Scenario search: finding diverse, plausible and comprehensive scenario sets for complex systems

Patrick Steinmann[1,3*], Judith Verstegen[2], George van Voorn[3], Sabin Roman[4], and Arend Ligtenberg[5]

[1]Faculty of Technology, Policy, and Management, Delft University of Technology, Delft, The Netherlands
[2]Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, Netherlands
[3]Biometris, Wageningen University & Research, Wageningen, The Netherlands
[4]Jožef Stefan Institute, Ljubljana, Slovenia
[5]Laboratory of Geo-information Science and Remote Sensing,
Wageningen University & Research, Wageningen, The Netherlands

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.

Keywords

Scenarios; many-objective optimisation; Exploratory Modelling & Analysis; scenario discovery; deep uncertainty

Code availability

All code used for performing the described exploration and optimization using the Python programming language (version 3.9) and various software packages can be found on GitHub under the URL https://github.com/steipatr/Scenario-Search. This repository was created by Patrick Steinmann in 2024 and includes data sets (70 KB) and analysis code (1.2 MB).

1. Introduction

Scenarios, or structured sets of plausible future narratives driven by external forces (Spaniol and 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

Correspondence:
Contact P. Steinmann at P.Steinmann@tudelft.nl

Cite this article as:
Steinmann, P., Verstegen, J., van Voorn, G., Roman, S., & Ligtenberg, A.
Scenario search: finding diverse, plausible and comprehensive scenario sets for complex systems
Socio-Environmental Systems Modelling, vol. 7, 18823, 2025, doi:10.18174/sesmo.18823

Socio-Environmental Systems Modelling
An Open-Access Scholarly Journal
http://www.sesmo.org

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 that 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, it may be that expert-driven approaches fail to identify some 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 (Lamontagne et al., 2018; Dolan et al., 2021). As McPhail et al. (2020) showed, the selection of scenarios for decision support can have a substantial impact on the quantitative outcomes of the subsequent decision. Thus, we identify a knowledge gap regarding how to generate 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 for the stated scenario generation purpose. Finally, we discuss some implications for scenario-based planning in particular and decision support in general.

2. Background

The Anthropocene is characterized by a wide variety of interdependent socio-technical environmental systems such as energy infrastructure, financial markets, and industrial agriculture. 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 and 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 relationships, and outcomes of interest underlying a decision-making context. These *deep uncertainties* (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 and 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; Rizzoli and Young, 1997) - 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 scenarios 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).

3. 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. We note here that, depending on the specific problem and decision context, the exact purpose of the scenario generation process may differ. In the following, we assume that the goal of the scenario generation process is to identify a small set of scenarios that summarizes a complex system's many plausible future states well, similar to the work done by Carlsen et al. (2016). In this sense, the scenario set forms an on-ramp for engaging with complexity in decision support processes (Wilkinson et al., 2013).

Diverse scenarios are meaningfully different alternatives to one another (Spaniol and Rowland, 2019), that is, they describe clearly distinguishable 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 and 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. Pruyt et al. (2018) highlighted how identifying small sets of diverse scenarios can be crucial, especially for rapid, interactive decision support in crisis situations.

At the same time, the presented alternative futures must be plausible, or within the scope of what could physically occur within the studied system - with no claim towards the probability of occurring (Wiek et al., 2013). In this sense, we follow the thinking of Wilson (1998) and Urueña (2019) - a future state is plausible if it is reasonable to believe that the state could happen, given certain initial conditions and causal relations of the system. 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 and 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 and Parker, 2014), as the models can systematically explore the implications of a large number of possible system configurations and assumptions (Bankes, 1993; Winsberg, 2010). As Wiek et al. (2013) point out, deeming a scenario to be plausible based purely on the theoretical evaluation of a model meets only the minimum threshold for plausibility. However, we believe this is more than offset by the explicit, testable nature of the simulation model underlying this evaluation, which also establishes the internal consistency of the scenario, as it results directly from the internal logic of the model. Consistency has been used as a measure of scenario plausibility by Lord et al. (2016), Tietje (2005), and Seeve and Vilkkumaa (2022), among others.

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, 2022, 2020; Zatarain Salazar et al., 2022). This ensures that the resulting analysis is robust to whichever future eventually ends up materializing (Rosenhead et al., 1972; Lempert et al., 2006).

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 (Davis et al., 2007; Chérel et al., 2015; Pruyt and Islam, 2015; Islam and Pruyt, 2016). 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 and Lempert, 2010; Kwakkel et al., 2013; Kwakkel, 2019; Stonedahl and Wilensky, 2011; Edali and Yücel, 2019; ten Broeke et al., 2021) or multiple outcomes of interest (Steinmann et al., 2020; Jafino and Kwakkel, 2021; Trindade et al., 2020; Student et al., 2020). In the present work, we build on ideas proposed by Verstegen et al. (2017b) to create a new method which straddles the two aforementioned lines of research – exploring the model's behavior space for extreme outcomes (similar to directed search; Halim et al., 2016), and then linking multiple of these extreme outcomes with their generative input parameter combinations to form maximally diverse, plausible, and comprehensive input-output scenario sets for complex systems.

As our proposed method relies heavily on the underlying simulation model to generate and analyze scenarios, choosing a suitable model is essential. To make the resulting scenarios trustworthy, the model should represent some real-world system to such a degree that all relevant parties consider it a valid representation thereof, and useful for the purpose of generating scenarios. To enable the optimization-based search for scenarios, the model should have independent input parameters, the ranges of which should be described in the model's documentation (Grimm et al., 2014, 2020). Crucially, these ranges should be such that every possible combination of input parameter values generates model behavior that is physically possible, even if it is unexpected or undesirable to stakeholders. Ideally, the model (and its associated input parameter ranges) would be built from the start with an exploratory modelling mindset (Auping, 2018), although more consolidative models (Bankes, 1993) can also be used with caution. We note here that a clear distinction must be made between the simulation model and the scenarios it may generate - the simulation model encodes knowledge about the studied system, while the scenarios are unique combinations of model inputs and outputs. We base this framing on work by Davis et al. (2007) on model-based scenario generation.

4. Methods

4.1 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 objective function that represents the relevant optimization criteria. In this case, the function is a simulation-based 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, as mentioned earlier.

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 preclude the unreachable ones), even though they are not known yet.

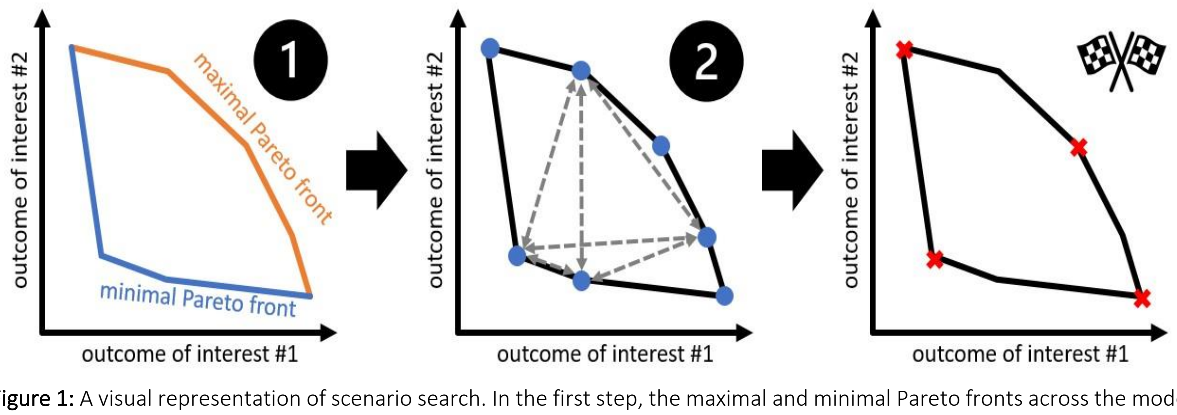

Figure 1: A visual representation of scenario search. In the first step, the maximal and minimal Pareto fronts across the model outcomes of interest are found through many-objective optimization. In the second step, all possible subsets of size *k* (here: *k* = 4) are generated from the points on the 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.

In the first step of scenario search (establishing comprehensiveness, see Figure 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. This optimization searches across the entire obtainable or feasible input parameter space to find parameter value combinations generating outcomes that maximize or minimize some outcomes of interest. We then join these individual fronts together to form the Pareto hull, which encompasses all plausible outcomes of the model. The Pareto hull is therefore the union of all points of all Pareto fronts. The model runs that constitute the Pareto hull dominate all other possible model runs regarding the outputs of interest, either positively or negatively. We note here that the term “optimisation”, which we use throughout this work, refers solely to the mathematical procedure, and not some desired system performance metric we wish to find.

In the second step (establishing diversity), we search the model outcomes on the Pareto hull for the most diverse subset of a desired size. Following previous work by Trutnevyte (2013) and Eker and Kwakkel (2018), 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 is a comparatively small subset of the entire data set, we can more easily calculate and compare the distances between the points. With this comparison, we identify the subset of most distant (i.e., diverse) model outputs, defined as the subset with the highest sum of intra-set pairwise distances. 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. Elsewhere, Lord et al. (2016) also advocated using between four and six scenarios to form a set, while Steinmann et al. (2024) used six and Kahagalage et al. (2024) used three scenarios, respectively.

4.2 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 groupings of cells (so-called patches) emerge across the grid. This combination of model simplicity and behavioral richness (Sun et al., 2016) makes Schelling’s segregation model an attractive case study for us.

The model is implemented in Python using the Mesa software package (Kazil et al., 2020; ter Hoeven et al., 2025). The square grid upon which the individual agents move is toroidal, and its size is 30 by 30 grid squares. The two groups of agents are randomly seeded in equal numbers at the start of each model run, with the remaining grid squares staying empty (but available as move destinations). The model runs for 100 time steps, although it often converges to a steady state earlier, depending on the specific input parameter combinations. For the parameter sweep and many-objective optimization, we specify the input space as [0.05,0.95] for *density* and [3,8] for *homophily*. 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 patches (contiguous regions of same-class neighboring cells) have emerged. These are the two objectives that 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. For the parameter sweep, we use 3000 function evaluations.

We perform the many-objective optimization using the ϵ-NSGA-II optimization algorithm (Kollat and Reed, 2006) implemented in the Platypus library (Hadka, 2015) for Python and controlled through the Exploratory Modelling and Analysis Workbench (Kwakkel, 2017). Based on testing for convergence using the hypervolume and epsilon progress metrics in the Platypus library, 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.

4.3 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 generally start by identifying external drivers of change. For these driving factors, high and low levels are determined. Across the factors' values, 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. In some more advanced matrix-based methods, such as Intuitive Logics (Wright et al., 2013), the number of drivers may be first reduced to only the most impactful ones through clustering, influence diagrams, and/or ranking methods. To generate our matrix-based scenarios, we follow the Massive Scenario Generation approach proposed by Davis et al. (2007) by sampling the corners of the model input space, representing the high(est) and low(est) levels for every axis. These corner points are then passed into the simulation model and paired with their resulting outputs. Scenario matrices thus reason from the drivers to the narratives, or, in a modelling sense, from the input to the output space.

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 that might create that world can then be 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 that 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 (Steinmann et al., 2024; Jafino and Kwakkel, 2021; Kwakkel et al., 2013; Rozenberg et al., 2014) 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 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. To establish the diversity of an entire scenario set, we calculate the pairwise distance between every pair of scenarios in the set. In the presented case study, this gives six distances per scenario set of four scenarios, with larger values being more desirable, as they indicate higher diversity.

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

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. This is based on the underlying notion of plausibility introduced previously - a future state is probable if it can be generated by the model. The further away a future state is from what the model can actually generate, the less plausible it is. This metric, therefore, punishes scenario generation methods that create impossible or fanciful scenarios according to the model. To facilitate the analysis, we construct the metric given in Equation 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. For our case study, this results in four distances per scenario set, one for each scenario in the set.

$$P(S_1) = \frac{1}{min_{B \in R} D(S_1, B) + 1} \tag{2}$$

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 3, we give the area calculation for a scenario set S with clockwise-ordered points in a two-dimensional Euclidean space. Note that for higher-dimensional scenarios (i.e., decision contexts with three or more relevant objectives), the calculation of the (hyper)volumes enclosed by the Pareto hull and scenario set may be more complicated due to their irregularity. Decomposition into simpler volumes may be useful in this regard. This calculation is only performed once, as it is performed on the level of the scenario set, not the individual constituent scenarios.

$$A(S) = \frac{1}{2} \sum_{i=1}^{|S|} S_{i,x}(S_{i+1,y} - S_{i-1,y}) \tag{3}$$

5. 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.

5.1 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 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 patches, 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 that 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.

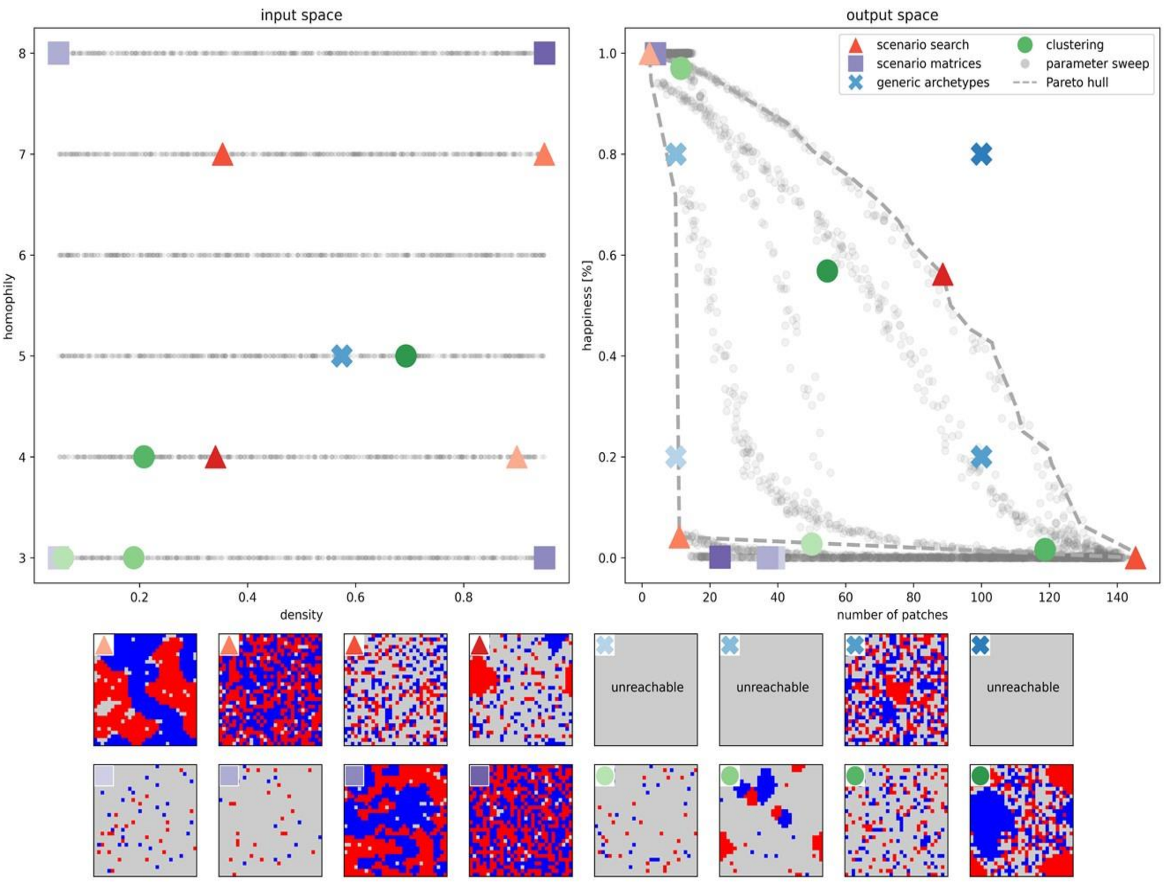

Figure 2: Input and output spaces of Schelling's segregation model. In each space, we show four sets of four scenarios each, one set per evaluated scenario generation method. The different methods are double-coded by color and marker shape, with the color hue distinguishing the four scenarios within 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.

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 showing medium-high density and regionally varying granularity.

5.2 Scenario Criteria

5.2.1 Diversity

The most diverse scenarios are created by scenario search (Figure 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 the fourth is far away. The 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.

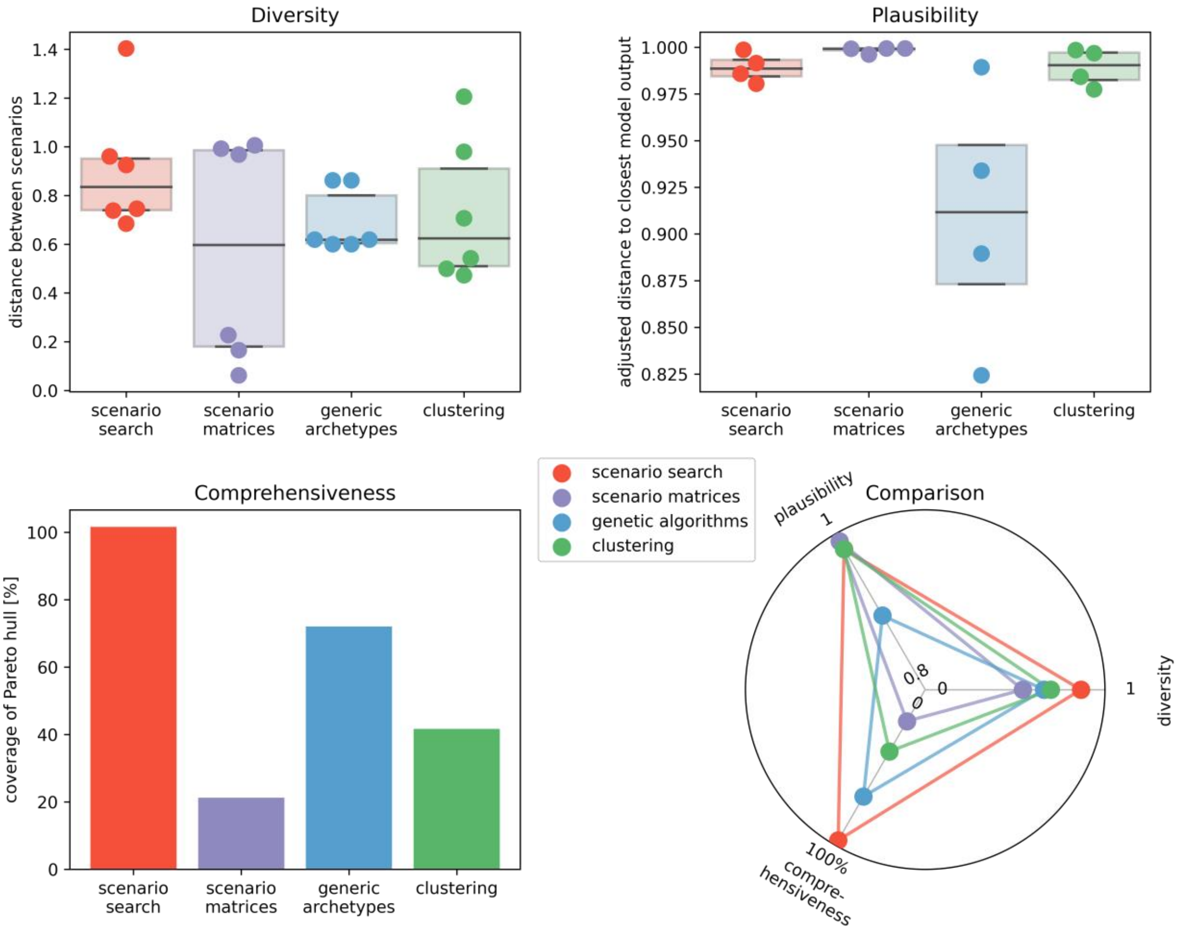

Figure 3: Evaluation of all four scenario generation methods against the three scenario set criteria. Where applicable, means and quartiles are represented with underlying box plots. The point markers are jittered to avoid overlap. The radar chart shows the criteria as polar axes in one figure, allowing overall comparison between the four different scenario generation methods.

The 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.

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.

5.2.2 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 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 that, 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 that 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.

5.2.3 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 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%.

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".

5.3 Comparison

When evaluating the four scenario generation methods across all three criteria (see radar chart in Figure 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.

Table 1: Rankings of the four scenario generation methods across the three scenario set criteria, with overall rankings computed using both multiplicative and additive scoring.

	diversity	plausibility	comprehensiveness	multiplicative 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)

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 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. Discussion

6.1 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 that 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, for our case study, 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 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 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 et al., 2017a). Furthermore, it is desirable to base such an analysis on future scenarios that could actually materialize. However, at least one, and potentially two, of the studied scenario generation methods produced scenarios that 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.

6.2 Equifinality

A key consideration in the analysis of any model's input-output mapping is equifinality, or the generation of identical model outputs by distinct inputs (Von Bertalanffy and Sutherland, 1974). In our analysis, we observed that the scenario matrices method exhibited strong equifinality, in that three of the four input parameter combinations generated almost identical outputs despite representing different corners of the input space. This highlights not only that the studied system is nonlinear and complex, but that this specific approach to generating scenario sets fails to account for this complexity.

Our proposed method also does not explicitly account for the possibility of equifinality. Because we search for scenarios that are far apart in the output space, it is not possible to identify equifinal scenarios, or scenarios that are similar but generated by different external drivers. However, these may be of interest to stakeholders and decision makers. Therefore, it might be appealing to use concurrent distance measurement in the input and output space, as proposed by Jafino and Kwakkel (2021), to include equifinality considerations in scenario search.

6.3 Model-based vs. model-free scenario generation

In our presented analysis, we found that scenario generation using an underlying simulation model produces scenario sets that are generally more diverse, plausible, and comprehensive than model-free methods. We believe this is because the simulation model is a formal, explicit representation of the knowledge, assumptions, and uncertainties about the studied system, the interactions and implications of which can be evaluated. This is beneficial in regard to all three stated desirable properties for scenario sets. However, there are also advantages to model-free approaches, which may be insufficiently captured by our criteria. By including more degrees of freedom, the range of future system states thought to be plausible may be larger, which we also saw evidence of in our work – one of the scenarios generated with the scenario matrices method lay far outside the scope of what could be considered plausible based on the simulation model. Whether this is desirable or not may depend on the specific decision context and how plausibility is understood. It might also be beneficial to combine multiple different scenario generation methods to build a “super set” of scenarios, which may be even more diverse and comprehensive than any of the individual methods.

6.4 Limitations

A number of limitations apply to our research. We applied scenario search and the other three scenario generation methods to just one case study, the Schelling model. We therefore cannot make any statements about the generalizability of our method. Future applications of scenario search across a wider range of complex system models and/or scenario generation processes may (in)validate our ideas. Conceptually, scenario search requires that a simulation model 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 and 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. 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 plausible scenarios.

7. Conclusions

Scenario-based decision-making relies on sets of scenarios that are diverse, plausible, and comprehensive. In this paper, we presented a novel approach for generating such scenario sets that 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, these scenarios be critically reviewed regarding their diversity, plausibility, and comprehensiveness.

Author CRediT statement

Patrick Steinmann Conceptualization, Funding acquisition, Methodology, Investigation, Writing – original draft
Judith Verstegen Conceptualization, Funding acquisition, Methodology, Investigation, Supervision, Writing – review and editing
George van Voorn Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review and editing
Sabin Roman Methodology, Supervision, Writing – review and editing
Arend Ligtenberg Funding acquisition, Methodology, Supervision, Writing – review and editing

Funding

This work was supported by Wageningen University & Research's graduate school for Production Ecology and Resource Conservation as well as the Swiss Existential Risk Initiative (now Pivotal Research).

Acknowledgements

We thank Nora Ammann, Manon Gouiran, and Silvana Hultsch for reviewing early manuscript drafts.

References

Auping, W. L. (2018). Modelling uncertainty: Developing and using simulation models for exploring the consequences of deep uncertainty in complex problems.

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

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

Bezold, C. (2009). Aspirational futures: Understanding threats, opportunities and visionary possibilities. Journal of Futures Studies, 13(4), 81–90.

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

Carlsen, H., Lempert, R., Wikman-Svahn, P., & Schweizer, V. (2016). Choosing small sets of policy-relevant scenarios by combining vulnerability and diversity approaches. Environmental Modelling & Software, 84, 155–164.

Chérel, G., Cottineau, C., & Reuillon, R. (2015). Beyond corroboration: Strengthening model validation by looking for unexpected patterns. PLOS ONE, 10(9), e0138212. https://doi.org/10.1371/journal.pone.0138212

Cook, M. (2004). Universality in elementary cellular automata. Complex Systems, 40.

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 (Technical Report No. TR-392). RAND National Security Research Division.

de Regt, H. W., & Parker, W. S. (2014). Introduction: Simulation, visualization, and scientific understanding. Perspectives on Science, 22(3), 311–317. https://doi.org/10.1162/POSC_e_00135

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. https://doi.org/10.1177/0308518X19877885

Derbyshire, J. (2022). Increasing preparedness for extreme events using plausibility-based scenario planning: Lessons from COVID-19. Risk Analysis, 42(1), 97–104. https://doi.org/10.1111/risa.13827

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. https://doi.org/10.1038/s41467-021-22194-0

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. https://doi.org/10.1016/j.simpat.2018.12.006

Eker, S., & Kwakkel, J. H. (2018). Including robustness considerations in the search phase of many-objective robust decision making. Environmental Modelling & Software, 105, 201–216.

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. Futures, 25(7), 739–755. https://doi.org/10.1016/0016-3287(93)90022-L

Gotts, N. M., van Voorn, G. A., Polhill, J. G., de Jong, E., Edmonds, B., Hofstede, G. J., & Meyer, R. (2019). Agent-based modelling of socio-ecological systems: Models, projects and ontologies. Ecological Complexity, 40, 100728. https://doi.org/10.1016/j.ecocom.2018.07.007

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S., Liu, C., Martin, B. T., Meli, M., Radchuk, V., et al. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. Ecological Modelling, 280, 129–139.

Grimm, V., Railsback, S. F., Vincenot, C. E., Berger, U., Gallagher, C., DeAngelis, D. L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., et al. (2020). The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. Journal of Artificial Societies and Social Simulation, 23(2).

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. https://doi.org/10.1016/j.gloenvcha.2012.12.006

Hadka, D. (2015). Platypus. https://github.com/Project-Platypus/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.

Helbing, D. (2013). Globally networked risks and how to respond. Nature, 497(7447), 51–59. https://doi.org/10.1038/nature12047

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. https://doi.org/10.1086/711501

Islam, T., & Pruyt, E. (2016). Scenario generation using adaptive sampling: The case of resource scarcity. Environmental Modelling & Software, 79, 285–299. https://doi.org/10.1016/j.envsoft.2015.09.014

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. SocioEnvironmental 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. https://doi.org/10.1016/j.envsoft.2021.105177

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. https://doi.org/10.1002/wcc.721

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.

Kazil, J., Masad, D., & Crooks, A. (2020). Utilizing Python for agent-based modeling: The Mesa framework. In Social, Cultural, and Behavioral Modeling: 13th International Conference, SBP-BRiMS 2020, Washington, DC, USA, October 18–21, 2020, Proceedings 13 (pp. 308–317). Springer.

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. https://doi.org/10.1016/j.advwatres.2005.07.010

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. https://doi.org/10.1002/ffo2.8

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

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

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.

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

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics (pp. 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. https://doi.org/10.1016/j.envsoft.2016.03.014

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. https://doi.org/10.1016/j.envsoft.2018.11.018

McPhail, C., Maier, H. R., Westra, S., Kwakkel, J. H., & Linden, L. (2020). Impact of scenario selection on robustness. Water Resources Research, 56(9). https://doi.org/10.1029/2019WR026515

Nakicenovic, 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.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. Science, 263(5147), 641–646. https://doi.org/10.1126/science.263.5147.641

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. https://doi.org/10.1111/padm.12872

Pruyt, E., & Islam, T. (2015). On generating and exploring the behavior space of complex models. System Dynamics Review, 31(4), 220–249. https://doi.org/10.1002/sdr.1544

Pruyt, E., Islam, T., & Arzt, T. (2018). On the spot and map: Interactive model-based policy support under deep uncertainty. In Policy Analytics, Modelling, and Informatics: Innovative Tools for Solving Complex Social Problems (pp. 315–342).

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

Rizzoli, A. E., & Young, W. J. (1997). Delivering environmental decision support systems: Software tools and techniques. Environmental Modelling & Software, 12(2–3), 237–249.

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. https://doi.org/10.1057/jors.1972.72

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. https://doi.org/10.1073/pnas.0711295105

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

Schelling, T. C. (1971). Dynamic models of segregation. The Journal of Mathematical Sociology, 1(2), 143–186. https://doi.org/10.1080/0022250X.1971.9989794

Seeve, T., & Vilkkumaa, E. (2022). Identifying and visualizing a diverse set of plausible scenarios for strategic planning. European Journal of Operational Research, 298(2), 596–610.

Siegenfeld, A. F., & Bar-Yam, Y. (2020). An introduction to complex systems science and its applications. Complexity, 1–16. https://doi.org/10.1155/2020/6105872

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. https://doi.org/10.1016/S0140-6736(21)00424-4

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. https://doi.org/10.1016/j.techfore.2020.120052

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(4), e194.

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

Stonedahl, F., & Wilensky, U. (2011). Finding forms of flocking: Evolutionary search in ABM parameter-spaces. In T. Bosse, A. Geller, & C. M. Jonker (Eds.), MultiAgent-Based Simulation XI (Vol. 6532, pp. 61–75). Springer. https://doi.org/10.1007/978-3-642-18345-4_5

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

Sun, Z., Lorscheid, I., Millington, J. D., Lauf, S., Magliocca, N. R., Groeneveld, J., Balbi, S., Nolzen, H., Müller, B., Schulze, J., et al. (2016). Simple or complicated agent-based models? A complicated issue. Environmental Modelling & Software, 86, 56–67.

ten Broeke, G., & Tobi, H. (2021). Mapping validity and validation in modelling for interdisciplinary research. Quality & Quantity, 55(5), 1613–1630. https://doi.org/10.1007/s11135-020-01073-8

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.

ter Hoeven, E., Kwakkel, J., Hess, V., Pike, T., Wang, B., Kazil, J., et al. (2025). Mesa 3: Agent-based modeling with Python in 2025. Journal of Open Source Software, 10(107), 7668.

Thompson, E. L., & Smith, L. A. (2019). Escape from model-land. Economics, 13(1), 20190040. https://doi.org/10.5018/economics-ejournal.ja.2019-40

Tietje, O. (2005). Identification of a small reliable and efficient set of consistent scenarios. European Journal of Operational Research, 162(2), 418–432.

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. https://doi.org/10.1016/j.envsoft.2020.104772

Trutnevyte, E. (2013). Expanse methodology for evaluating the economic potential of renewable energy from an energy mix perspective. Applied Energy, 111, 593–601.

Uruena, S. (2019). Understanding “plausibility”: A relational approach to the anticipatory heuristics of future scenarios. Futures, 111, 15–25.

Verstegen, J. A., Jonker, J. G. G., Karssenberg, D., van der Hilst, F., Schmitz, O., de Jong, S. M., & Faaij, A. P. C. (2017a). 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. (2017b). Locating the position of a scenario projection in solution space. In The 20th AGILE Conference on Geographic Information Science, Wageningen, NL.

Von Bertalanffy, L., & Sutherland, J. W. (1974). General systems theory: Foundations, developments, applications. IEEE Transactions on Systems, Man, and Cybernetics, (6), 592–592.

Voros, J. (2017). Big history and anticipation. In R. Poli (Ed.), Handbook of Anticipation (pp. 1–40). Springer. https://doi.org/10.1007/978-3-319-31737-3_95-1

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. https://doi.org/10.1002/1099-1360(200001/05)9:1/3<11::AID-MCDA264>3.0.CO;2-3

Wiek, A., Withycombe Keeler, L., 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., 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.

Wilson, I. (1998). Mental maps of the future: An intuitive logics approach to scenarios. In Learning from the future: Competitive foresight scenarios (pp. 81–108).

Winsberg, E. B. (2010). Science in the age of computer simulation. The University of Chicago Press.

Wolters, H., van den Born, G. J., Dammers, E., & Reinhard, S. (2018). Deltascenario’s voor de 21e eeuw, actualisering 2017 (Technical report). 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. https://doi.org/10.1016/j.techfore.2012.09.003

Zatarain Salazar, J., Castelletti, A., & Giuliani, M. (2022). Multi-objective robust planning tools. In Oxford Research Encyclopedia of Environmental Science. Oxford University Press. https://doi.org/10.1093/acrefore/9780199389414.013.626