



Delft University of Technology

## Improving scenario discovery by bagging random boxes

Kwakkel, J. H.; Cunningham, S. C.

### DOI

[10.1016/j.techfore.2016.06.014](https://doi.org/10.1016/j.techfore.2016.06.014)

### Publication date

2016

### Document Version

Final published version

### Published in

Technological Forecasting and Social Change

### Citation (APA)

Kwakkel, J. H., & Cunningham, S. C. (2016). Improving scenario discovery by bagging random boxes. *Technological Forecasting and Social Change*, 111, 124-134. <https://doi.org/10.1016/j.techfore.2016.06.014>

### 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.



## Improving scenario discovery by bagging random boxes



J.H. Kwakkel <sup>\*</sup>, S.C. Cunningham

*Delft University of Technology, Faculty of Technology, Policy and Management, Delft, The Netherlands*

### ARTICLE INFO

#### Article history:

Received 24 January 2015

Received in revised form 10 June 2016

Accepted 12 June 2016

Available online 30 June 2016

#### Keywords:

Scenario discovery  
Robust decision making  
Exploratory modeling  
Deep uncertainty

### ABSTRACT

Scenario discovery is a model-based approach to scenario development under deep uncertainty. Scenario discovery relies on the use of statistical machine learning algorithms. The most frequently used algorithm is the Patient Rule Induction Method (PRIM). This algorithm identifies regions in an uncertain model input space that are highly predictive of model outcomes that are of interest. To identify these regions, PRIM uses a hill-climbing optimization procedure. This suggests that PRIM can suffer from the usual defects of hill climbing optimization algorithms, including local optima, plateaus, and ridges and valleys. In case of PRIM, these problems are even more pronounced when dealing with heterogeneously typed data. Drawing inspiration from machine learning research on random forests, we present an improved version of PRIM. This improved version is based on the idea of performing multiple PRIM analyses based on randomly selected features and combining these results using a bagging technique. The efficacy of the approach is demonstrated using three cases. Each of the cases has been published before and used PRIM. We compare the results found using PRIM with the results found using the improved version of PRIM. We find that the improved version is more robust to new data, can better cope with heterogeneously typed data, and is less prone to overfitting.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### 1. Introduction

Scenario discovery (Bryant and Lempert, 2010) is an approach for addressing the challenges of characterizing and communicating deep uncertainty associated with simulation models (Dalal et al., 2013). Deep uncertainty is encountered when the different parties to a decision do not know or cannot agree on the system model that relates actions to consequences, the exogenous inputs to the system model (Lempert et al., 2003). Decision problems under deep uncertainty often involve decisions that are made over time in dynamic interaction with the system (Hallegatte et al., 2012). When confronted by deep uncertainty, it is possible to enumerate the possibilities (e.g. sets of model inputs, alternative relationships inside a model), without ranking these possibilities in terms of perceived likelihood or assigning probabilities to them (Kwakkel et al., 2010). Scenario discovery addresses the challenge posed by deep uncertainty by exploring the consequences of the various deep uncertainties associated with a simulation model through conducting a series of computational experiments (Bankes et al., 2013). The resulting data set is subsequently analyzed using statistical machine learning algorithms in order to identify regions in the uncertainty space that are of interest (Bryant and Lempert, 2010, Kwakkel et al., 2013). These identified regions, which are typically characterized by only a small subset of the deeply uncertain factors, can subsequently

be communicated to the actors involved in the decision problem. Preliminary experiments with real world decision makers suggest that scenario discovery results are decision relevant and easier to interpret for decision makers than probabilistic ways of conveying the same information (Parker et al., 2015).

Currently, the main statistical rule induction algorithm used for scenario discovery is the Patient Rule Induction Method (PRIM) (Friedman and Fisher, 1999). PRIM can be used for data analytic questions, where the analyst tries to find combinations of values for input variables that result in similar characteristic values for the outcome variables. Specifically, PRIM identifies one or more of hyper-rectangular subspaces of the model input space within which the values of a single output variable are considerably different from its average values over the entire model input space. These subspaces are described as hyper-rectangular boxes of the model input space. To identify these boxes, PRIM uses a non-greedy, or patient, and hill climbing optimization procedure.

There are two key concerns when using PRIM for scenario discovery. The first concern is the interpretability of the results. Ideally the subspaces identified through PRIM should be composed of only a small subset of the uncertainties considered. If the number of uncertainties that jointly define the subspace is too large, interpretation of the results becomes challenging for the analyst (Bryant and Lempert, 2010). But, perhaps even more importantly, communicating such results to the stakeholders involved in the process becomes substantially more challenging (Parker et al., 2015). The second concern is that the uncertainties in the subset should be significant. That is, PRIM should

<sup>\*</sup> Corresponding author.

E-mail address: [j.h.kwakkel@tudelft.nl](mailto:j.h.kwakkel@tudelft.nl) (J.H. Kwakkel).

only include uncertain factors in the definition of a subspace that are truly predictive for the characteristic values of the outcome variable. This concern is particularly important given that PRIM uses a lenient hill climbing optimization procedure for finding the subspaces. As such, PRIM suffers from the usual defects associated with hill climbing. The main defect is that hill climbing can only find a local optimum. Moreover, PRIM can get stuck on a plateau where the performance does not change resulting in an early stop of the optimization. PRIM can also get stuck by ridges and valleys which prevent the hill climbing algorithm from further improving the performance. Together, these defects imply that there might exist boxes that offer a better description of the data, but which cannot be found by the hill climbing optimization algorithm.

In current scenario discovery practice, the interpretability concern is addressed primarily by performing PRIM in an interactive manner. By keeping track of the route followed by the lenient hill climbing optimization procedure used in PRIM, the so-called peeling trajectory, a manual inspection can reveal how the number of uncertainties that define the subspace varies as a function of density (precision) and coverage (recall). This allows for making a judgment call by the analyst balancing interpretability, coverage, and density. To avoid the inclusion of spurious uncertainties in the subset, Bryant and Lempert (2010) propose a resampling procedure and a quasi-p-values test. This resampling test assesses how often essentially the same subspace is found by running PRIM on randomly selected subsets of the data. The quasi-p-value test, essentially a one sided binomial test, is an estimate of the likelihood that a given uncertainty is included in the definition of the subspace purely by chance.

In this paper, we investigate an alternative approach that addresses the interpretability concern and the significance concern simultaneously. This alternative approach is inspired by the extensive work that has been done with Classification and Regression Trees (CART) (Breiman et al., 1984) and related classification tree algorithms. The basic idea behind this alternative is to perform multiple runs of the PRIM algorithm based on randomly selected features (Breiman, 2001) and combining these results using a bagging technique (Breiman, 1996). The resulting algorithm is known as random forest (Breiman, 2001). The idea of random feature selection is that all the data is used, but rather than including all uncertainties as candidate dimensions, only a randomly selected subset is used. So, instead of repeatedly running PRIM on randomly selected data as currently done in the resampling procedure suggested by Bryant and Lempert (2010), this random feature selection procedure randomly selects the uncertainties instead. Bagging is an established approach in machine learning for combining multiple versions of a predictor into an aggregate predictor (Breiman, 1996). The expectation is that this random boxes approach will outperform normal PRIM, analogous to how a random forest outperforms a single classification tree.

To demonstrate the proposed approach and assess its efficacy compared to the normal use of PRIM in the context of scenario discovery, we apply it to three cases. In particular, we apply it to the same data as used in the paper of Bryant and Lempert (2010) in which Scenario Discovery was first proposed, the case study of Rozenberg et al. (2013), and the case used by Hamarat et al. (2014). The first case covers continuous uncertain factors, the second case covers discrete uncertain factors, and the third case has continuous, discrete, and categorical uncertain factors. This allows for a comparison between the original algorithm and the proposed approach across cases with differently typed uncertain factors.

The remainder of this paper is structured accordingly. In Section 2, we present a review of the scenario discovery literature. In Section 3, we outline the method in more detail. More specifically, we introduce PRIM in Section 3.1, random forests in Section 3.2, and the combined approach in Section 3.3. Section 4 contains the results. We discuss the results in Section 5. Section 6 contains the conclusions.

## 2. Prior research

Scenario discovery was first put forward by Bryant and Lempert (2010). Their work builds on earlier work on the use of PRIM and CART in the context of Robust Decision Making (Lempert et al., 2006; Groves and Lempert, 2007; Lempert et al., 2008). Scenario discovery forms the analytical core of Robust Decision Making (Walker et al., 2013). Many examples of the use of scenario discovery in the context of Robust Decision Making can be found in the literature (Lempert et al., 2006; Lempert and Collins, 2007; Dalal et al., 2013; Hamarat et al., 2013; Matrosov et al., 2013a, 2013b; Auping et al., 2015; Eker and van Daalen, 2015). Robust Decision Making aims at supporting the design of policies that perform satisfactory across a very large ensemble of future states of the world. In this context, scenario discovery is used to identify the combination of uncertainties under which a candidate policy performs poorly, allowing for their iterative improvement. The use of scenario discovery for Robust Decision Making suggests that it could also be used in other planning approaches that design plans based on an analysis of the conditions under which a plan fails to meet its goals (Walker et al., 2013). Specially, Kwakkel et al. (2015) and Kwakkel et al. (2016) suggest that the vulnerabilities identified through scenario discovery can be understood as a multi-dimensional generalization of adaptation tipping points (Kwadijk et al., 2010), which are a core concept in the literature on dynamic adaptive policy pathways (Haasnoot et al., 2013).

Increasingly, scenario discovery is used more general as a bottom up model-based approach to scenario development. (Gerst et al., 2013; Kwakkel et al., 2013; Rozenberg et al., 2013; Halim et al., 2015; Greeven et al., 2016). There exists a plethora of scenario definitions, typologies, and methodologies (Bradfield et al., 2005; Börjeson et al., 2006). Broadly, three schools can be distinguished: the *La Prospective* school developed in France; the Probabilistic Modified Trends school originating at RAND; and the intuitive logic school typically associated with the work of Shell (Bradfield et al., 2005; Amer et al., 2013). Scenario discovery can be understood as a model-based approach to scenario development belonging to the intuitive logic school (Bryant and Lempert, 2010).

Scenario discovery aims to address several shortcomings of other scenario approaches. First, the available literature on evaluating scenario studies has found that scenario development is difficult if the involved actors have diverging interests and worldviews (van 't Klooster and van Asselt, 2006; European Environmental Agency, 2009; Bryant and Lempert, 2010). Rather than trying to achieve consensus, facilitate a process of joint sense-making to resolve the differences between worldviews, or arbitrarily imposing one particular worldview, scenario discovery aims at making transparent which uncertain factors actually make a difference for the decision problem at hand. An illustration of this is offered by Kwakkel et al. (2013) who capture two distinct mental models of how copper demands emerges in two distinct System Dynamics models and apply scenario discovery to both models simultaneously. Similarly, Pruyt and Kwakkel (2014) apply scenario discovery to three models of radicalization processes, which encapsulates three distinct mental models of how home grown terrorists emerge.

Another shortcoming identified in the evaluative literature is that scenario development processes have a tendency to overlook surprising developments and discontinuities (van Notten et al., 2005; Goodwin and Wright, 2010; Derbyshire and Wright, 2014). This might be at least partly due to the fact that many intuitive logic approaches move from a large set of relevant uncertain factors to a smaller set of drivers or megatrends. The highly uncertain and high impact drivers form the scenario logic. In this dimensionality reduction, interesting plausible combinations of uncertain developments are lost. In contrast, scenario discovery first systematically explores the consequences of all the relevant factors, and only then performs a dimensionality reduction in light of the resulting outcomes – thus potentially identifying surprising

results that would have been missed with traditional intuitive logic approaches.

There is an emerging interest in the identification of exemplar scenarios from the large database of simulation model results. Scenario discovery aims at summarizing the combination of uncertain factors that jointly explain a particular type or class of outcome of interest. This information can be of great relevance to decision-making. However, there are situations where one wants to investigate a specific simulation result in more detail. Lord et al. (2016) put forward an approach to assist in this scenario selection problem that can complement scenario discovery. Similarly, Halim et al. (2015) put forward an optimization based worst-case scenario discovery approach. Comes et al. (2015) discuss the scenario selection task in the context of time sensitive decisions. Trutnevvyte et al. (2016) highlight the importance of scenario selection in the context of scenario discovery and call for more research in this direction.

Another emerging topic is adaptive sampling, which was first suggested by Lempert et al. (2008) as being of potential relevance to scenario discovery. A first step towards investigating this is offered by Pruyt and Islam (2016) and Islam and Pruyt (2015) who explore the use of adaptive sampling to improve the diversity of types of model behavior available in the dataset.

Various methodological developments have taken place since scenario discovery was first put forward. Dalal et al. (2013) suggest preprocessing the input to scenario discovery using Principal Components Analysis (PCA). This can help in identifying succinct descriptions of the cases of interest if they are dependent on specific correlations between uncertain factors. They discuss in depth issues surrounding interpretability in case of PCA preprocessing. Kwakkel et al. (2013) apply this PCA-PRIM technique to a dataset of time series generated by a System Dynamics simulation model of the copper system, but struggle with the interpretability of the results. Kwakkel and Jaxa-Rozen (2016) investigate various alternative objective functions to better cope with heterogeneously typed input data and multivariate outcome variables. Guivarch et al. (2016) investigate a modification to PRIM pertaining how it behaves in case an analyst is trying to find multiple boxes in the same data set. Parker et al. (2015) offers a first investigation of the interpretability of scenario discovery results by decision makers and find that decision makers are able to use scenario discovery results, even if communicated purely quantitatively, for arriving at better decisions that when given the same information in a probabilistic manner. Hallegatte et al. (2015) use f-PRIM, an extension of PRIM, to perform automated scenario discovery.

In all methodological developments of scenario discovery in general and with respect to PRIM in particular, however, the basic approach for dealing with interpretability and accurate identification of uncertain factors has not been modified. That is, all these studies rely on the analysts making a tradeoff between coverage, density, and number of restricted uncertain factors, supported by resampling and quasi-p values. This paper, in contrast, explicitly puts forward an alternative approach for handling the interpretability concern and the significance concern.

### 3. Method

In this section, we first introduce PRIM and random forest, followed by an outline of how we combine these two into a more sophisticated version of PRIM based on random feature selection and bagging.

#### 3.1. Prim

Before offering a detailed mathematical exposition of the algorithm, we first offer a high level visual outline of the algorithm. Fig. 1 offers this visual explanation. The aim of PRIM is to find a rectangular box that has a high concentration of points of interest (denoted in red). We start with a box that contains all the data points (top left axes in Fig. 1).

Next, we consider removing a small slice of data along the top and bottom, and left and right ((the grey shaded areas in axes in the second and third rows in Fig. 1). This gives four candidate boxes. PRIM will select the one that results in the most increase on the objective function, which is typically the mean of the data remaining. In this particular example, removing along the top results in a higher concentration of red points, so removing from the top is a better choice. Likewise, removing from the top is better than either removing from the left or right. This results in a new box  $B_{l+1}$  (shown in the bottom row on the left). Now the procedure is repeated until a user specified stopping condition is met.

In the mathematical description of PRIM below, we follow the exposition as given by Friedman and Fisher (1999). Given a learning set  $\mathcal{L} = \{x_i, y_i\}_1^N$  where  $y$  is some output variable, and  $S_j$  is the set of all possible values for the input variable  $x_j$ .

$$\{x_i \in S_j\}_{j=1}^n \quad (1)$$

$S_j$  could be real values, discrete values, or categorical values. So, the entire input domain  $S$  can be represented by the  $n$ -dimensional product space

$$S = S_1 \times S_2 \times \dots \times S_n \quad (2)$$

The goal of PRIM is to find a subset  $R$  of the input domain  $S$ , so  $R \subset S$ , for which the average value  $\bar{Y}$  is substantially different from the average over the entire domain. For interpretability, one would like to specify the subset  $R$  with simple logical conditions, or rules, based on the values of the individual input variables  $\{x_j\}_1^n$ . That is, the subset  $R$  is the union of a set of simple sub regions  $\{B_k\}_1^K$ .  $B_k$  is a box  $k$  within  $S$ :

$$B_k = s_{1k} \times s_{2k} \times \dots \times s_{nk} \quad (3)$$

where  $s_{jk}$  is a subset of the possible values of input variable  $x_j$ ; so  $\{s_{jk} \subseteq S_j\}_1^n$ . A given box  $B_k$  is then described by the intersection of the subsets of values of each input variables  $x_j$ .

$$x \in B_k = \bigcap_{j=1}^n (x_j \in s_{jk}) \quad (4)$$

in case of real or discretely valued input variables, the subsets are contiguous sub-intervals:

$$s_{jk} = [t_{jk}^-, t_{jk}^+] \quad (5)$$

in case of categorical valued input variables,  $s_{jk}$  is any possible subset of the categories  $S_j$ :

$$s_{jk} \subset S_j \quad (6)$$

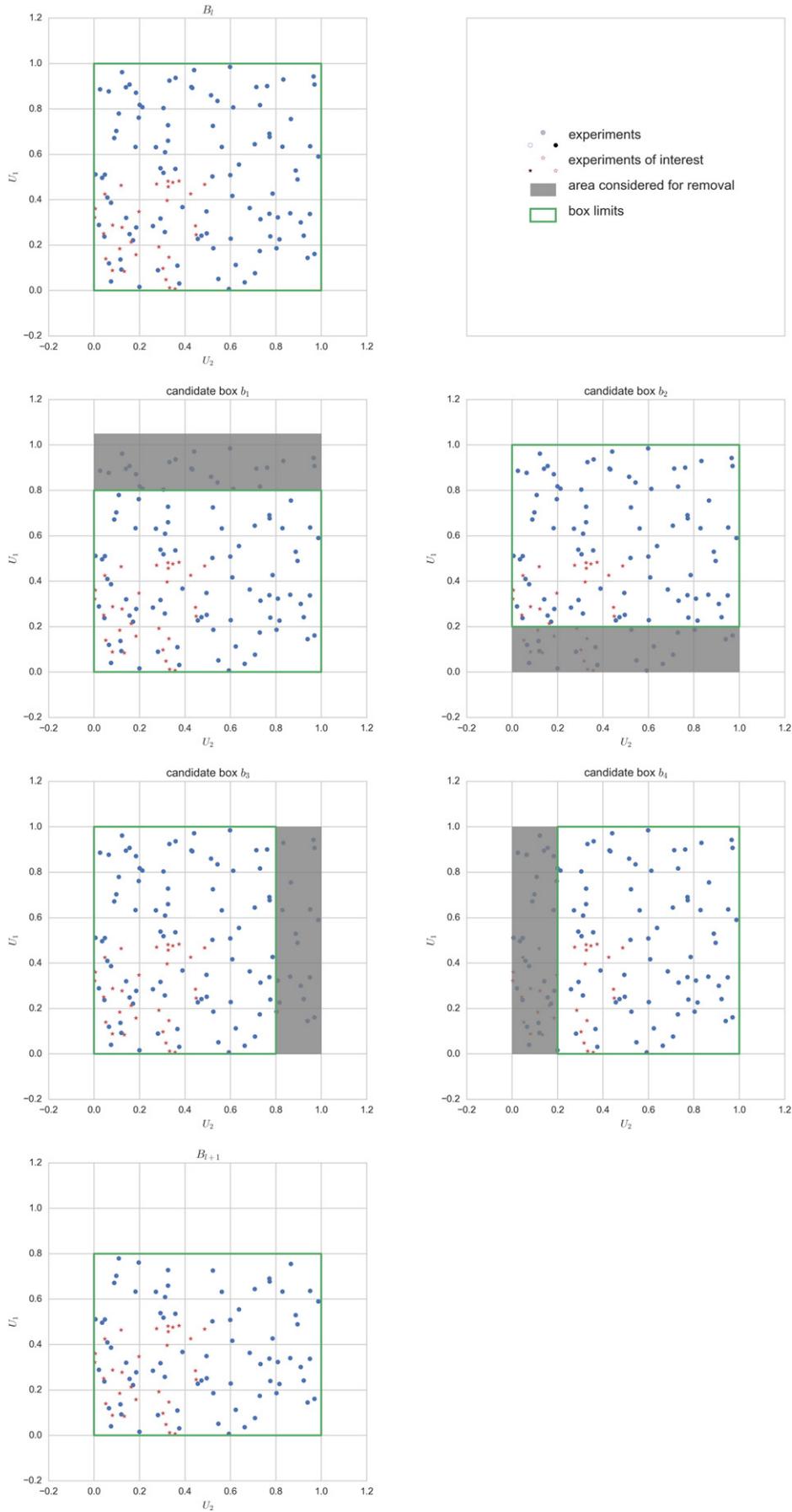
It is possible that the subset or sub interval  $s_k$  for any variable  $j$  is equal to the entire set or interval  $S_j$ , so  $s_{jk} = S_j$ , in which case this variable  $x_j \in S_j$  can be omitted from the box definition. The box definition then becomes

$$x \in B_k = \bigcap_{s_{jk} \neq S_j} (x_j \in s_{jk}) \quad (7)$$

The input variables  $x_j$  for which  $s_{jk} \neq S_j$  define the box  $B_k$ .

In order to find a given box  $B_k$ , PRIM uses a lenient hill climbing optimization procedure. Following Friedman and Fisher (1999), we consider here only the maximization case. The objective function of this optimization procedure in the default version of PRIM is to maximize  $\bar{Y}_{B_k}$ . That is, the average of  $y_i$  in  $B_k$ .

PRIM uses a lenient optimization procedures based on recursive top-down peeling, followed by bottom-up recursive pasting. An intuitive understanding of peeling is that recursively a small slice from the top or bottom of a given box is removed. Pasting is the converse procedure,



**Fig. 1.** Visual explanation of the PRIM algorithm.

where recursively a small slice is added back to the box. As also shown in Fig. 1, the optimization procedure starts with an initial box  $B_l$  that covers all the data. Iteratively a small sub-box  $b$  within  $B_l$  is removed. The algorithm first identifies all candidate boxes  $b_j$  that are eligible for removal. Next, Friedman and Fisher (1999) suggest several alternative criteria that can be used to select the best box  $b^*$  from these. The simplest criterion is to choose the box  $b^*$  that has the largest output mean value for the new box resulting from removing  $b$  from  $B$ . However, in case of heterogeneously typed data, this simple criterion is flawed. The average for a candidate box  $b_j$  in case of real valued variable will typically be based on more data points than for a categorical variable. To correct for this, the mass has to be taken into consideration as well. Friedman and Fisher (1999) therefore suggest a more lenient criterion that could be used instead:

$$b^* = \arg \max_{b \in C(b)} \frac{\text{ave}[y_l | x_l \in B_l - b] - \text{ave}[y_l | x_l \in B_l]}{\beta_{B_l} - \beta_{B_l - b}} \quad (8)$$

where the index  $l$  specifies the current box  $B_l$  in the peeling trajectory, and  $B_l - b$  is the box resulting from the removal of candidate sub-box  $b$  from  $B_l$ . So, the more lenient criterion is to select a sub-box  $b^*$  by looking at the change in the average divided by the change in the mass. In the paper, this more lenient criterion is used. Kwakkel and Jaxa-Rozen (2016) offer an in depth exploration of the impact of various alternative objective functions and suggest that the more lenient function used here is a suitable function to use as a default.

$$B_{l+1} \leftarrow B_l - b^* \quad (9)$$

Where the index  $l$  denotes the order of the box  $B$  in the peeling trajectory (see below). This peeling procedure is repeated recursively on each new smaller box until the mass of the box  $\beta_{B_k}$  falls below a user specified threshold. The mass of the box is simply the number of data points inside the box  $B_k$  divided by the total number of data points  $N$ . This threshold is a user defined parameter, and is in scenario discovery typically selected through trial and error.

$$\beta_{B_k} = \frac{1}{N} \sum_{i=1}^N 1(x_i \in B_k) \leq \beta_0 \quad (10)$$

The results of this recursive peeling is a succession of boxes, where each box is slightly smaller than the previous, has a slightly smaller mass than the previous, and shows an increase on the objective function. This succession of boxes is called a peeling trajectory in scenario discovery:

$$\{\bar{y}_l, \beta_l\}_1^L \quad (11)$$

Each candidate sub-box in  $C(b)$  is defined by a single input variable  $x_j$ . In case of a real valued or discrete valued input variable, this variable provides two candidate sub-boxes  $b_{j-}$  and  $b_{j+}$ . These two candidate boxes border respectively the upper and lower boundary of the current box  $B$  on the  $j$ -th input:

$$\begin{aligned} b_{j-} &= \{x | x_j < x_{j(\alpha)}\} \\ b_{j+} &= \{x | x_j < x_{j(1-\alpha)}\} \end{aligned} \quad (12)$$

where  $x_{j(\alpha)}$  is the  $\alpha$ -quantile of the values of  $x_j$  within the current box, and  $x_{j(1-\alpha)}$  is the  $(1-\alpha)$ -quantile. The value for  $\alpha$  is a user-defined parameter and typically quite small (0.05–0.1). This parameter controls the leniency of the peeling for real valued and discrete valued data.

In case of a categorical valued input variable, this variable provides a number of candidate boxes. This number is equal to the cardinality of the set of categories (i.e.  $|S_{jk}|$ ) remaining in definition of the box  $B$  minus one. So, in case of a categorical variable  $x_j$  where  $|S_{jk}| = 5$ , the

number of candidate boxes contributed by this categorical variable will be 4.

$$b_{jm} = \{x | x_j = s_{jm}\}, s_{jm} \in S_{jk} \quad (13)$$

The PRIM algorithm results in a peeling trajectory. In the context of scenario discovery, the analyst selects an appropriate box from this peeling trajectory. For this, three criteria are used: coverage, density, and interpretability. (Bryant and Lempert, 2010). Coverage is the fraction of all the cases that are of interest that fall within a box. A coverage of 1 means that all of the cases of interest are contained in a given box. A coverage of 0 means that none of the cases of interest are contained in the box. Density is the fraction of cases within a box that is of interest. A density of 1 means that all of the cases in the box are of interest. A density of 0 means that none of the cases in the box are cases of interest. Interpretability is more difficult to operationalize. One solution is to use as a proxy the number of uncertain factors that make up a box definition. By calculating coverage, density, and the number of uncertain factors that make up the box definition for each candidate box in the peeling trajectory, an analyst can make an informed choice on which of the boxes from the peeling trajectory to use.

### 3.2. Random forest

Random forest is a well-established machine learning technique that can be used for both classification and regression. It is an ensemble technique, where a set of simpler learners are combined to produce a single much more powerful classifier. The original version of random forest used the CART algorithm (Breiman et al., 1984) for the simpler learners. There are no principal reasons to not use other learners instead. Each simple learner is trained on a random subset of the data and a random subset of features.

#### 3.2.1. Random feature selection

Drawing inspiration from the work of Dietterich (2000) on random split selection, the work of Ho (1998) on the ‘random subspace’ technique, and the work of Amit and Geman (1997) who build decision trees by randomly selecting a subset of features at each split in the tree, Breiman (2001) proposes to train learners based on randomly selected features. Adopting the notation used for describing PRIM, the idea is that at each split in the tree, a subset of  $k$  variables  $s_k$  is randomly selected, so  $\{s_k \in S\}_{k=1}^k \subset S$ , and the best split from this subset is used.

#### 3.2.2. Bagging

Bagging or bootstrap aggregating is a technique first proposed by Breiman (1996) for generating several alternative versions of a predictor and then combining them into a single aggregate predictor. In the typical case, a learning set  $\mathcal{L}$  is used to train a predictor  $\varphi$  (so  $\varphi(x, \mathcal{L})$ ). So in case of PRIM, a given box  $B_k$  is such a predictor  $\varphi$ . Bootstrap aggregating is then a procedure for generating repeated bootstrap samples  $\{\mathcal{L}^{(B)}\}$  from  $\mathcal{L}$  and train a predictor on each of these samples:  $\{\varphi(x, \mathcal{L}^{(B)})\}$ . The bootstrap samples are generated by drawing at random, but with replacement, from  $\mathcal{L}$ . Next these individual predictors have to be combined to create the aggregate predictor  $\varphi_B$ . In case of numerical values, one takes the average value over the predictors

$$\varphi_B(x) = \text{ave}_B \varphi(x, \mathcal{L}^{(B)}) \quad (14)$$

while in case of classification, the individual classifiers vote to arrive at  $\varphi_B(x)$ . Bagging can be applied in combination with any predictor.

#### 3.2.3. Interpretability

Random forests are known to be very effective classifiers, but the exact rules that are being used in classification are impenetrable. In the context of scenario discovery, interpretability is key. Breiman

(2001) offers one approach for addressing the interpretability problem. He suggests using the resulting ensemble for calculating the input variable importance. This metric can be calculated by taking the out-of-bag data, randomly permute the  $m$ -th input variable, and then run it through the associated set of rules. This gives a set of predictions for class membership, which by comparing it with the true label gives a misclassification rate. This misclassification rate is the “percent increase in misclassification rate as compared to the out-of-bag rate” (Breiman, 2001). Based on this input variable importance score, a new classifier can be made which uses only the  $n$  most important input variables, or, slightly more sophisticated, one generates a series of boxes where step-wise the  $n$ -th most important input variable is added to the training data.

### 3.3. Bagging random boxes

Having presented both PRIM and random forest, we can now outline the modified PRIM procedure we are proposing.

1. Take a random bootstrap sample  $\mathcal{L}_k$  from  $\mathcal{L}$ , as discussed under bagging
2. Select a random subset of variables on which to train PRIM; given that PRIM is a patient rather than a greedy strategy, we use random feature selection prior to training, rather than at each successive step in the peeling procedure.
3. Train PRIM following the procedure outlined in Section 3.1
4. Assess the quality of the resulting box  $B_k$  using  $\mathcal{L}$  (see Breiman (1996) on using the entire learning set  $\mathcal{L}$  as a test set)

The outlined procedure will result in a single box  $B_k$ . Following the bagging procedure, a number of these boxes can be generated, and used as an aggregate predictor. However, this aggregate predictor will have a black box character. That is, it is not trivial to specify the classification rules used by the aggregate predictor. Given the importance of interpretability of the results of applying PRIM in a scenario discovery context, this is a problem.

A first approach for addressing the interpretability problem is to apply the procedure suggested by Breiman (2001) in the context of random forests. A second approach is suggested by Friedman and Fisher (1999), where one would take the peeling trajectories of each of the individual boxes  $B_k$  and identify for each the trade-off between the mean of the points inside the box and the mass of the box. In a scenario discovery context, this is adapted by looking at the trade-off between coverage and density. Given a set of boxes, it is possible to identify the Pareto front of all the peeling trajectories. Next, the analyst can inspect this front and make an informed choice for a particular box on it. The result of this procedure is that out of the ensemble of boxes that is being generated, only a single box is used. This single box is easily interpreted.

The presented random boxes approach can be used in combination with recent methodological advances on scenario discovery. The procedure is compatible with PCA-PRIM Dalal et al. (2013), can be adapted to use any of the objective functions discussed by Kwakkel and Jaxa-Rozen (2016), can be combined with f-PRIM (Hallegatte et al., 2015), and can be used in combination with the suggested modifications for multiple boxes of (Guivarch et al., 2016).

## 4. Results

In this section, we demonstrate the presented approach by applying it to three cases. The first case is the original case used by Bryant and Lempert (2010) when they first introduced scenario discovery. This case contains only continuous uncertain factors. The second case is based on the work of Rozenberg et al. (2013), where the standard scenario discovery approach was used with discrete uncertain factors. As the third case, we use the case of Hamarat et al. (2014). This case has continuous, discrete, and categorical uncertain factors. We thus explicitly are reusing cases that have been published before. This allows for

comparing the results of the presented approach with the original scenario discovery approach. Moreover, by moving from a relatively simple cases with homogenous data types to a more complex case with heterogeneous data types, we get insight into the performance of both PRIM and the random boxes approach for both.

In order to compare the performance of the new approach with the results from a normal application of PRIM, we perform three analyses. We randomly split the dataset into a training sample containing 60% of the data and a test sample containing the remaining 40% of the data. Next we train both the random boxes approach and normal prim on the training data. We assess how well the identified rules generalize using the test data. We also include a normal prim analysis on all the data. In the comparison, we will use coverage density trade off curves (Bryant and Lempert, 2010).

### 4.1. Bryant & Lempert data

Bryant and Lempert (2010) demonstrated scenario discovery using a renewable energy case that considers the potential impact of the so called 25 by 25 policy on the development of greenhouse gas emissions and economic costs. This 25 by 25 policy aims at 25% renewable sources for electricity and motor fuels by 2025 in the United States. The aim of applying scenario discovery to this case is to reveal the conditions under which the 25 by 25 policy results in unacceptably high economic costs. The dataset on which scenario discovery is being applied was generated using a simulation model (Toman et al., 2008) that calculates the social costs given assumptions pertaining the performance of various technologies and consumer behavior among others. Table 1 lists the 9 uncertain factors and their ranges that have been explored using a Latin Hypercube sampling strategy. The resulting dataset as used both here and in the original work of Bryant and Lempert (2010) contains 882 cases.

Fig. 2 shows the results of applying both normal PRIM and the random boxes approach to the data. As can be seen, the random boxes approach performs better on the test data than the normal PRIM approach. This suggests that the random boxes approach produces results that are generalize better than the results from the normal PRIM procedure. For comparison we also included the results of applying PRIM to the complete dataset, as is normally done in scenario discovery. If we compare the random boxes results with the normal PRIM results, we can draw two important conclusions. First, the differences are quite small. Some of the candidate boxes found using the random boxes approach are better in terms of both coverage and density than results found through applying PRIM to the entire data set. This leads to a second point, the common practice of applying PRIM to the entire dataset can result in over fitting. That is, in results that will not stand up to the inclusion of additional data.

Table 2 shows the feature scores resulting from the random boxes approach. This feature score is the misclassification rate. So a feature score of 0.2 means that if this uncertainty is randomly permuted on average 20% of the observations will be misclassified as a result. So, the higher the score, the more important a given uncertain factor is in predicting whether a given experiment produces good or bad outcomes.

**Table 1**

The uncertain factors and their ranges (adapted from Bryant and Lempert, 2010).

Uncertain factor	Range
Biofuel production costs	67–134 dollar per unit
Low-cost biomass supply	450–1000 millions of tons
Feedstock supply distribution	0–1
Biofuel yield	80–100 gal per ton
Oil supply elasticity	0.2–06
Transportation demand elasticity	–0.2 to –0.8
Electricity co-product	0–2 kWh per gallon
Shift in oil supply curve	–10–10% change
Biomass backstop price	90–200 \$ per ton

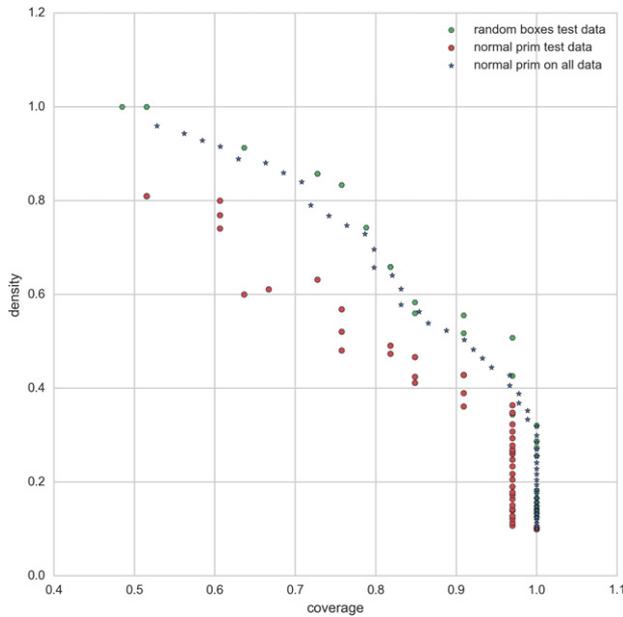


Fig. 2. Coverage versus density trade off curve for a normal PRIM on all data, a normal PRIM on the test data, and the random boxes approach on the test data for the dataset of Bryant and Lempert (2010).

The first four factors in this table match with the four factors identified by Bryant and Lempert (2010) in their original application of scenario discovery. The jump in feature score between cellulosic costs and electricity coproduction offers additional support for their choices.

A more detailed investigation of the results found using the random boxes approach suggests that it produces results that are qualitatively similar to those found in the original study of (Bryant and Lempert, 2010). That is, the random boxes approach identifies the same set of uncertain factors as being jointly responsible for cases with high social costs. The exact quantitative limits are somewhat different, explaining the difference in performance in terms of coverage and density. The same is true for the feature scoring approach.

4.2. Rozenberg data

Rozenberg et al. (2013) use scenario discovery for the bottom up development of Shared Socio-economic Pathways (SSPs). The scientific community in the context of climate change research has developed a new set of scenarios. These are based on a combination of the level of radiative forcing and plausible trajectories of future global developments. The radiative forcing is described in reference concentration pathways (RCP), while the trajectories for global development are described in as shared socio-economic pathways (SSPs). Any given climate change scenario is a combination of an RCP and an SSP. SSPs describe different socioeconomic conditions, different vulnerabilities, and different greenhouse gas emissions. The SSP's have been developed over the last few

Table 2  
Feature scores.

Uncertain factor	Feature score
Transportation demand elasticity	0.1998
Biomass backstop price	0.1204
Low-cost biomass supply	0.1157
Biofuel production costs	0.0703
Electricity coproduction	0.0339
Biofuel yield	0.0316
Shift in oil supply curve	0.0151
Feedstock supply distribution	0.0096
Oil supply elasticity	0.0082

years following essentially a scenario logic approach. Rozenberg et al. (2013) point out that such a top down approach to scenario development might miss the most important driving forces for the different SSPs. Hence, they suggest a bottom up, or backward, approach where one first identifies a large set of potential drivers for challenges to adaptation and mitigation, followed by a simulation based exploration of these drivers, and the subsequent a posteriori identification of which drivers matter when. This third stage relies on scenario discovery.

To demonstrate this bottom up approach, Rozenberg et al. (2013) use the IMACLIM-R model (Sassi et al., 2010). This model projects the long-term evolution of the global economy, given exogenous trends such as population and other exogenous parameters such as annual improvement in energy efficiency. The model has a substantial set of exogenous parameters and trends that can be varied by a user. To manage the potential combinatoric explosion, Rozenberg et al. (2013) group these parameters into four groups, in line with the existing SSP literature. These groups are globalization, environmental stress, carbon dependence and equity. Next the make alternative sets of internally consistent assumptions for each of these groups. They perform a full factorial analysis on these assumptions resulting in a set of 286 cases. The model outcomes from these 286 cases are subsequently scored in terms of challenges to mitigation and challenges to adaptation. To mimic the existing SSP's, Rozenberg et al. (2013) impose a rule based clustering of the cases such that a large majority of cases can uniquely be assigned to represent on of five SSPs. Next, PRIM is used to discover the key drivers for each of the five clusters. Here we only consider the application the SSP 1.

We follow the same procedure as in the previous case. The results of training PRIM on a training data set and testing its performance on a test data set, the results from the random boxes approach, and the results of applying PRIM to the entire dataset are shown in Fig. 3. As can be seen, the performance of PRIM on the test data is poor. The random boxes approach performs substantially better, and is quite close the results found by applying PRIM to the entire dataset. This again suggests that PRIM can easily be overfitted, in particular when confronted with smaller data sets.

We now compare the box definition as used in the original work with the closest box in terms of coverage and density from both the random boxes approach and normal PRIM on the test data. These boxes are annotated in Fig. 3. The associated box definition and coverage and

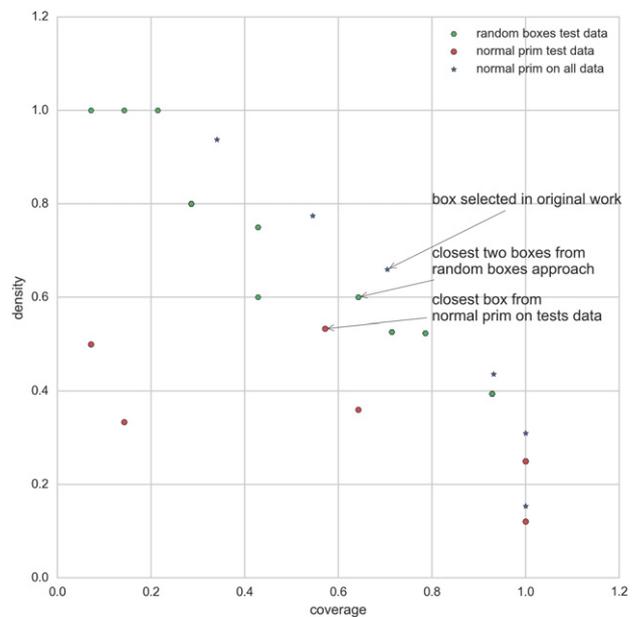


Fig. 3. Coverage versus density trade off curve for a normal PRIM on all data, a normal PRIM on the test data, and the random boxes approach on the test data for the dataset of Rozenberg et al. (2013).

**Table 3**  
Comparison of results.

	Original box		Closest for random boxes		Closest for test data
<b>Coverage</b>	0.71	<b>Coverage</b>	0.64	<b>Coverage</b>	0.57
<b>Density</b>	0.66	<b>Density</b>	0.60	<b>Density</b>	0.53
	<b>Range</b>	<b>Box 1</b>	<b>Range</b>		<b>Range</b>
<b>Behaviors</b>	1	<b>Inequalities</b>	1	<b>Behaviors</b>	1
<b>Inequalities</b>	1			<b>Inequalities</b>	1
<b>Population</b>	0–1	<b>Box 2</b>	<b>Range</b>	<b>Population</b>	0–1
		<b>Population</b>	0–1		
		<b>Behaviors</b>	1		
		<b>Technologies</b>	1		

density information is shown in Table 3. As can be seen, the results between the random boxes approach and the other two are different. Interestingly, the random boxes approach has two alternative box definitions with the same coverage and density. The first of these is based on the inequalities factor, while the second is based on population, behavior, and technologies. Apart from this last factor, the others are also used in both normal PRIM and in PRIM on the test data. If we dig deeper and also look at the feature scores, shown in Table 4, we observe that the highest scoring features are in fact those three factors, as also identified in the original paper.

Based on the results shown in Tables 3 and 4, we conclude that PRIM is sensitive to overfitting. In case PRIM is overfitted, the identified box definitions will not hold up in light of new data. Moreover, the identified box limits can severely misrepresent the key uncertainties that truly characterize the cases of interest.

4.3. Hamarat et al. data

Hamarat et al. (2014) use scenario discovery for developing an adaptive policy for steering the European energy system towards more sustainable functioning. Their starting point is the current emission-trading scheme (ETS). Using an integrated System Dynamics model of the system, they explore the performance of ETS across a wide range of uncertain factors. Next, scenario discovery is used to reveal combinations of uncertain factors under which ETS perform poorly. In light of these vulnerabilities, a set of adaptive actions is added to complement the ETS policy in order to improve the overall performance of the policy across all the uncertain factors.

The simulation model that is used represents the EU energy system including interconnections and congestion. The EU is split in seven regions. Nine power generation technologies are included, ranging from existing coal, gas and nuclear power generation, to gas and coal in combination with carbon capture and storage, and up to sustainable technologies such as solar and wind. The model outputs include the fraction of renewables in the energy generation portfolio over time, the average total costs of energy production, and the reduction of CO<sub>2</sub> emissions. The behavior of the model for these outputs is explored across 46 uncertain factors, a high level summary of which is given in Table 5. We used a Latin hypercube and performed 10,000 experiments. As can be seen, the uncertain factors are a mix of continuous variables, integer variables, and categorical variables.

**Table 4**  
Feature scores.

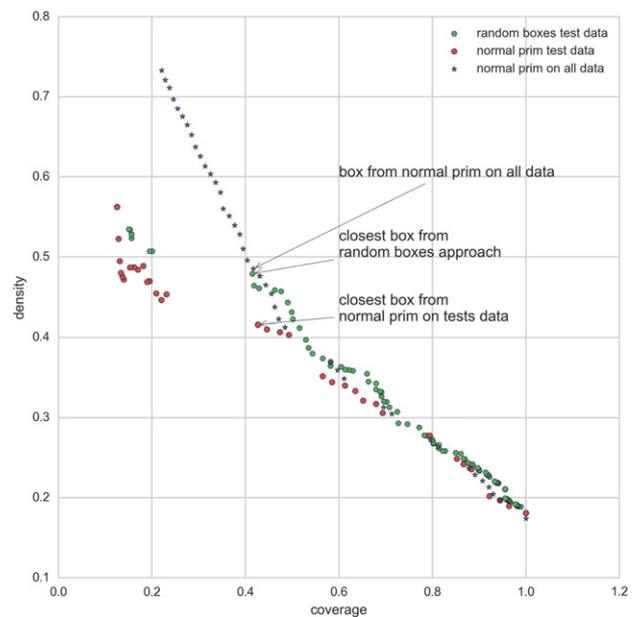
Uncertain factor	Feature score
Behaviors	0.2628
Population	0.1573
Inequalities	0.1369
Convergence	0.0822
Fossils	0.0486
Technologies	0.0465
Capital markets	0.0225

**Table 5**  
Specification of the uncertainties to be explored.

Name	Description
Economic lifetime	The average lifetime of each energy generation technology
Learning curve	The future performance increase of each energy generation technology
Economic growth	Six alternative economic growth scenarios are considered
Electrification rate	The rate of electrification of the economy is explored by means of six different electrification trends.
Physical limits	The effect of physical limits on the penetration rate of a technology is unknown. Two different behaviors are considered.
Preference weights	Investor perspectives on technology investments are treated as being deeply uncertain. Growth potential, technological familiarity, marginal investment costs and carbon abatement are possible decision criteria.
(Battery) storage	Storage of energy mainly in relation to sustainable energy generation technologies
Time of nuclear ban	A forced ban for nuclear energy in many EU countries is expected between 2013 and 2050. The time of the nuclear ban is varied between 2013 and 2050.
Price-demand elasticity	The elasticity of demand to price changes

We use the same approach as in the previous two cases. The results are shown in Fig. 4. As can be seen, the random boxes peeling trajectory dominates the normal PRIM on virtually all locations. This implies that the random boxes approach is likely to produce candidate boxes that are robust to new data. That is, the random boxes approach helps in preventing the inclusion of spurious uncertainties in the box definition. This figure also suggests that the normal PRIM procedure can get stuck in a local optimum, confirming the suggestion made in the introduction that PRIM can suffer from the usual defects of hill climbing optimization procedures.

Next, we compare three boxes that are quite close in terms of coverage and density. These boxes are indicated in Fig. 4. The box suggested by the random boxes approach represents a scenario where sustainable energy technology rapidly improves. In contrast, the performance of normal PRIM on the test data is abysmal. The box consists of 36 uncertain factors, the majority of which are barely restricted. This indicates



**Fig. 4.** Coverage versus density trade off curve for a normal PRIM on all data, a normal PRIM on the test data, and the random boxes approach on the test data for the dataset of Hamarat et al. (2013).

strong over fitting on part of PRIM. If we look at normal PRIM on the entire dataset, we see a scenario, which is broadly consistent with the one found using the random boxes approach. Here too, sustainable energy technology advances rapidly, and has a long lifetime.

Table 6 shows the feature scores for each of the individual uncertainties. The high scoring features largely correspond with those found in both normal PRIM on the entire dataset and those found in the random boxes approach. There are some subtle differences, for the random boxes approach and normal PRIM select slightly different factors from the high scoring features.

Based on the results of this third case study, we conclude that the problem of overfitting is clearly present here. The results for normal PRIM on the test data are very poor. The uncertain factors in the box definition do not adequately explain the cases of interest. The random boxes approach produces much better insight into the various uncertain factors that characterize the cases of interest.

#### 4.4. Synthesis of results

We compared the performance of normal PRIM fitted to the entire dataset, the performance of normal PRIM on a test data set, and the random boxes approach. In each of the three cases, the random boxes approach is superior to the performance of normal PRIM on the test data (Note that the test data and training data has been kept the same for both). This suggests that normal PRIM is sensitive to overfitting and that a random boxes approach can help in avoiding this. In this, it offers an alternative to quasi-p values as suggested by Bryant and Lempert (2010).

If we compare the random boxes approach and normal PRIM applied to the entire dataset, we observe that the boxes selected in the original work, and the closest box from the random boxes approach produce broadly consistent results in all three cases. This implies that the resampling statistics and quasi p-values suggested by Bryant and Lempert (2010) are effective in addressing both the interpretability concern as well as mitigate the risk of over fitting.

Compared to normal PRIM, the random boxes approach produces additional information through feature scoring. This is useful information for it gives insight into the relative importance of the different uncertain factors in predicting whether or not a given experiment produces a case of interest. Moreover, this information can be used diagnostically. One can compare a given box definition to this list to check

whether it uses the high scoring features. The function of the feature scores as calculated by the random boxes approach serve a purpose quite similar to the resampling statistics suggested by Bryant and Lempert (2010). They are, however, quite different in several respects. First, the resampling statistics apply only to a user-selected box, while the feature scores are generic. Second, the resampling statistics indicate the fraction of boxes generated on a random subset of the data that has essentially the same box limits as the user selected box. In contrast, the feature scores percent increase in misclassification rate if the given uncertain factor is not included as part of a box definition.

There is another potential benefit of the random boxes approach. In Fig. 4, we see that the random boxes peeling trajectory and the peeling trajectory of PRIM on the whole data set track one another quite closely at first, but a large gap opens up in the top left corner, immediately after the box selected in the original work. This gap might be indicative of over training on part of normal PRIM. This interpretation is supported by the fact that the performance of normal PRIM on the test data, also performs poorly in the top left corner. This suggests that the random boxes approach in conjunction with normal PRIM on the total dataset might be of use in supporting the selection of a box from the peeling trajectory that is less likely to be overfitted.

The analysis in this paper is based on three previously published cases. Future works should test the efficacy of the random boxes approach on more cases in order to assess whether this approach is always useful or whether its efficacy is case dependent. Given the success of random forest, however, we speculate that the random boxes approach will virtually always add value.

## 5. Conclusion

In this paper we explored a way of improving PRIM, which is the dominant algorithm currently used for scenario discovery. When using PRIM for scenario discovery, the analyst faces a twofold challenge. First, the results should be interpretable by the analyst and communicable to the client. Second, the results found by PRIM should be truly predictive of the computational experiments of interest. In this paper, we explored a new approach, which addresses these two challenges. This new random boxes approach is inspired by random forest. A random forest is a collection of Classification and Regression Trees, where each tree is trained on a random subset of the data, and where at each split in the tree, a random subset of uncertain factors is considered.

**Table 6**  
Feature scores for the 46 uncertainties.

Uncertain factor	Feature score	Uncertain factor	Feature score
SWITCH electrification rate	0.2058	Investors desired excess capacity investment	0.0315
SWITCH economic growth	0.2005	Economic lifetime hydro	0.0312
SWITCH physical limits	0.1685	Progress ratio biomass	0.0284
Time of nuclear power plant ban	0.1496	Demand fuel price elasticity factor	0.0277
Economic lifetime wind	0.1092	SWITCH market price determination	0.0274
Progress ratio wind	0.0774	Progress ratio ngcc	0.0274
Progress ratio pv	0.0708	Investment proportionality constant	0.0261
SWITCH interconnection capacity expansion	0.0666	Economic lifetime gas	0.0261
SWITCH storage for intermittent supply	0.0665	Maximum battery storage uncertainty constant	0.0232
Progress ratio nuclear	0.0607	Starting construction time	0.0207
Uncertainty initial gross fuel costs	0.0588	Economic lifetime igcc	0.0201
SWITCH preference carbon curve	0.0555	Price volatility global resource markets	0.0168
SWITCH low reserve margin price markup	0.0512	Year	0.0163
SWITCH carbon cap	0.0413	Maximum no storage penetration rate pv	0.0153
Weight factor technological familiarity	0.0399	Economic lifetime ngcc	0.0149
Economic lifetime coal	0.0373	Progress ratio coal	0.0144
Economic lifetime pv	0.0363	Weight factor carbon abatement	0.0137
Price demand elasticity factor	0.0358	Progress ratio igcc	0.0112
Weight factor marginal investment costs	0.0348	Economic lifetime biomass	0.0094
Progress ratio gas	0.0346	Weight factor technological growth potential	0.0094
SWITCH TGC obligation curve	0.0339	Progress ratio hydro	0.0083
Economic lifetime nuclear	0.0328	Maximum no storage penetration rate wind	0.0083
SWITCH lookup curve TGC	0.0318	Policy	0
SWITCH carbon price determination	0.0317	Model	0

The predictions of the resulting trees are aggregated through a voting system or by taking the average across the trees. Random forest outperforms individual trees. We adapted this approach by replacing CART with PRIM, and train each instance of PRIM on a random subset of the data and a random subset of the uncertain factors.

From our analysis of three previously published cases, we conclude that the resulting random boxes approach does indeed outperform individual PRIM. In each of the three cases, the performance of PRIM on the test data was dominated by the performance of the random boxes results on the test data. That is, we have shown that it is possible to improve on the results found through normal PRIM by adopting a random boxes approach. We also have found that the resampling statistics and quasi-p values suggested by Bryant and Lempert (2010) can be used successfully to counteract the risk of overfitting. For the use of PRIM in scenario discovery, together, this implies that users should at a minimum always use and report the resampling statistics and quasi-p values. When confronted with more complicated datasets with heterogeneously typed uncertain factors, or with relatively small data sets, the analysis could benefit from the use of the random boxes approach.

The presented random boxes approach addresses the interpretability concern through both feature scores and the identification of the Pareto front. Feature scores can help in deciding to drop certain uncertainties from the box definition, making interpretation easier. The Pareto front peeling trajectory, which in our case dominated the peeling trajectory of normal PRIM on the test data, helps in finding high coverage high density boxes.

The random boxes approach addresses the risk of overfitting in two ways. First, the presented approach relies on training the random boxes algorithm on a training data set, and assessing the adequacy of the result through a test set. The resulting Pareto front in terms of coverage and density thus only contains box definitions that have generalized well on the test data. Second, the feature scores offer an additional tool, in addition to quasi p-values, that analysts can use to mitigate the risk of overfitting. Low scoring features should be excluded from the definition of a given box.

There are several directions for future work. A first direction is the interpretability of the ensemble of random boxes. In this paper, we addressed this through feature scores and the Pareto front. This interpretability concern also exists in case of random forest. A more thorough analysis for how this is addressed in the literature on random forest might reveal additional techniques that can be adapted to also work with the random boxes approach.

In this paper, we explored a way of improving PRIM through ideas derived from random forest. An alternative direction that could be investigated is to assess the extent to which PRIM could be improved by combining it with Adaboost (Freund and Schapire, 1997). Adaboost is an alternative to random forest. Like random forest, Adaboost is an ensemble method. In contrast to random forest, each ensemble member is generated based on reweighting the training data in light of the performance of the previously trained classifier. So, any observation that is misclassified by the first classifier is weighted more heavily in training the next classifier. This approach can be adapted to PRIM in a relatively straightforward manner.

Both combining Adaboost with PRIM and the random boxes approaches explored in this paper keep the peeling and pasting procedure used in PRIM intact. There is no a priori reason why other optimization procedures could not be used instead of the lenient hill climbing used by PRIM. For example, the peeling trajectories shown in Fig. 4 suggest the use of a multi-objective optimization procedure where this peeling trajectory is identified directly by the algorithm, rather than emerging from the route followed by the hill climbing optimization procedure. That is, it might be possible to perform scenario discovery by optimizing coverage, density, and the number of restricted dimensions jointly given box limits. These box limits would then be the decision variables used in the optimization. This idea can be implemented relatively

straightforwardly using either simulated annealing or a genetic algorithm. Such an approach to improving PRIM might be particularly appealing given the evidence provided in this paper about local optima.

## Acknowledgements

We thank Ben Bryant and Julie Rozenberg for making the data from their work available to us. We also thank Caner Hamarat for sharing the model, allowing us to generate a new dataset ourselves. An earlier version of this paper has been presented at the 2014 PICMET conference. Based on comments and suggestions received there, we substantially enhanced the paper. This research was supported by the Dutch National Science foundation (NWO), VENI grant 451-13-018.

## References

- Amer, M., Daim, T.U., Jetter, A., 2013. A review of scenario planning. *Futures* 46, 23–40.
- Amit, Y., Geman, D., 1997. Shape quantization and recognition with randomized trees. *Neural Comput.* 9 (7), 1545–1588.
- Auping, W., Pruyt, E., Kwakkel, J.H., 2015. Societal ageing in the Netherlands: a robust system dynamics approach. *Syst. Res. Behav. Sci.* 32 (4), 485–501.
- Banques, S.C., Walker, W.E., Kwakkel, J.H., 2013. Exploratory Modeling and Analysis. *Encyclopedia of Operations Research and Management Science*. S. Gass and M. C. Fu, Berlin, Germany, Springer.
- Börjesson, L., Höjer, M., K.H., D., Ekvall, T., Finnveden, G., 2006. Scenario types and techniques: towards a user's guide. *Futures* 38 (7), 723–739.
- 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.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–23.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Wadsworth, Monterey, CA.
- Bryant, B.P., Lempert, R.J., 2010. Thinking inside the box: a participatory computer-assisted approach to scenario discovery. *Technol. Forecast. Soc. Chang.* 77, 34–49.
- Comes, T., Wijngaards, N., van de Walle, B., 2015. Exploring the future: runtime scenario selection for complex and time-bound decisions. *Technol. Forecast. Soc. Chang.* 97, 29–46.
- Dalal, S., Han, B., Lempert, R., Jaycocks, A., Hackbarth, A., 2013. Improving scenario discovery using orthogonal rotations. *Environ. Model. Softw.* 48, 49–64.
- Derbyshire, J., Wright, G., 2014. Preparing for the future: development of an 'antifragile' methodology that complements scenario planning by omitting causation. *Technol. Forecast. Soc. Chang.* 215–225.
- Dieterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Mach. Learn.* 40, 139–157.
- Eker, S., van Daalen, C., 2015. A model-based analysis of biomethane production in the Netherlands and the effectiveness of the subsidization policy under uncertainty. *Energy Policy* 82, 178–196.
- European Environmental Agency, 2009. Looking Back on Looking Forward: A Review of Evaluative Scenario Literature. European Environmental Agency, Luxembourg.
- Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* 55 (1), 119–139.
- Friedman, J.H., Fisher, N.L., 1999. Bump hunting in high-dimensional data. *Stat. Comput.* 9 (2), 123–143.
- Gerst, M.D., Wang, P., Borsuk, M.E., 2013. Discovering plausible energy and economic futures under global change using multidimensional scenario discovery. *Environ. Model. Softw.* 44, 76–86.
- Goodwin, P., Wright, G., 2010. The limits of forecasting methods in anticipating rare events. *Technol. Forecast. Soc. Chang.* 77, 355–368.
- Greeven, S., Kraan, O., Chappin, E.J.L., Kwakkel, J.H., 2016. The emergence of climate change mitigation action by society – an agent-based scenario discovery study. *J. Artif. Soc. Soc. Simul.* 19 (3). <http://dx.doi.org/10.18564/jasss.3134>.
- Groves, D.G., Lempert, R.J., 2007. A new analytic method for finding policy-relevant scenarios. *Glob. Environ. Chang.* 17 (1), 73–85.
- Guivarch, C., Rozenberg, J., Schweizer, V., 2016. The diversity of socio-economic pathways and CO<sub>2</sub> emissions scenarios: insights from the investigation of a scenarios database. *Environ. Model. Softw.* 80, 336–353.
- 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. *Glob. Environ. Chang.* 23 (2), 485–498.
- Halim, R.A., Kwakkel, J.H., Tavasszy, L.A., 2015. A scenario discovery study of the impact of uncertainties in the global container transport system on European ports. *Futures*.
- Hallegatte, S., Shah, A., Lempert, R., Brown, C., Gill, S., 2012. Investment Decision Making Under Deep Uncertainty: Application to Climate Change. *The World Bank*.
- Hallegatte, S., Bangalore, M., Bonzanigo, L., Fay, M., Tamaro, K., Narloch, U., Rozenberg, J., Treguer, D., Vogt-Schilb, A., 2015. Shock Waves – Managing the Impact of Climate Change on Poverty. *The World Bank*, Washington, DC.
- Hamarat, C., Kwakkel, J.H., Pruyt, E., 2013. Adaptive robust design under deep uncertainty. *Technol. Forecast. Soc. Chang.* 80 (3), 408–418.

- Hamarat, C., Kwakkel, J.H., Pruyt, E., Loonen, E., 2014. An exploratory approach for adaptive policymaking by using multi-objective robust optimization. *Simul. Model. Pract. Theory* 46, 25–39.
- Ho, T.K., 1998. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (8), 832–844.
- Islam, T., Pruyt, E., 2015. Scenario generation using adaptive sampling – the case of resource scarcity. *Environ. Model. Softw.*
- Kwadijk, J.C.J., Haasnoot, M., Mulder, J.P.M., Hoogvliet, M.M.C., Jeuken, A.B.M., van der Krogt, R.A.A., van Oostrom, N.G.C., Schelfhout, H.A., van Velzen, E.H., van Waveren, H., de Wit, M.J.M., 2010. Using adaptation tipping points to prepare for climate change and sea level rise: a case study in the Netherlands. *Wiley Interdiscip. Rev. Clim. Chang.* 1 (5), 729–740.
- Kwakkel, J.H., Jaxa-Rozen, M., 2016. Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. *Environ. Model. Softw.*
- Kwakkel, J.H., Walker, W.E., Marchau, V.A.W.J., 2010. Classifying and communicating uncertainties in model-based policy analysis. *Int. J. Technol. Policy Manage.* 10 (4), 299–315.
- Kwakkel, J.H., Auping, W.L., Pruyt, E., 2013. Dynamic scenario discovery under deep uncertainty: the future of copper. *Technol. Forecast. Soc. Chang.* 80 (4), 789–800.
- Kwakkel, J.H., Haasnoot, M., Walker, W.E., 2015. Developing dynamic adaptive policy pathways: a computer-assisted approach for developing adaptive strategies for a deeply uncertain world. *Clim. Chang.* 132 (3), 373–386.
- Kwakkel, J.H., Walker, W.E., Haasnoot, M., 2016. Coping with the wickedness of public policy problems: approaches for decision making under deep uncertainty. *J. Water Resour. Plan. Manag.*
- Lempert, R.J., Collins, M., 2007. Managing the risk of uncertain threshold response: comparison of robust, optimum, and precautionary approaches. *Risk Anal.* 24 (4), 1009–1026.
- Lempert, R.J., Popper, S., Bankes, S., 2003. *Shaping the Next One Hundred Years: New Methods for Quantitative, Long Term Policy Analysis*. RAND, Santa Monica, CA.
- Lempert, R.J., Groves, D.G., Popper, S.W., Bankes, S.C., 2006. A general, analytic method for generating robust strategies and narrative scenarios. *Manag. Sci.* 52 (4), 514–528.
- Lempert, R.J., Bryant, B.P., Bankes, S., 2008. *Comparing Algorithms for Scenario Discovery*. RAND, Santa Monica, CA, USA.
- Lord, S., Halfgott, A., Vervoort, J.M., 2016. Choosing diverse sets of plausible scenarios in multidimensional exploratory futures techniques. *Futures* 77, 11–27.
- Matrosov, E.S., Padula, S., Harou, J.J., 2013a. Selecting portfolios of water supply and demand management strategies under uncertainty – contrasting economic optimisation and 'robust decision making' approaches. *Water Resour. Manag.* 27 (4), 1123–1148.
- Matrosov, E.S., Woords, A.M., Harou, J.J., 2013b. Robust decision making and info-gap decision theory for water resource system planning. *J. Hydrol.* 494, 43–58 (28 June).
- Parker, A.M., Srinivasan, S.V., Lempert, R.J., Berry, S.H., 2015. Evaluating simulation-derived scenarios for effective decision support. *Technol. Forecast. Soc. Chang.* 91, 64–77 (February).
- Pruyt, E., Islam, T., 2016. On generating and exploring the behavior space of complex models. *Syst. Dyn. Rev.*
- Pruyt, E., Kwakkel, J.H., 2014. Radicalization under deep uncertainty: a multi-model exploration of activism, extremism and terrorism. *Syst. Dyn. Rev.* 30 (1–2), 1–28.
- Rozenberg, J., Guivarch, C., Lempert, R.J., Hallegatte, S., 2013. Building SSPs for climate policy analysis: a scenario elicitation methodology to map the space of possible future challenges to mitigation and adaptation. *Clim. Chang.* 122 (3), 509–522.
- Sassi, O., Hourcade, J.-C., Crassous, R., Gitz, V., Waisman, H., Guivarch, C., 2010. IMACLIM-R: a modeling framework for sustainable development issues. *Int. J. Glob. Environ. Issues* 10 (1–2), 5–24.
- Toman, M., Griffin, J., Lempert, R., 2008. *Impacts on U.S. Energy Expenditures and Greenhouse-Gas Emissions of Increasing Renewable Energy Use*. RAND Corporation, Santa Monica, CA.
- Trutnevyte, E., Guivarch, C., Lempert, R.J., Strachan, N., 2016. Reinvigorating the scenario technique to expand uncertainty consideration. *Clim. Change.*
- van 't Klooster, S.A., van Asselt, M.B.A., 2006. Practising the scenario-axes technique. *Futures* 38 (1), 15–30.
- van Notten, P.W.F., Slegers, A.M., van Asselt, M.B.A., 2005. The future shocks: on discontinuity and scenario development. *Technol. Forecast. Soc. Chang.* 72 (2), 175–194.
- Walker, W.E., Haasnoot, M., Kwakkel, J.H., 2013. Adapt or perish: a review of planning approaches for adaptation under deep uncertainty. *Sustainability* 5 (3), 955–979.

**Jan Kwakkel** works as an associate professor at Delft University of Technology. His research focuses on model based decision support for decision making under deep uncertainty. He has applied a variety of methods and techniques in different domains. His current focus is climate change in the water sector.

**Scott Cunningham** received a Ph.D. in Science, Technology and Innovation Policy from the Science Policy Research Unit. He has worked as a knowledge discovery analyst for software and database companies, and has helped in advising the British and Malaysian governments on their national profiles of scientific activity. He currently works at the Delft University of Technology, as an associate professor.