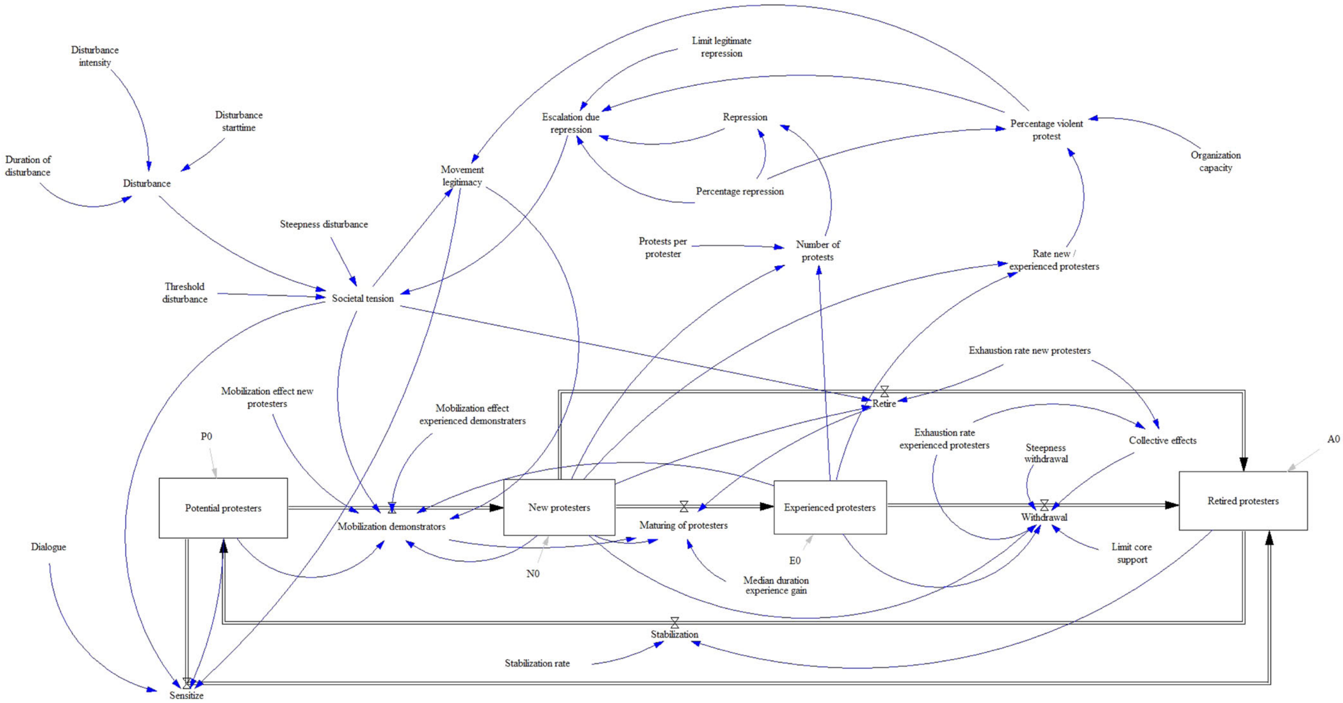

FIGURE 1 System dynamics model of a fictional protest movement.

protesters and a lack of organizational capacity positively influence the rate of protests that escalate (Gustafson, 2020). Furthermore, repression is an instrument aimed at limiting the negative effects of protests, however, it causes more protests that lack organizational capacity to escalate into violence. These dynamics have an intensifying impact on the conflict situation as they positively influence the legitimacy of the movement and the escalation due to repression, if the level of repression is higher than what the level of escalation legitimizes.

3.2 | Simulation experiments

To generate a wide range of plausible future protest movement dynamics, we simulated the protest movement model described above using a variety of parameter settings, each producing a unique plausible future behavior of the movement. In total, we conducted 1000 simulation experiments with unique input parameter value combinations. This number proved sufficient to generate diverse and interesting clusters without making the cluster computation process excessively long. As the model is deterministic, we did not perform any replications of the individual combinations. The sampling of the input parameter value combinations, as well as the processing and storage of the simulation experiments, was done in Python using the Exploratory Modeling & Analysis Workbench (Kwakkel, 2017). The simulation experiments themselves were run in Vensim DSS 8.1 using the Runge-Kutte 4 Auto integration technique over 2000 time steps. Of these time steps, the first 200 time steps (those before the exogenously introduced disturbance) were considered part of the model's warm-up period, and discarded before analysis.

To create the unique input parameter value combinations used for the simulation experiments, we used Latin Hypercube Sampling (McKay et al., 1979) to uniformly cover the input space. The parameter ranges defining this space are given in Table 1. We empirically identified these parameter ranges, first based on literature and the model design, and then by exploring the parameter space for interesting and diverse model dynamics.

Based on discussions with stakeholders, domain experts, and model exploration, we identified two model outcomes of interest, namely the *number of protests*, and the *number of experienced protesters*. These two outcomes may be considered decision-relevant in the context of the fictional protest movement depicted by our model.

3.3 | Multidimensional time series clustering

To identify underlying behavioral patterns across the simulation experiments, we clustered the resulting data across the two aforementioned outcomes of interest. The simulation data was first preprocessed individually per outcome using the standard scaler in the Python package scikit-learn (Pedregosa et al., 2011), thus removing the mean and scaling all model output time series to unit variance. This was necessary as the observations differed in their amplitudes and means by multiple orders of magnitude.

For the clustering step, we used k-means clustering with a Dynamic Time Warping distance metric, implemented in the

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Input parameter ranges for sampling.

Model parameter	Range	Unit
Duration of disturbance	[1, 15]	Days
Disturbance intensity	[0.4, 1.8]	Dimensionless
Threshold disturbance	[0.4, 0.8]	Dimensionless
Steepness disturbance	[1, 8]	Dimensionless
P0	[40, 100]	Protesters
N0	[50, 100]	Protesters
E0	[0, 80]	Protesters
A0	[0, 60]	Protesters
Mobilization effect new protesters	[0, 0.5]	Protesters/ protester/day
Mobilization effect experienced protesters	[0, 0.5]	Protesters/ protester/day
Exhaustion rate new protesters	[0, 0.1]	Protesters/day
Exhaustion rate experienced protesters	[0.02, 0.1]	Protesters/day
Median duration experience gain	[14, 35]	Days
Limit core support	[10, 100]	Protesters
Steepness	[5, 10]	Dimensionless
Stabilization rate	[0.002, 0.008]	Protesters/ protester/day
Protests per protester	[0.04, 0.1]	Protests/protester/day
Organization capacity	[5, 10]	Dimensionless
Limit legitimate repression	[0, 0.6]	Protests/repression
Percentage repression	[0.1, 0.3]	Protests/repression

Python package tslearn (Tavenard et al., 2020). The clustering was performed across both outputs of interest simultaneously. To reduce computation demands in the clustering phase, we restricted the clustering to the first 800 postdisturbance time steps of the model runs, where the largest differences in dynamics could be observed. We chose to search for $k = 5$ clusters. This number was chosen empirically based on subsequent analysis steps, but has some grounding in literature. Notably, Lord et al. (2016) advocated using between four and six scenarios, as three scenarios may be interpreted as a high-moderate-low arrangement to the detriment of the resulting decision (Goodwin et al., 2019), and seven approaches the limit of human working memory (Miller, 1956). In the original paper describing behavior-based scenario discovery, the authors chose six clusters (Steinmann et al., 2020), while in another application of behavior-based scenario discovery, Kahagalage et al. (2024) explicitly chose three scenarios to frame the resulting clusters as high, moderate, and low performance levels of the studied system.

3.4 | Generative input subspaces per cluster

To identify the regions of the input space (also called subspaces) from which the majority of each cluster's included simulation experiments originate, we applied a rule induction technique called the Patient Rule Induction Method or PRIM (Friedman & Fisher, 1999), again implemented in the Exploratory Modeling & Analysis Workbench, with an updated objective function (Kwakkel & Jaxa-Rozen, 2016). In our case, we applied PRIM to the input parameter combinations associated with each cluster's constituent model outcomes, in line with previous work by Steinmann et al. (2020). To facilitate later analysis, we restricted the rule induction algorithm to a small number of input parameter space dimensions which had comparatively greater influence on model dynamics and clustering outcomes. Building on the work by Weinans et al. (2024) on identifying behavior-relevant parameters, we used visual inspection of model input-output data as well as global sensitivity analysis using the PAWN method (Pianosi et al., 2016) to select these input dimensions. The latter was implemented in the Python package SALib (Herman & Usher, 2017; Iwanaga et al., 2022). PAWN, named after the two lead developers, measures the influence of a specific model input as the variation in a model output's cumulative distribution function (CDF) when the uncertainty about that input is removed. This is done by comparing the unconditional and conditional (i.e., with a fixed value for the specific model input) CDFs using the Kolmogorov–Smirnov statistic. To ensure comparability across the clusters, we chose the first box in the PRIM peeling trajectory of each cluster with a density of 80% as that cluster's representative subspace, or, if such a box could not achieved, the highest-density box available.

To verify that the identified input parameter regions were predictive for their associated clusters' dynamics, we reconstructed each cluster's dynamics from their respective input subspaces. To do so, we drew 50 samples, again using Latin Hypercube Sampling, from each cluster's associated input region, and then used the previously described Python-Vensim simulation setup to run simulation experiments for each of those samples. We then visually compared the resulting model dynamics with those originally identified through the clustering algorithm.

Furthermore, we created verbal narratives for each scenario by manually assessing each cluster's dynamics, the underlying parameter ranges, and the model's causal relations. In writing the narratives, we relied on the scenario criteria proposed earlier, as well as the scenario narrative evaluations proposed by Dhami et al. (2022). The resulting narratives may be more accessible or suitable for certain decision support contexts or communication channels than the quantitative, data-heavy model outputs underlying them.

Finally, to inspect the resulting scenarios describing the fictional protest movement's plausible future developments, as well as their separability, we used a combination of visual and statistical analysis techniques. Hakanen et al. (2023) refer to such approaches as "coordinated multiple views," and consider them especially suitable for high-dimensional model outputs. We leveraged recent advances in in the visual inspection of many-objective optimization outcomes

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Filipic & Tusar, 2018; Osika et al., 2023) to develop two novel visual representations of multiclass scenario subspaces in the form of pairwise grid plots and parallel coordinate plots. Furthermore, we explored the representation of the subspaces as intersecting sets using a so-called upset plot (Lex et al., 2014), an extension of initial attempts at quantifying overlap by Steinmann et al. (2020). Upset plots are widely used in the biomedical domain (Gadhave et al., 2019) to analyze sets and their various intersections. We apply this idea to the subspaces and the sampled points they contain. As some points lie within two (or more) subspaces' boundaries, their dynamics may be attributed to multiple clusters, which might in turn introduce ambiguity into the decision support process. By quantifying the box subspace intersections, a better understanding of the separation between the different subspaces can be obtained.

4 | RESULTS

4.1 | Clustering outcomes

The results of the multivariate clustering applied to our model's two decision-relevant outcomes are shown in Figure 2 for all five clusters together in the top row, and individually below. It is apparent that each cluster shows similar internal dynamics, but that these dynamics are qualitatively different across the clusters, especially when considering both outcomes of interest. Furthermore, the dynamics of the number of protests and number of experienced protesters for each cluster are quite similar. This can be explained by the fact that experienced protesters are a large driver of protest activity in our model. The labels given to each cluster are the narrative names introduced later in Section 4.4—for consistency, we use them throughout the manuscript. These labels were partially manually assigned, and partially generated using AI based on the verbal narrative descriptions. We note here that the chosen colors do not imply any qualitative assessment of the clusters - for example, cluster Intensity is not necessarily more desirable than cluster Indifference, despite the green and red colors, respectively. We further observe that the clusters are not evenly sized. The Waves cluster contains 314 model runs, cluster Hidden Threat contains 82 model runs, cluster Intensity contains 70 model runs, cluster Indifference contains 183 model runs, and cluster Flickering contains 351 model runs. The data underlying this clustering was generated in roughly 2 min, and the clustering itself took roughly 20 min, both on a standard workstation. We note here that, due to the uniform sampling, the relative frequency of certain model runs and the associated cluster sizes do not imply anything about the real-world probabilities of the different clusters emerging in the model.

In Figure 3, we visualize the 1000 input parameter combinations in the model's input space, colored by the cluster their resultant model outputs were assigned to. In the scatter plot in the lower triangle, we observe that distinct regions of colored points are visible. For example, orange points, denoting cluster Waves, seem largely associated with higher stabilization rate values and lower values for the exhaustion rate of new protesters. In the filled kernel density estimate plots on the diagonal, the relative change in assigned clusters over each variable's range is shown. Here too certain trends are

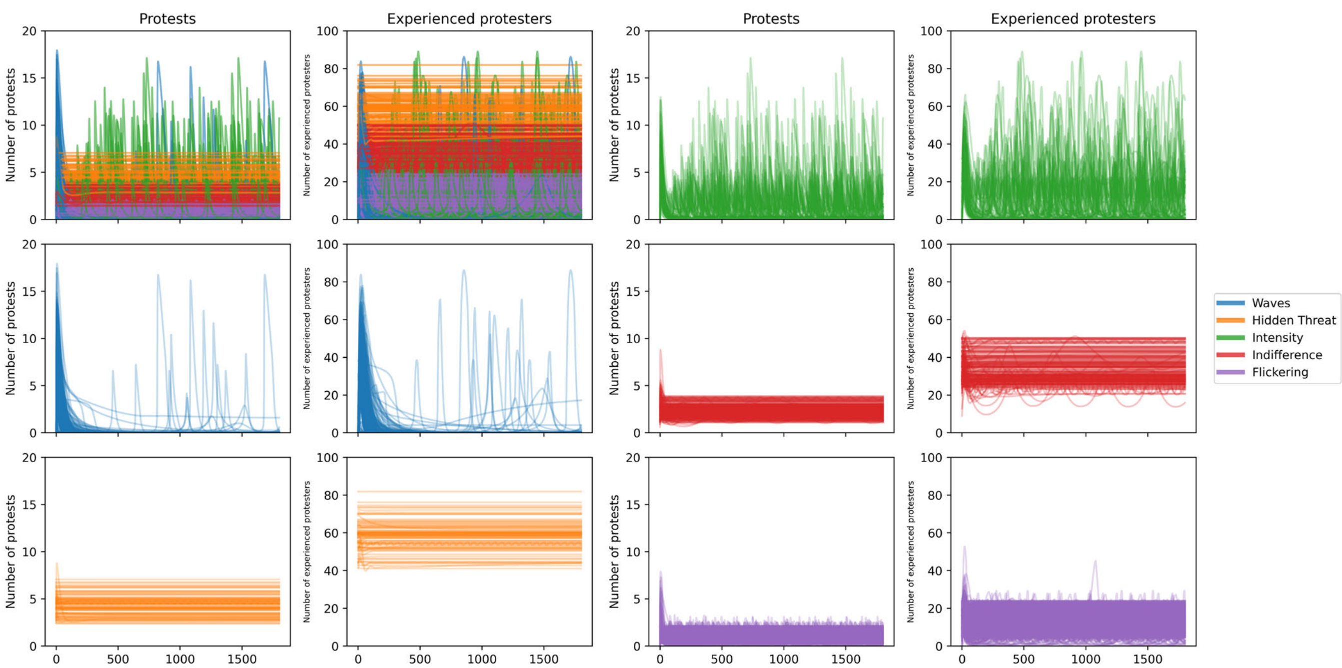

FIGURE 2 Identified clusters ($k = 5$) for 1000 model runs across two outcomes of interest (number of protests, and number of experienced protesters). The top-left pair of figures shows all clusters together for the two outcomes of interest, and the remaining pairs of figures show the individual clusters, paired by color. The x-axes represent time over the model run, and the y-axes the number of protests and experienced protesters, respectively. The cluster labels are the narrative names introduced in Section 4.4.

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technishe Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

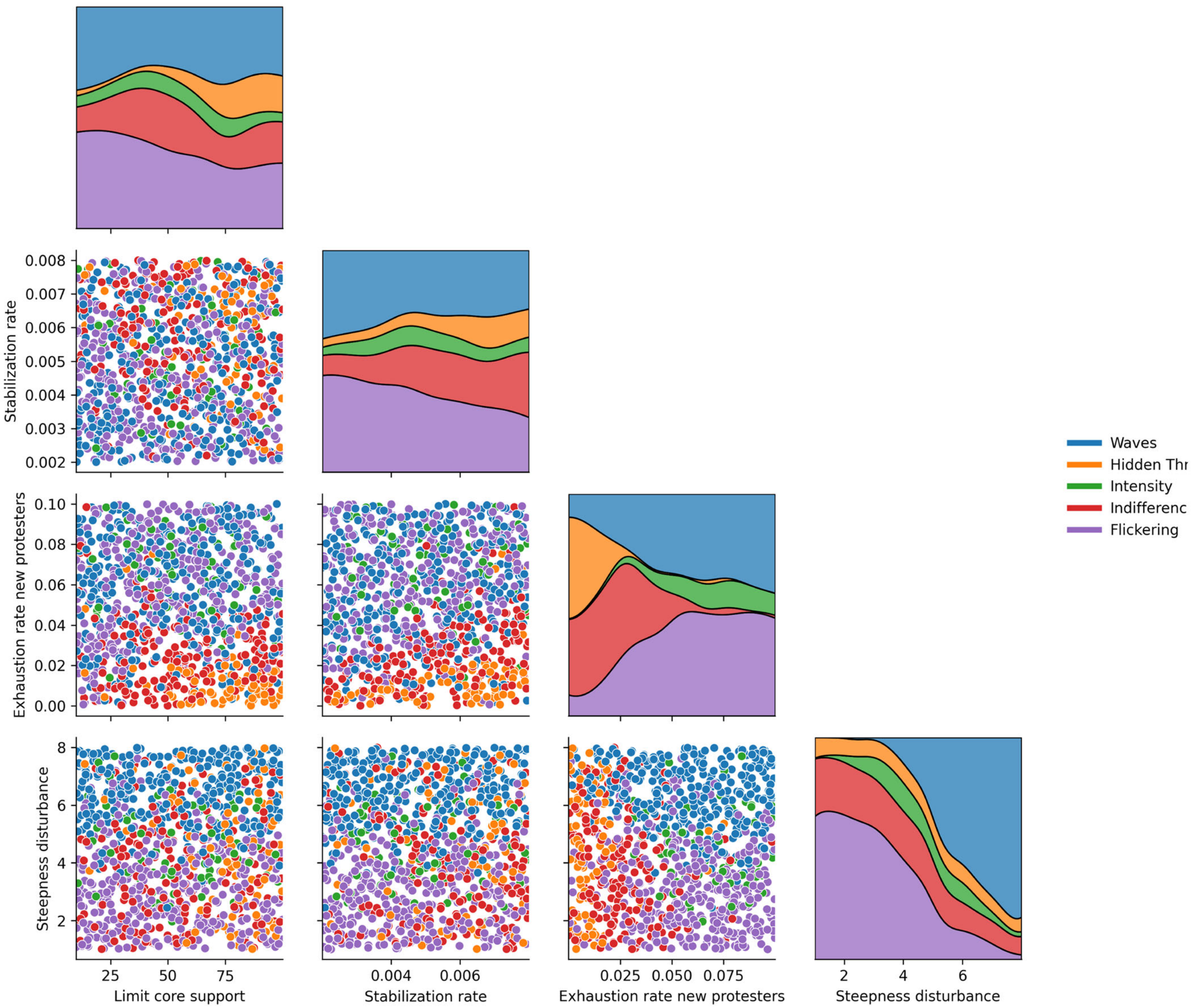

FIGURE 3 Pair grid plot for model inputs and their associated output clusters. Every subplot represents a pair of input parameters, except for the subplots on the diagonal which refer to a single input parameter. In the lower triangle, every point represents a single model run's input parameter values, the color of the point indicating the cluster the resulting outputs were assigned to. The subplots on the diagonal are filled kernel density estimates, and represent the relative distribution of the different clusters as the parameter value changes. Cluster colors match with previous figure.

apparent. For example, cluster Waves is only present above values of 4 for steepness disturbance, while cluster Flickering virtually disappears around a value of 6 for the same input parameter. It is also apparent that certain input parameters have very little effect on some clusters. For example, the number of model runs assigned to cluster Intensity is virtually invariant to the value of the stabilization rate parameter. In the Supporting Information S1: File S1, we also present contour plots for each cluster individually.

4.2 | Generative input subspaces per cluster

When performing rule induction with PRIM for each identified cluster, the subspaces documented in Table 2 result. These regions of the model's input space were found by PRIM to be predictive of the associated cluster, that is, an input parameter combination sampled from within one of these regions is commonly assigned to that region's associated cluster in the output space. The probability of this assignment is ≥80%, as this was the chosen density threshold for the rule induction. Table 3 gives summary coverage, density and interpretability statistics for each induced subspace.

4.3 | Cluster reconstruction

To verify that the subspaces identified with PRIM for each cluster actually represent their clusters' dynamics well, we reconstruct each cluster from its input subspace by performing another set of

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Induced input parameter subspaces for each cluster.

Model parameter	Waves	Hidden threat	Intensity	Indifference	Flickering
Steepness disturbance	[5.25, 7.99]	[1.41, 7.82]	[2.46, 6.14]	[1.00, 4.94]	[1.00, 3.71]
Exhaustion rate new protesters	[0.033, 0.100]	[0.000, 0.017]	[0.031, 0.094]	[0.014, 0.039]	[0.027, 0.100]
Limit core support	[10.03, 99.96]	[66.95, 99.96]	[30.78, 80.61]	[30.19, 92.39]	[10.03, 99.96]
Stabilization rate	[0.002, 0.008]	[0.002, 0.008]	[0.003, 0.007]	[0.004, 0.008]	[0.002, 0.007]

TABLE 3 PRIM details for each cluster's subspace.

Cluster	Coverage	Density	Interpretability
Waves	0.67	0.82	2
Hidden threat	0.61	0.76	4
Intensity	0.51	0.27	4
Indifference	0.30	0.81	4
Flickering	0.56	0.81	3

Note: Coverage represents the percentage of cluster members included in the found subspace, density the ratio of included cluster members over total points in the box, and interpretability the number of restricted dimensions necessary. Abbreviation: PRIM, patient rule induction method.

simulation experiments for each cluster's generative input subspace. The number of experiments for each cluster is based on the cluster sizes described in Section 4.1. In Figure 4, we show the original and reconstructed dynamics for the first outcome of interest, the number of protests. We restrict ourselves to one outcome of interest here because the two outcomes are highly correlated (evident in Figure 2). It is visually apparent that for every cluster, the reconstructed dynamics are similar to the original ones. This indicates that the identified regions are indeed predictive for the associated clusters' dynamics.

4.4 | Narratives

Based on the identified clusters and their associated input parameter regions, the following short narratives may be associated with each identified cluster.

- *Waves*: A short but intense period of protests immediately follows the initial external disturbance. Afterwards, there is a period of relative calm, with few to none daily protests. Subsequent to this quiet period, the protest movement regains momentum and re-establishes daily protests at levels comparable to the initial activity period. The number of experienced protesters closely matches this pattern—an initial surge in numbers, followed by the movement almost completely dying out, only to return to manpower levels comparable to the first phase of protests. This dynamic is driven by a high steepness of the disturbance, and a rapid exhaustion rate of new protesters.
- *Hidden threat*: The movement maintains a relatively low, but constant frequency of protests. There is no substantial change over time. While few protests are ever observed, a relatively large number of experienced protesters forms a hard core of the movement, keeping it alive, and potentially serving as a breeding ground for a future escalation. The main drivers of this scenario are a very low exhaustion rate of new protesters, as well as a high threshold value for the core support of the movement.
- *Intensity*: The protest movement maintains a high activity rate. Although the absolute number of protests varies, the movement is consistently active. The number of experienced protesters is also relatively high, although it too shrinks and grows rapidly. This behavior pattern is driven by intermediate values for the steepness of disturbance, stabilization rate, and limit of the movement's core support, and a high value for the exhaustion rate of new protesters.
- *Indifference*: Apart from a minor initial spike, the movement never gains the momentum or followers necessary for sustained and intense protest activity, staying consistently at low activity levels. However, there is also no sigh of the movement completely collapsing. The dynamics of this scenario can be traced to a low value for the exhaustion rate of new protesters and steepness of disturbance, and intermediate to high values for the limit of core support and stabilization rate.
- *Flickering*: While the protest movement never establishes high activity and support levels, it frequently gains and loses momentum at a lower level, showing repeated signs of life and warranting constant attention. The key drivers of this behavior pattern are a low steepness of disturbance, low to intermediate value for stabilization rate, and intermediate to high value for the exhaustion rate of new protesters.

As the five four-dimensional subregions described in the table are difficult to intuitively visualize, we present them in another pairwise grid plot in Figure 5. This allows the spatial positions of the five boxes to be visually analyzed for every pair of input parameter dimensions individually. In a confirmation of observations already made based on Figure 3, it is apparent that not all input dimensions are predictive for every cluster. For example, the Waves cluster is essentially independent of the value of the core support limit. However, certain interactions and alignments between the different boxes are also visible, notably for the exhaustion rate of new protesters (Hidden Threat and Indifference are clearly separated from

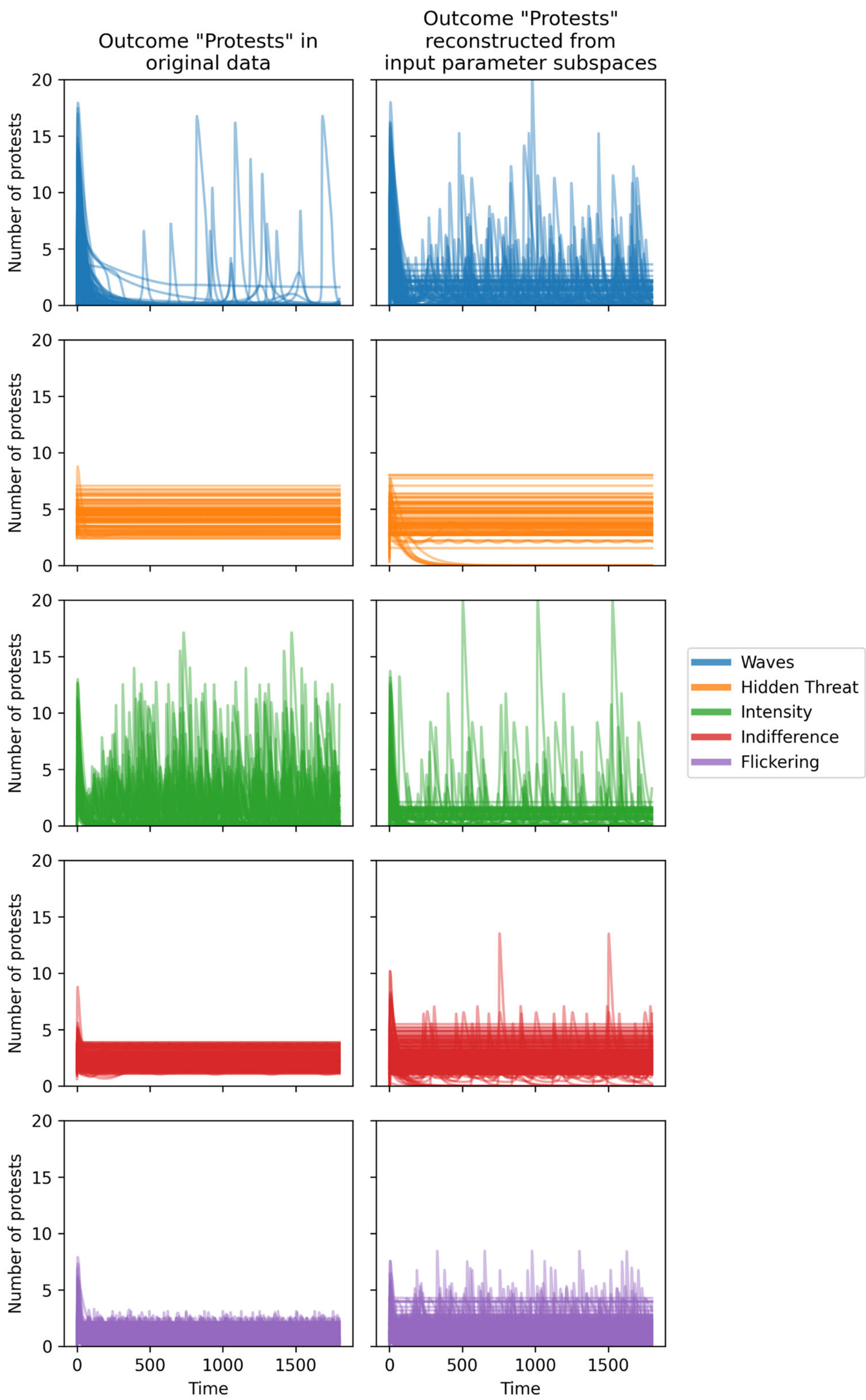

FIGURE 4 Comparison of clusters identified in the originally generated data set, and ensembles of model runs generated from each cluster's input parameter subspace, for a single outcome of interest. The pairwise behavioral similarity shows that the identified subspaces are indeed predictive for each cluster's unique dynamics. Cluster colors match with previous figures.

Waves, Intensity and Flickering, which overlap considerably) and the steepness disturbance (Waves and Flickering are clearly separated, while Intensity and Indifference overlap). In the plot combining these two parameters, a distinct separation between all five clusters is visible, apart from minor overlap at the boundaries of some clusters.

In Figure 6, we provide a novel visualization of PRIM subspace regions using a parallel coordinate plot with shaded areas

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technishe Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

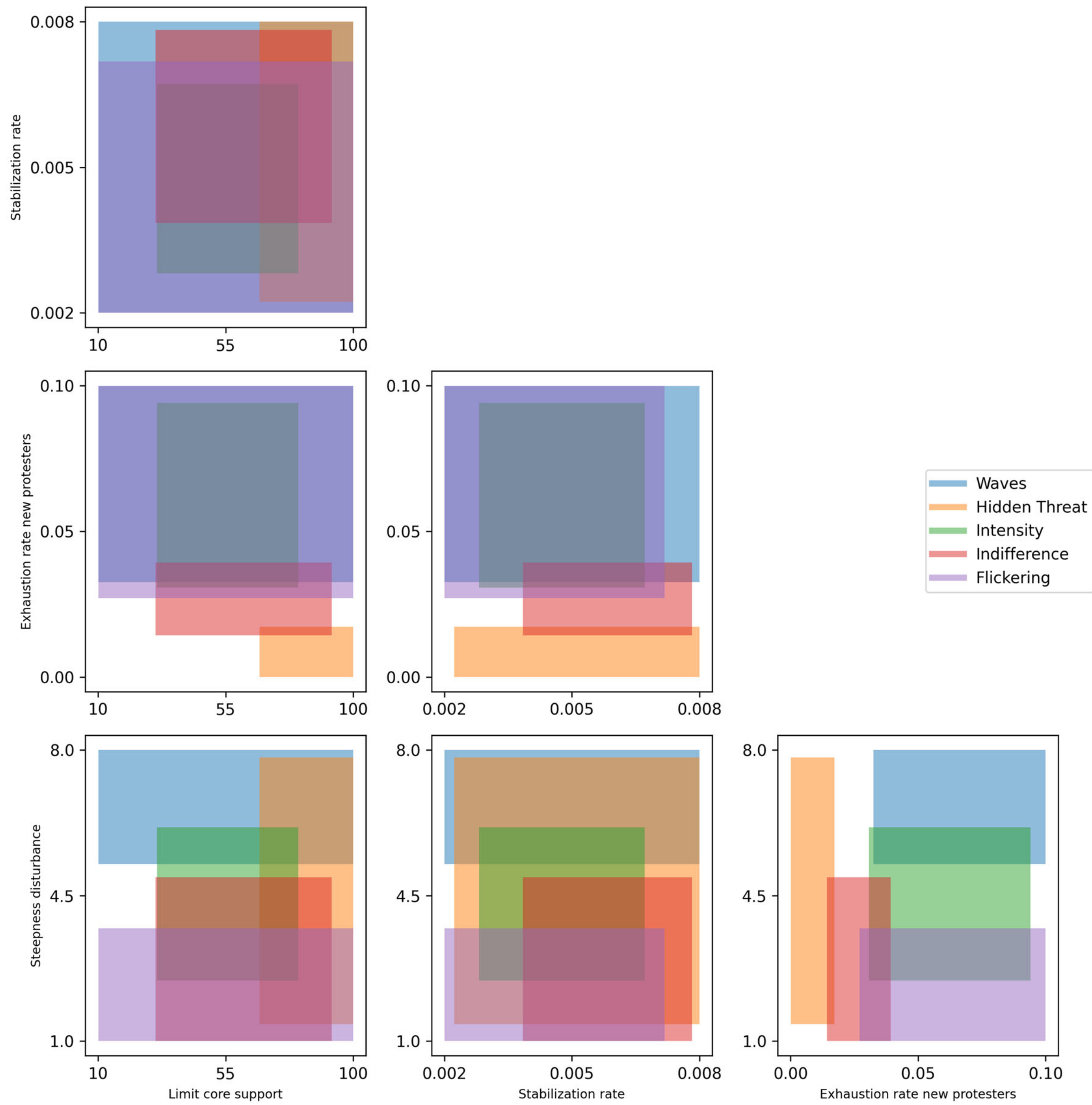

FIGURE 5 Generative input subspaces for each cluster, shown for the four sensitive input parameter axes shown in Figure 3. Each box denotes a two-dimensional shadow of the four-dimensional hypercube which contains most of the input parameter combinations belonging to a specific cluster. Cluster colors match with previous figures.

representing each cluster's predictive range for each studied input parameter axis. This allows a direct comparison of the similarities between different clusters' subregions, as they are shown in the same figure in their entirety (as opposed to the lower-dimensional shadows in Figure 5. For example, we may observe that clusters 2 and 3 are roughly comparable for the first two plotted axes, but diverge markedly otherwise. The figure also makes it apparent how some clusters are very narrowly demarcated along certain axes (such as cluster Hidden Threat for the third axis *Exhaustion rate new protesters*), and very widely along others (e.g., cluster Hidden Threat along the fourth axis *Steepness disturbance*). We note here that these parameter ranges, as described in Section 3.2, were chosen for the dynamics they created with our stylized model rather than any real-world equivalence.

4.5 | Cluster separability

To better understand the overlaps and spatial interactions of the five clusters' input subspace regions, we represent the model runs in each subregion as a set, and analyze the intersections of these sets with an upset plot in Figure 7. On the bottom, the individual clusters form one row each, the bars on the left-hand side showing how many model runs lie within the input parameter subspace associated with that cluster. The dots and lines toward the right represent intersections or overlaps between these subspaces, with the bars above the dots showing how great these intersections are. For example, the Hidden Threat and Indifference subspaces contain 66 and 68 model runs, respectively, but have an overlap of only one model run. Thus, they are well-separated. This can be corroborated by considering

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

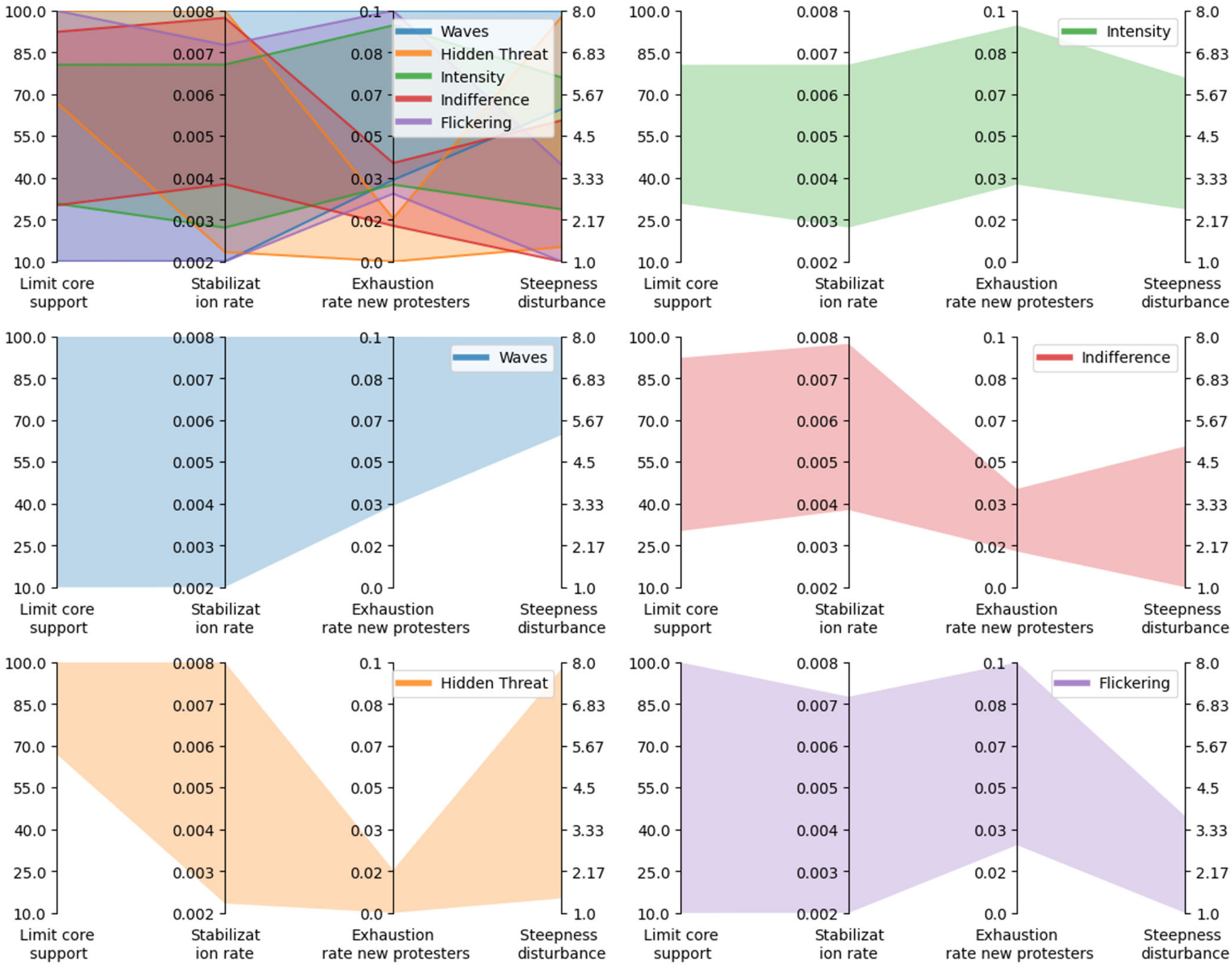

FIGURE 6 Generative input subspaces for the four sensitive input parameter axes shown in Figure 3 as parallel coordinates. The top-left figure shows all clusters together, and the remaining figures the individual input subspaces per cluster. Colors match with previous figures.

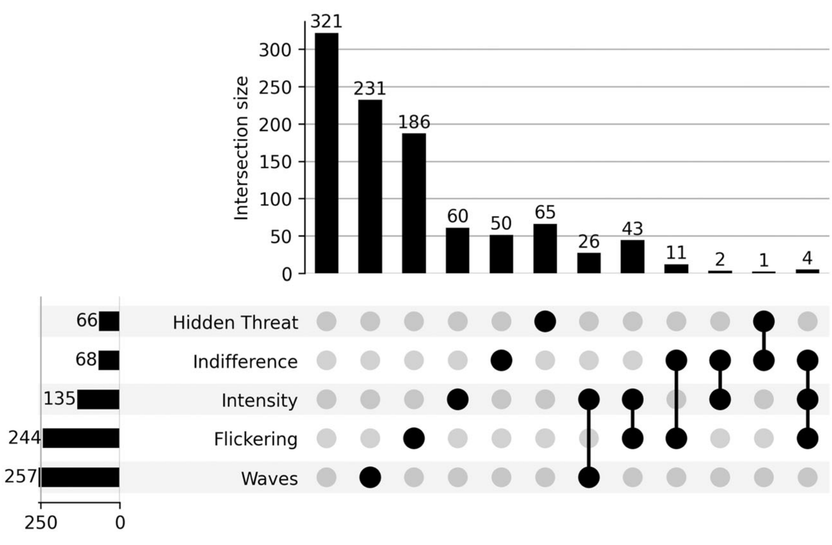

FIGURE 7 Upset plot of the generative input subspaces for each cluster, showing input parameter combinations sampled for the 1000 model runs which lie within the subspace boundaries of one or more of the boxes shown in Figure 5. Unlisted intersections did not occur.

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 5, where it is visible that these two clusters have only minimal overlap in the parameter dimension *Exhaustion rate new protesters*. The upset plot thus adds a quantitative underpinning to the visual inspection possible with pairwise grid plot in Figure 5.

A number of observations can be drawn from this figure. First, most model runs only lie within a single cluster's input space subregion. This is desirable, as it indicates that the subregions are well separated and distinct. The greatest overlap is between clusters 2 and 4. This makes intuitive sense, as the respective dynamics of these two clusters do not differ much qualitatively, only quantitatively in terms of their amplitude. It is therefore not surprising that they are close together in the input space. However, we also notice that 321 model runs are not associated with any cluster. In other words, for almost one-third of all studied plausible future dynamics of the protest movement, it is unclear to which scenario they belong. By extension, this implies that one-third of the model's input space is not covered by any cluster's subspace. Finally, only four model runs lie within the subspaces of three different clusters.

5 | DISCUSSION

In the presented work, we extended behavior-based scenario discovery to multivariate outcomes in a case study of protest movement dynamics. Furthermore, we introduced a number of novel analytical techniques for making sense of the resulting model-generated scenarios of protest intensity and movement support. In this section, we discuss observations from our case study, and implications of our methodological innovations.

5.1 | Observations from case study

Our application of multidimensional behavior-based scenario discovery to the case study of a fictional protest movement yielded five distinct and interesting scenarios. Each scenario provides a unique description of a plausible future development of the movement regarding two decision-relevant objectives, the number of protests per day and the number of experienced protesters. It is noteworthy that none of the scenarios represent conventional scenario archetypes such as utopia or dystopia—every generated scenario poses its own challenges to decision makers. This is a key advantage of generating scenarios with simulation models (Guivarch et al., 2017), rather than creating them directly using structured methods such as futures cones or Intuitive Logics. It may also improve the robustness of the resulting decisions and subsequent policing actions, as they have been evaluated against a wider range of challenging scenarios (Lamontagne et al., 2018).

The extension of behavior-based scenario discovery (Steinmann et al., 2020) to multidimensional model outputs over time generates scenarios which may be more useful for decision support processes. For example, through the addition of a second dimension, a clear difference emerged between clusters 1 and 3, which have comparable protest numbers, but differ substantially regarding the number of protesters. While our case study included only two decision-relevant dimensions, more dimensions could easily be added, creating even richer and more complex scenarios. This may be especially interesting for output dimensions which are less closely correlated than the ones we used in our case study.

When performing the rule induction for each cluster's generative input subspace, PRIM generates an entire trajectory of possible subspaces, which differ regarding density, coverage, and interpretability. We chose the first box to reach a density of 80% (or, failing that, the highest-density box available), giving a relatively small but highly predictive result for each cluster. The 80% density threshold was reached for three out of five clusters, with a fourth cluster reaching 75% density. Using small, high-density boxes is one possible solution to PRIM's orthogonality constraints (see Auping, 2018; Quinn et al., 2017), and provided relatively good reconstructions of each cluster's dynamics from its associated subspaces. However, it also had the drawback of relatively low coverage values for the individual subspaces (see Table 3, which by extension left almost one-third of all model runs not assigned to any cluster. These runs stem from the uncovered or white regions visible in Figure 5. Care must be taken when interpreting these unassigned model runs—it is not that their dynamics are individually fundamentally different from those included in the five found subspaces, but that their generative input parameter combinations are difficult to distinguish from those of model runs assigned to other clusters. In effect, the uncovered regions may be seen as a sort of fuzzy boundary space between the high-density subspaces of the five scenarios.

Whether it is acceptable to have such a degree of uncertainty surrounding the scenarios and their respective input subspaces will depend on the decision support context. In domains which are accustomed to operating under uncertainty, such as public safety and security, the incompleteness of the scenarios may be more acceptable than in aviation or nuclear energy, for example. We note here that the choice of density threshold value is itself a many-objective optimization problem between minimizing the number of model runs not included in any scenario's generative subspace, maximizing the number of runs included in a single scenario's generative subspace, and generating roughly equally-sized scenarios.

The verbal narratives which we wrote to accompany each cluster's associated dynamics and identified input parameter region complemented the underlying quantitative data. However, this writing was a relatively free-form process, as we found little guidance in the literature on how to craft a scenario narrative based on a simulation model's dynamics. Length, style, and focus of these narratives could be adjusted. However, it is unclear how this might affect the perception of the scenarios, with associated impacts on the resulting decisions.

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5.2 | Methodological innovations

Scenario discovery was originally proposed by Bryant and Lempert (2010) for creating a single scenario based on a single outcome of interest's value at a single point in time. Subsequently, the method has been enhanced by Gerst et al. (2013) (among others) for multi-dimensional outcomes of interest, and by Kwakkel et al. (2013) for outcomes of interest over time. This work combines these two enhancements to enable the generation of multidimensional, temporal scenarios. This makes scenario discovery increasingly useful for decision support in complex, real-world problem contexts, which are commonly multidimensional and temporally sensitive (Kwakkel & Auping, 2021; Osika et al., 2023). Furthermore, by clustering the model's outputs in multiple dimensions, the resulting clusters may be associated with more separable input subspaces, as the outputs can more easily be differentiated. The outputs of our method are also more comparable to those resulting from other scenario methods, in that they are multidimensional and include narratives. This raises the possibility of using an ensemble of scenario methods, and thus creating a scenario "superset" containing plausible futures generated with a variety of methods. This may further improve the robustness and effectiveness of the resulting decisions, as they have been evaluated against a wider range of scenarios.

By visually and statistically inspecting both the individual scenario regions and their relations to one another across multiple plot types, a more holistic understanding of the scenarios can be gained than through comparison of their output dynamics alone. This has a number of potential benefits. First, the differences between the scenarios become clearer, especially regarding their underlying driving forces. Second, the effectiveness of policy interventions for each scenario can be assessed more accurately, and interventions targeting specific scenarios can be designed. This allows for more granular and adaptive policy design, which is a key tool for coping with modern societal challenges in the face of uncertainty (Kwakkel & Haasnoot, 2019). Finally, the visual representation of the scenarios using the various presented methods may facilitate communication and explanation of the quantitative, multidimensional data underlying the scenarios.

6 | CONCLUSION

We extended behavior-based scenario discovery with multivariate time series clustering to generate multidimensional scenarios with underlying drivers and associated narratives. The resulting scenarios were all distinct and uniquely challenging for potential policy interventions, as they represented qualitatively distinct alternative future dynamics across multiple decision-relevant objectives. We also introduced a number of novel analytical approaches for evaluating the generated scenario sets, including pairwise grid plots and parallel coordinate plots of the scenario input subspaces, upset plots of the model runs lying within the scenario subspaces, and reconstruction of cluster dynamics from those subspaces. This multimethod approach to cluster evaluation highlights the differences and similarities between the various scenarios, thus improving upon the state of the art in analyzing the scenarios generated with multiclass scenario discovery. These methodological innovations can strengthen scenario generation for high-quality public policy decision-making, including mitigation of safety and security risks connected to adverse developments of protest movements.

The consideration of more than one policy objective when generating scenarios for decision support more closely reflects real-world decision contexts, and makes the outputs of behavior-based scenario discovery more comparable to those of more conventional scenario generation methods. This also opens the door to multi-method scenario generation, in which scenario "supersets" are generated using a variety of methods to get the most diverse and comprehensive possible set. This may improve the robustness and usefulness of the resulting decisions. A key area of research in this regard is a more structured approach to generating integrated verbal narratives (and associated names) from (sets of) quantitative time series.

ACKNOWLEDGMENTS

The authors are indebted to the participants of their stakeholder workshops for their feedback and improvement suggestions, and to Guido Veldhuis, Kees van Dongen and three anonymous reviewers for suggesting improvements to this manuscript. This work was funded through the TNO research program "Kennisopbouw Politie, Politiedata & Intelligence" (Knowledge Development Police, Police Data & Intelligence), in collaboration with the Real Time Intelligence Lab.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All data and code are available upon request from the corresponding author.

ORCID

Patrick Steinmann http://orcid.org/0000-0002-6204-6253
Koen van der Zwet http://orcid.org/0000-0001-6395-7393

REFERENCES

Alsulami, A., Glukhov, A., Shishlenin, M., & Petrovskii, S. (2022). Dynamical modelling of street protests using the yellow vest movement and Khabarovsk as case studies. *Scientific Reports*, *12*(1), 20447.

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures*, *46*, 23–40.

Auping, W. L. (2018). *Modelling uncertainty: Developing and using simulation models for exploring the consequences of deep uncertainty in complex problems*. Delft University of Technology. https://doi.org/10.4233/uuid:0e0da51a-e2c9-4aa0-80cc-d930b685fc53

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations Research*, *41*(3), 435–449. https://doi.org/10.1287/opre.41.3.435

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bankes, S. (2011). The use of complexity for policy exploration. In P. Allen, S. Maguire, & B. McKelvey (Eds.), *The SAGE handbook of complexity and management* (pp. 570–589). Sage Publications.

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. https://doi.org/10.1016/j.futures.2005.01.003

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. https://doi.org/10.1016/j.techfore.2009.08.002

Crawford, M. M., & Wright, G. (2022). The value of mass-produced COVID-19 scenarios: A quality evaluation of development processes and scenario content. *Technological Forecasting and Social Change*, *183*, 121937.

Dalal, S., Han, B., Lempert, R., Jaycocks, A., & Hackbarth, A. (2013). Improving scenario discovery using orthogonal rotations. *Environmental Modelling & Software*, *48*, 49–64. https://doi.org/10.1016/j.envsoft.2013.05.013

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., & Morgan, J. (2022). Is seeking certainty in climate sensitivity measures counterproductive in the context of climate emergency? The case for scenario planning. *Technological Forecasting and Social Change*, *182*, 121811.

Dhami, M. K., Wicke, L., & Önkal, D. (2022). Scenario generation and scenario quality using the cone of plausibility. *Futures*, *142*, 102995.

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.

Filipic, B., & Tusar, T. (2018). A taxonomy of methods for visualizing Pareto front approximations. In H. Aguirre (Ed.), *GECCO '18: Proceedings of the genetic and evolutionary computation conference* (pp. 649–656). Association for Computing Machinery New York NY United States.

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, *9*(2), 123–143. https://doi.org/10.1023/a:1008894516817

Gadhave, K., Strobelt, H., Gehlenborg, N., & Lex, A. (2019). Upset 2: From prototype to tool. In A. Endert, B. Fisher, & W. Willett (Eds.), *Proceedings of the IEEE Information Visualization Conference–Posters (InfoVis' 19)*, 7. IEEE.

Gall, T., Vallet, F., & Yannou, B. (2022). How to visualise futures studies concepts: Revision of the futures cone. *Futures*, *143*, 103024.

Gallo, G. (2013). Conflict theory, complexity and systems approach. *Systems Research and Behavioral Science*, *30*(2), 156–175.

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

Godet, M. (2000). The art of scenarios and strategic planning. *Technological Forecasting and Social Change*, *65*(1), 3–22.

Goodwin, P., Gönül, M. S., & Önkal, D. (2019). When providing optimistic and pessimistic scenarios can be detrimental to judgmental demand forecasts and production decisions. *European Journal of Operational Research*, *273*(3), 992–1004.

Gotts, N. M., van Voorn, G. A. K., 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*(Part B), 100728. https://doi.org/10.1016/j.ecocom.2018.07.007

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), 9. https://doi.org/10.18564/jasss.3134

Guivarch, C., Lempert, R., & Trutnevyte, E. (2017). Scenario techniques for energy and environmental research: An overview of recent developments to broaden the capacity to deal with complexity and uncertainty. *Environmental Modelling & Software*, *97*, 201–210.

Guivarch, C., Rozenberg, J., & Schweizer, V. (2016). The diversity of socio-economic pathways and CO_2 emissions scenarios: Insights from the investigation of a scenarios database. *Environmental Modelling & Software*, *80*, 336–353. https://doi.org/10.1016/j.envsoft.2016.03.006

Gustafson, D. (2020). Hunger to violence: Explaining the violent escalation of nonviolent demonstrations. *Journal of Conflict Resolution*, *64*(6), 1121–1145.

Hakanen, J., Gold, D., Miettinen, K., & Reed, P. M. (2023). Visualisation for decision support in many-objective optimisation: State-of-the-art, guidance and future directions, In *Many criteria optimization and decision analysis: State-of-the-art, present challenges, and future perspectives* (pp. 181–212). Springer International Publishing.

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. https://doi.org/10.1016/j.futures.2015.09.004

Herman, J., & Usher, W. (2017). Salib: An open-source python library for sensitivity analysis. *The Journal of Open Source Software*, *2*(9), 97.

Hidayatno, A., Jafino, B. A., Setiawan, A. D., & Purwanto, W. W. (2020). When and why does transition fail? A model-based identification of adoption barriers and policy vulnerabilities for transition to natural gas vehicles. *Energy Policy*, *138*, 111239.

Hitch, C. J. (1960 *On the choice of objectives in systems studies*. Rand Corporation.

Iwanaga, T., Usher, W., & Herman, J. (2022). Toward salib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environmental Systems Modelling*, *4*, 18155.

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.

Kahagalage, S. D., Turan, H. H., Elsawah, S., & Gary, M. S. (2024). Exploratory modelling and analysis to support decision-making under deep uncertainty: A case study from defence resource planning and asset management. *Technological Forecasting and Social Change*, *200*, 123150.

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. https://doi.org/10.1016/j.envsoft.2012.007

Kunc, M. (2024). Integrating system dynamics and scenarios: A framework based on personal experience. *Futures & Foresight Science*, *6*, e174. https://doi.org/10.1002/ffo2.174

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. https://doi.org/10.1016/j.envsoft.2017.06.054

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. (2021). Reaction: A commentary on Lustick and Tetlock (2021). *Futures & Foresight Science*, *3*(2), e84.

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. https://doi.org/10.1016/j.techfore.2012.09.012

Kwakkel, J. H., & Haasnoot, M. (2019). Supporting DMDU: A taxonomy of approaches and tools. In V. Marchau, W. Walker, P. Bloemen, &

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

S. Popper (Eds.), *Decision Making Under Deep Uncertainty*. Springer. https://doi.org/10.1007/978-3-030-05252-2_15

Kwakkel, J. H., & Jaxa-Rozen, M. (2016). Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. *Environmental Modelling & Software*, *79*, 311–321. https://doi.org/10.1016/j.envsoft.2015.11.020

Kwakkel, J. H., Walker, W. E., & 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), 01816001. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000626

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. https://doi.org/10.1002/2017ef000701

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., Engelbrecht, F., Fischer, E., Fyfe, J., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S., & Zhou, T. (2021). Future global climate: Scenario-based projections and near-term information. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelek, R. Yu, & B. Zhou (Eds.), *Climate change 2021: The physical science basis. contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change* (pp. 553–672). Cambridge University Press. https://doi.org/10.1017/9781009157896.006

Lempert, R. J., Bryant, B. P., & Bankes, S. C. (2008). *Comparing algorithms for scenario discovery*. 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. https://doi.org/10.1287/mnsc.1050.0472

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 Corporation. https://www.rand.org/pubs/monograph_reports/MR1626.html

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). Upset: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1983–1992.

Litchfield, R. C., Fan, J., & Brown, V. R. (2011). Directing idea generation using brainstorming with specific novelty goals. *Motivation and Emotion*, *35*, 135–143.

Lord, S., Helfgott, A., & Vervoort, J. M. (2016). Choosing diverse sets of plausible scenarios in multidimensional exploratory futures techniques. *Futures*, *77*, 11–27.

Lustick, I. S., & Tetlock, P. E. (2021). The simulation manifesto: The limits of brute-force empiricism in geopolitical forecasting. *Futures & Foresight Science*, *3*(2), e64.

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? *Energy Economics*, *33*(4), 619–631. https://doi.org/10.1016/j.eneco.2010.10.007

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. https://doi.org/10.1080/00401706.1979.10489755

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Nowack, M., Endrikat, J., & Guenther, E. (2011). Review of Delphi-based scenario studies: Quality and design considerations. *Technological Forecasting and Social Change*, *78*(9), 1603–1615.

Osika, Z., Salazar, J. Z., Roijers, D. M., Oliehoek, F. A., & Murukannaiah, P. K. (2023). What lies beyond the Pareto front? a survey on decision-support methods for multi-objective optimization. arXiv preprint arXiv:2311.11288.

Parker, A. M., Srinivasan, S. V., Lempert, R. J., & Berry, S. H. (2015). Evaluating simulation-derived scenarios for effective decision support. *Technological Forecasting and Social Change*, *91*, 64–77. https://doi.org/10.1016/j.techfore.2014.01.010

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Popper, S. W. (2019). Robust decision making and scenario discovery in the absence of formalmodels. *Futures & Foresight Science*, *1*(3–4), e22.

Pot, W., Scherpenisse, J., & 't Hart, P. (2023). Robust governance for the long term and the heat of the moment: Temporal strategies for coping with dual crises. *Public Administration*, *101*(1), 221–235.

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. https://doi.org/10.1016/j.envsoft.2017.02.017

Rosenhead, J., Elton, M., & Gupta, S. K. (1972). Robustness and optimality as criteria for strategic decisions. *Journal of the Operational Research Society*, *23*, 413–431.

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. https://doi.org/10.1007/s10584-013-0904-3

Schoemaker, P. J. H. (1993). Multiple scenario development: Its conceptual and behavioral foundation. *Strategic Management Journal*, *14*(3), 193–213. https://doi.org/10.1002/smj.4250140304

Sluijs, T., Lokkers, L., Özsezen, S., Veldhuis, G. A., & Wortelboer, H. M. (2021). An innovative approach for decision-making on designing lifestyle programs to reduce type 2 diabetes on Dutch population level using dynamic simulations. *Frontiers in Public Health*, *9*, 652694.

Spaniol, M. J., & Rowland, N. J. (2019). Defining scenario. *Futures & Foresight Science*, *1*(1), e3. https://doi.org/10.1002/ffo2.3

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., Verstegen, J., van Voorn, G., Roman, S., & Ligtenberg, A. (n.d.). Scenario Search: Finding Diverse, Plausible and Comprehensive Scenario Sets for Complex Systems. Under revision.

Sterman, J. (2000). *Business dynamics: Systems thinking and modelling in a complex world*. McGraw-Hill Education.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, *10*(2–3), 291–330. https://doi.org/10.1002/sdr.4260100214

Sterman, J. D. (2002). All models are wrong: Reflections on becoming a systems scientist. *System Dynamics Review*, *18*(4), 501–531. https://doi.org/10.1002/sdr.261

Student, J., Kramer, M. R., & Steinmann, P. (2020a). Coasting: Model description, global sensitivity analysis, and scenario discovery. *MethodsX*, *7*, 101145.

Student, J., Kramer, M. R., & Steinmann, P. (2020b). Simulating emerging coastal tourism vulnerabilities: An agent-based modelling approach. *Annals of Tourism Research*, *85*, 103034.

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technishe Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Ruswurm, M., Kolar, K., & Woods, E. (2020). Tslearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research*, *21*(118), 1–6. http://jmlr.org/papers/v21/20-091.html

Veldhuis, G. A., de Reus, N. M., & Keijser, B. M. (2020). Concept development for comprehensive operations support with modeling and simulation. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, *17*(1), 99–116.

Veldhuis, G. A., Smits-Clijsen, E. M., & van Waas, R. P. M. (2023). Techniques to enhance the public policy impact of qualitative system dynamics models. *System Dynamics Review*. https://doi.org/10.1002/sdr.1758

Vermeulen, S. J., Challinor, A. J., Thornton, P. K., Campbell, B. M., Eriyagama, N., Vervoort, J. M., Kinyangi, J., Jarvis, A., Laderach, P., & Ramirez-Villegas, J. (2013). Addressing uncertainty in adaptation planning for agriculture. *Proceedings of the National Academy of Sciences United States of America 110*(21), 8357–8362.

Voros, J. (2003). A generic foresight process framework. *Foresight*, *5*(3), 10–21.

Weinans, E., van Voorn, G., Steinmann, P., Perrone, E., & Marandi, A. (2024). An exploration of drivers of opinion dynamics. *Journal of Artificial Societies and Social Simulation*, *27*(1).

Wiek, A., Keeler, L. W., Schweizer, V., & Lang, D. J. (2013). Plausibility indications in future scenarios. *International Journal of Foresight and Innovation Policy*, *9*(2–3–4), 133–147.

Wilkinson, A., & Eidinow, E. (2008). Evolving practices in environmental scenarios: A new scenario typology. *Environmental Research Letters*, *3*(4), 045017.

Wilkinson, A., Kupers, R., & Mangalagiu, D. (2013). How plausibility-based scenario practices are grappling with complexity to appreciate and address 21st century challenges. *Technological Forecasting and Social Change*, *80*(4), 699–710.

van der Zwet, K., Barros, A. I., van Engers, T. M., & Sloot, P. M. A. (2022). Emergence of protests during the COVID-19 pandemic: Quantitative models to explore the contributions of societal conditions. *Humanities and Social Sciences Communications*, *9*(1), 68.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Steinmann, P., van der Zwet, K., & Keijser, B. (2024). Simulation-based generation and analysis of multidimensional future scenarios with time series clustering. *Futures & Foresight Science*, *6*, e194. https://doi.org/10.1002/ffo2.194

25735152, 2024, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ffo2.194 by Technische Universiteit Delft, Wiley Online Library on [16/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License