# Improving Scenario Discovery by Bagging Random Boxes

J.H. Kwakkel, S.C. Cunningham, E. Pruyt
Delft University of Technology, Faculty of Technology, Policy and Management, Delft, The Netherlands

*Abstract*--**Scenario discovery is a novel participatory model-based approach to scenario development in the presence of deep uncertainty. Scenario discovery relies on the use of statistical machine-learning algorithms. The most frequently used algorithm is the Patient Rule Induction Method. This algorithm identifies regions in the uncertain model input space that are highly predictive of producing model outcomes that are of interest. To identify these regions, PRIM in essence 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 alleys. 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 through a case study of scenario discovery for the transition of the European energy system towards more sustainable functioning, focusing on identifying scenarios where the transition fails.**

## I. INTRODUCTION

Scenario discovery is a relatively novel approach for addressing the challenges of characterizing and communicating deep uncertainty associated with simulation models [1]. The basic idea is that the consequences of the various deep uncertainties associated with a simulation model are systematically explored through conducting series of computational experiments [2] and that the resulting data set is analyzed to identify regions in the uncertainty space that are of interest [3, 4]. These identified regions can subsequently be communicated through e.g. narratives to the decisionmakers and other actors involved. Scenario discovery is an analytical process which can be embedded in a participatory process supporting ”deliberation with analysis”[5].

A motivation for the use of scenario discovery is that the available literature on evaluating scenario studies has found that scenario development is difficult if the involved actors have diverging interests and worldviews [3, 6]. Another shortcoming identified in this literature is that scenario development processes have a tendency to overlook surprising developments and discontinuities [7-9].

Scenario discovery is an approach that offers support for decisionmaking under deep uncertainty. Deep uncertainty is encountered when the different parties to a decision do not know or cannot agree on the system model that relates consequences to actions and uncertain model inputs [10], or when decisions are adapted over time [11]. In these cases, it is possible to enumerate the possibilities (e.g. sets of model inputs, alternative relationships inside a model, etc.), without ranking these possibilities in terms of perceived likelihood or assigning probabilities to the different possibilities [12].

Although scenario discovery can be applied on its own [4, 13, 14], it is also a key step in Robust Decision Making (RDM) [1, 15-17]. RDM aims at supporting the design of robust policies. That is, policies that perform satisfactorily across a very large ensemble of future worlds. In this context, scenario discovery is used to identify the combination of uncertainties under which a candidate policy performs poorly, allowing for the iterative improvement of this policy. This particular use of scenario discovery 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 [18].

Currently, the main statistical rule induction algorithm used for scenario discovery is the Patient Rule Induction Method (PRIM) [19], although other algorithms, such as Classification and Regression Trees (CART) [20], are sometimes used [14, 21]. The main merit of PRIM is its interactive usage, which helps to overcome its main weakness of restricting too many dimensions. CART has a tendency to generate many and asymmetric boxes which are difficult to interpret. However, CART can be used for multiclass problems, while PRIM cannot [14]. 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, one seeks a set of 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. PRIM describes these subspaces in the form of ‘boxes’ of the model input space. To identify these boxes, PRIM uses a lenient or patient hill climbing optimization procedure. The most frequently employed implementation of PRIM that is being used for scenario discovery is the one provided by Bryant in the scenario discovery toolkit written in R [22].

There are two key concerns in scenario discovery. In no particular order, the first concern is the interpretability of the results. That is, 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 [3]. But, perhaps even more importantly, communicating such results to the stakeholders involved in the process becomes substantially more challenging [23]. The second concern is that the uncertainties in the subset should be significant. That is, PRIM should not include spurious uncertainties in the

definition of the identified subspace. 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, namely local optima, plateaus, and ridges and alleys.

In current 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 [3] 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 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 both concerns simultaneously. This alternative approach is inspired by the extensive work that has been done with CART and related classification tree algorithms. The basic idea behind this alternative approach is to perform multiple runs of the PRIM algorithm based on randomly selected uncertainties [24] and combining these results using a bagging technique [25]. 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, this 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 [25]. The inspiration for this alternative approach comes from [24], who combines random feature selection and bagging with CART [20], resulting in the by now well established random forest machine learning technique.

To assess the efficacy of the proposed alternative approach, we perform a case study. Given that as of yet no established bench mark cases are available for comparing alternative scenario discovery procedures, we use a case we have been working on for other purposes. The case is about the transition of the European energy system towards more sustainable functioning. Here, we focus on identifying scenarios where the transition fails. We apply both the standard scenario discovery procedure and the alternative approach to this case and compare the results.

The remainder of this paper is structured accordingly. In Section 2, we outline the method in more detail. Section 3 introduces the case. Section 4 contains the results. We discuss the results in Section 5. Section 6 contains the conclusions.

## II. 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.

### *A. 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 row 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 removes more data points than removing data from the right, so removing from the top is a better choice. 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 [19]. Given a learning set $\mathcal{L} = \{\mathrm{x_i}, \mathrm{y_i}\}_1^{\mathrm{N}}$ where $y_i$ is some output variable, and $\mathrm{S_j}$ is the set of all possible values for the input variable $\mathrm{x_j}$.

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

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

(2) $$S = S_1 \times S_2 \times \cdots \times S_n$$

The goal of PRIM is to find a subset $R$ of the input domain $S$, so $R \subset S$, for which

(3) $$\bar{f}_R = ave_{x \in R} f(x) = \frac{\int_{x \in R} f(x)p(x)dx}{\int_{x \in R} p(x)dx} \ll \bar{f}$$

where $\bar{f}$ is the average over the entire input space:

(4) $$\bar{f} = \int f(x)p(x)dx$$

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 the entire input domain $S$:

(5) $$B_k = s_{1k} \times s_{2k} \times \cdots \times s_{nk}$$

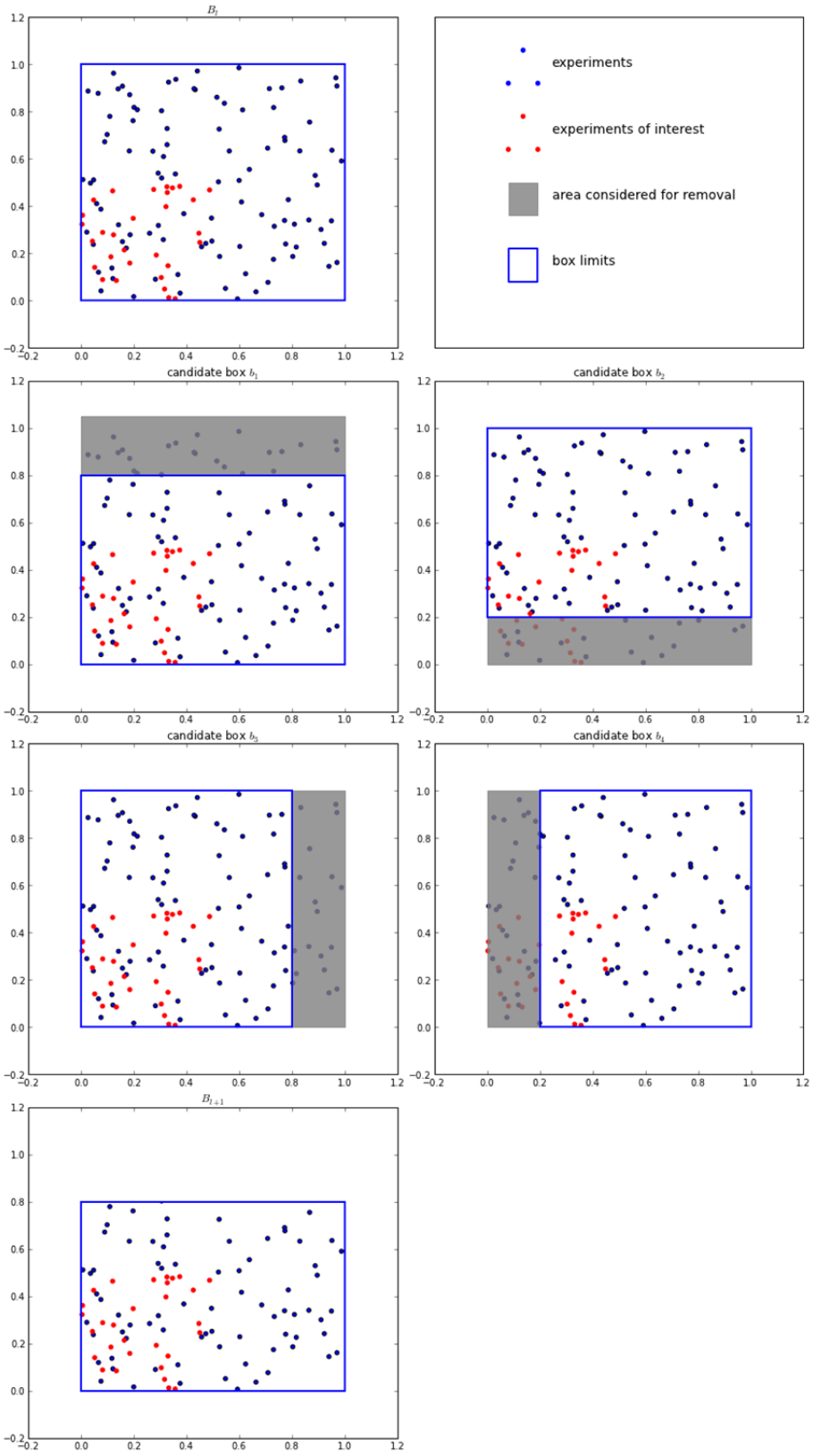


Fig. 1. Visual explanation of PRIM algorithm.

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 than described by the intersection of the subsets of values of each input variables $x_j$.

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

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

(7) $$s_{jk} = [t_{jk}^-, t_{jk}^+]$$

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

(8) $$s_{jk} \subset S_j$$

It is possible that the subset or sub interval $s_{jk}$ for any variable 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

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

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 [19], we consider here only the maximization case. The objective function of this optimization procedure in the default version of PRIM then becomes

(10) $$\bar{f}_{B_k} = \frac{\int_{x \in B_k} f(x)p(x)dx}{\int_{x \in B_k} p(x)dx}$$

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, 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 $b_j$ which are eligible for removal, and in the version presented by Friedman and Fisher [19] choses the box $b^*$ that has the largest output mean value for the new box resulting from removing $b$ from $B$.

(11) $$b^* = \arg\max_{b \in C(b)} ave\ [y_i | x_i \in B]$$

Where $C(b)$ is the class of sub-boxes $b_j$ eligible for removal. Given $b^*$, the box $B$ is updated:

(12) $$B_{l+1} \longleftarrow B_l - b^*$$

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.

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

The results of this recursive peeling is a succession of boxes 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:

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

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:

(15) $$b_{j-} = \{x | x_j < x_{j(\alpha)}\}$$
$$b_{j+} = \{x | x_j < x_{j(1-\alpha)}\}$$

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.

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

As indicated, Friedman and Fisher [19] select the candidate sub-box $b^*$ that has the maximum average for $B - b^*$. However, in case of heterogeneously typed data, this 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 [19] therefore suggest a more lenient criterion that could be used instead:

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

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 case, this more lenient criterion is used.

*B. 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, classification and regression trees (CART), are combined to produce a single much more powerful classifier. The individual trees in a forest are generated using the CART algorithm [20], and are trained on a random subset of the data and a random subset of features. The individual trees are combined through bootstrap aggregating, or bagging [25]. Below we briefly discuss each of these three components.

*CART*

Classification and regression tree (CART) [20] is an example of a decision tree learning algorithms. Other examples include ID3 [26], C4.5 [27], and C5.0 The aim in decision tree learning in general is to create a classifier that predicts the value of a target variable based on a set of input variables. In case of decision tree learning, this takes the form of a decision tree where at each node a given input variable is used to split the data set. CART can be used for both classification and regression problems. In case of regression problems, the tree is used to predict the value of the outcome of interest as a function of the set of input variables. In case of classification problems, the tree is used to predict class membership as a function of the set of input variables. CART differs from ID3 in that it uses gini-impurity instead of entropy as a basis for selecting which variable to split on. Through recursive partitioning, CART produces a set of disjoint rules that jointly cover the entire input space. Both CART and PRIM are rule induction procedures, but in contrast to PRIM, CART and other decision tree learners are greedy procedures.

*Random feature selection*

Drawing inspiration from the work of Dietterich [28] on random split selection, the work of Ho [29] on the 'random subspace' technique, and the work of Amit and German [30] who build decision trees by randomly selecting a subset of features at each split in the three, Breiman [24] proposes to build classification trees 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.

*Bagging*

Bagging or bootstrap aggregating is a technique first proposed by Breiman [25] for generating several alternative versions of a predictor and then combining them into a single aggregate predictor. In the typically 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

$$\varphi_B(x) = \text{ave}_B \varphi(x, \mathcal{L}^{(B)}) \tag{18}$$

while in case of classification, the individual classifiers vote to arrive at $\varphi_B(x)$. Bagging can be applied in combination with any predictor. In case of random forest, CART is used for the individual predictors $\varphi$.

*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 [24] offers one approach for addressing the interpretability problem. He suggests to use the resulting ensemble for calculating the input variable importance. This metric can be calculated by taking the out-off-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" [24]. 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.

*C. 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 0 with a the more lenient objective function
4. Asses the quality of the resulting box $B_k$ using $\mathcal{L}$ (see Breiman [25] 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.

An alternative to calculating the importance of features is suggested by Friedman and Fisher [19]. In this approach, 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 on this trade-off curve. Next, the analyst can inspect this resulting Pareto front and make an informed choice for a particular box on this Pareto Front. The result of this procedure is that out of the ensemble of boxes that is being generated, only a single box is used. As such this single box is easily interpreted. This completes the presentation of the method.

## III. CASE

Here we introduce the case we use for assessing the efficacy of the outlined method. The European Union (EU) has targets for the reduction in carbon emissions and the share of renewable technologies in the total energy production by 2020 [31]. The main aim is to reach 20% reduction in carbon emission levels compared to 1990 levels and to increase the share of renewables to at least 20% by 2020. However, the energy system includes various uncertainties related to technology lifetimes, economic growth, costs, learning curves, investment preferences and so on. For instance, precise lifetimes of technologies are not known and expected values are used in planning decisions. Furthermore, it is deeply uncertain how the economic conditions, which have a direct influence on the energy system, will evolve. Thus, it is of great importance to take these uncertainties into consideration when analyzing the energy system, and preparing policies for meeting the EU targets.

In order to meet the 2020 goals, the EU adopted the European Emissions Trading Scheme (ETS) for limiting the carbon emissions [31]. ETS imposes a cap-and-trade principle that sets a cap on the allowed greenhouse gas emissions and an option to trade allowances for emissions. However, current emissions and shares of renewables show a fragile progress of reaching the 2020 targets. It is necessary to take additional actions for steering the transition toward cleaner energy production. This requires a better handling of the uncertainties in the energy system and more robust policies that can promote renewable technologies.

In this study, a System Dynamics [32-34] model is used for simulating the plausible futures of the EU electricity system. The model represents the power sector in the EU and includes congestion on interconnection lines by distinguishing seven different regions in the EU. These are northwest (NW), northeast (NE), middle (M), southwest (SW), southeast (SE) of Europe, United Kingdom and Ireland (UKI) and Italy (I). Nine power generation technologies are included. These are: wind, PV solar, solid biomass, coal, natural gas, nuclear energy, natural gas with Carbon Capture and Sequestration (CCS), coal gasification with CCS, and large scale hydro power. The model endogenously includes mechanisms and processes related to the competition between technology investments, market supply-demand dynamics, cost mechanisms, and interconnection capacity dynamics. Not only endogenous mechanisms but also various exogenous variables are included. Fig. 2 shows the main sub-models that constitute this model at an aggregate level. These are, installed capacity, electricity demand, electricity price, profitability and levelised costs of electricity. At an aggregated level, there are two main factors that drive new capacity investments: electricity demand and expected profitability. An increase of the electricity demand leads to an increase in the installed capacity, which will affect the electricity price. This will cause a rising demand, in turn resulting in more installed capacity. On the other hand, decreasing electricity prices will lead to lower profitability and less installed capacity, which will result in electricity price increases. Each sub-model has more detailed interactions within itself and with the other sub-models and exogenous variables and these causal relationships drive the main dynamics of the EU electricity system.

Fig. 2 is a graphical representation of the main causal relationships between the main submodels. In order to run computational simulations, these relationships are translated into a system of differential equations, which are implemented in Vensim [35]. The model includes 33 ordinary differential equations, 499 auxiliary equations, and 632 variables. It is beyond the scope of this paper to include all the equations and variables separately. More detail on the model can be found in [36], including a detailed descriptions of each equation and variable.

We are interested in exploring and analyzing the influence of a set of deeply uncertain input variables on the key output variables. In order to explore the uncertainty space, not only parametric but also structural uncertainties are included in the analysis. For exploring structural uncertainties, several alternative model formulations have been specified and a switch mechanism is used for switching between these alternative formulations. Parametric uncertainties are explored over pre-defined ranges. Table 1 provides an overview of the uncertainties, 46 in total, that are analyzed and their descriptions.

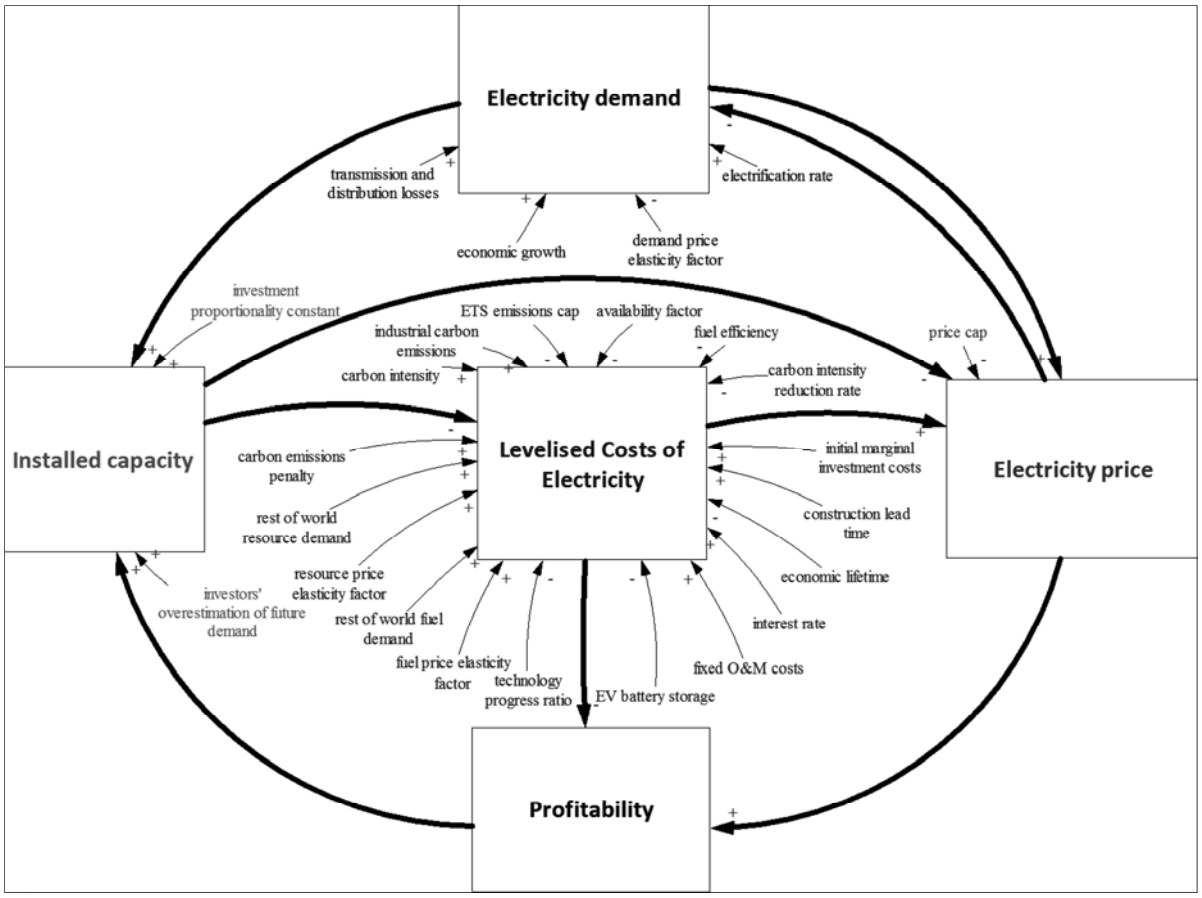


Fig. 2. A diagram of the main causal loops of the EU energy model.

TABLE 1: SPECIFICATION OF THE UNCERTAINTIES TO BE EXPLORED

| Name | Description |
|---|---|
| **Economic lifetime** | For each technology, the average lifetimes are not known precisely. Different ranges for the economic lifetimes are explored for each technology. |
| **Learning curve** | It is uncertain for different technologies how much costs will decrease with increasing experience. Different progress ratios are explored for each technology. |
| **Economic growth** | It is deeply uncertain how the economy will develop over time. Six possible developments of economic growth behaviors 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** | For wind and PV solar, the availability of (battery) storage is difficult to predict. A parametric range is explored for this uncertainty. |
| **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** | A parametric range is considered for price – demand elasticity factors. |

## IV. RESULTS

In order to explore the behavior of the European energy system in the presence of the emission trading scheme under uncertainty, we generated 5000 computational experiments using Latin Hypercube sampling. These 5000 experiments systematically cover the uncertainty space spanned by the 46 uncertainties included in this analysis. Fig. 3 shows the results of these 5000 experiments for two key performance indicators, namely the fraction of renewables and the carbon emission reduction fraction. The envelope shows the bandwidth of outcomes as encountered across the ensemble

of experiments. The Gaussian Kernel Density Estimate (KDE) at the left shows the distribution of the terminal values. A KDE can be understood as a continuous alternative to a histogram. The fraction renewables specifies the fraction of renewables in the total energy mix. As can be seen, in most runs this fraction increases over time, although there is still a substantial number of experiments where the fraction of renewables in 2050 is lower than in 2010. The carbon emission reduction fraction specifies by how much carbon emissions have been reduced as compared to the value in 2010. Note that a score above 0 means a decrease in emissions, while a score below 0 means an increase in emissions. From the KDE, we infer that in most experiments the emissions slightly increase as compared to the 2010 levels.

Taking the results for both the fraction of renewables and the carbon emissions reduction fraction together, it appears that the current ETS policy is far from effective in achieving the intended shift towards more sustainable energy generation. Even if the fraction of renewables where to increase, this increase does not appear to coincide with a substantial reduction of carbon emissions. Quite the opposite, in most experiments carbon emissions increase.

To perform scenario discovery, it is necessary to classify the results into results that are of interest and results that are not of interest. Typically in scenario discovery, the results where the policy fails are classified as being of interest. Here, we have chosen to focus on the subset of cases where the fraction of renewables in 2050 is lower than in 2010.

To test the random boxes approach, we compare it with a normal PRIM analysis. Both are parameterized in the exact same way. We generate 1000 random boxes, where each boxes uses 15 randomly selected uncertainties. We train both the normal PRIM and the random boxes approach on the dataset with 5000 experiments. As a test set, we generate a separate dataset containing 2000 experiments, again generated using Latin Hypercube sampling. Below we use the test set as the basis of comparison. It provides insight into how well the results from both the normal PRIM and the random boxes approach generalize.

Fig. 4 shows the peeling trajectories resulting from the normal prim and the random boxes approach. For the random boxes approach, we only show the best possible solution in terms of density and coverage (labeled Pareto front). 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.

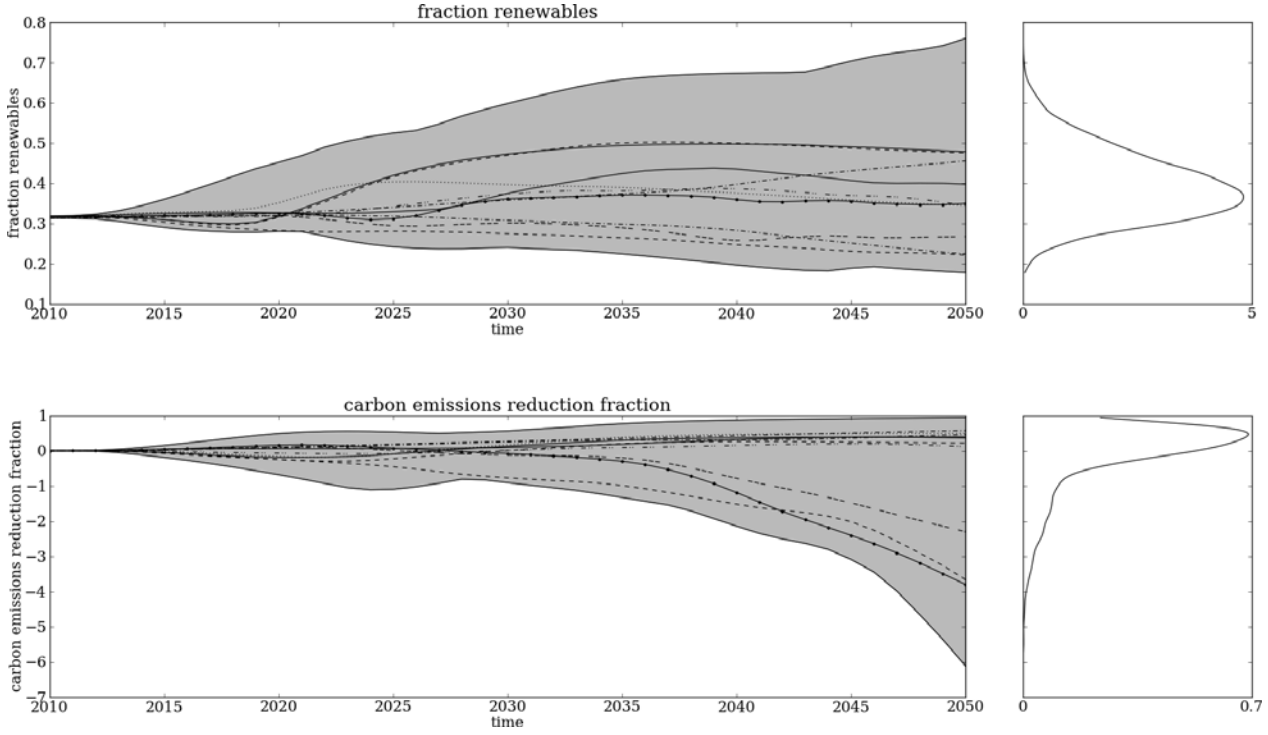


Fig. 3. Envelopes with traces for the fraction of renewables and the carbon emission reduction fraction, and Gaussian Kernel Density estimates for their terminal values.

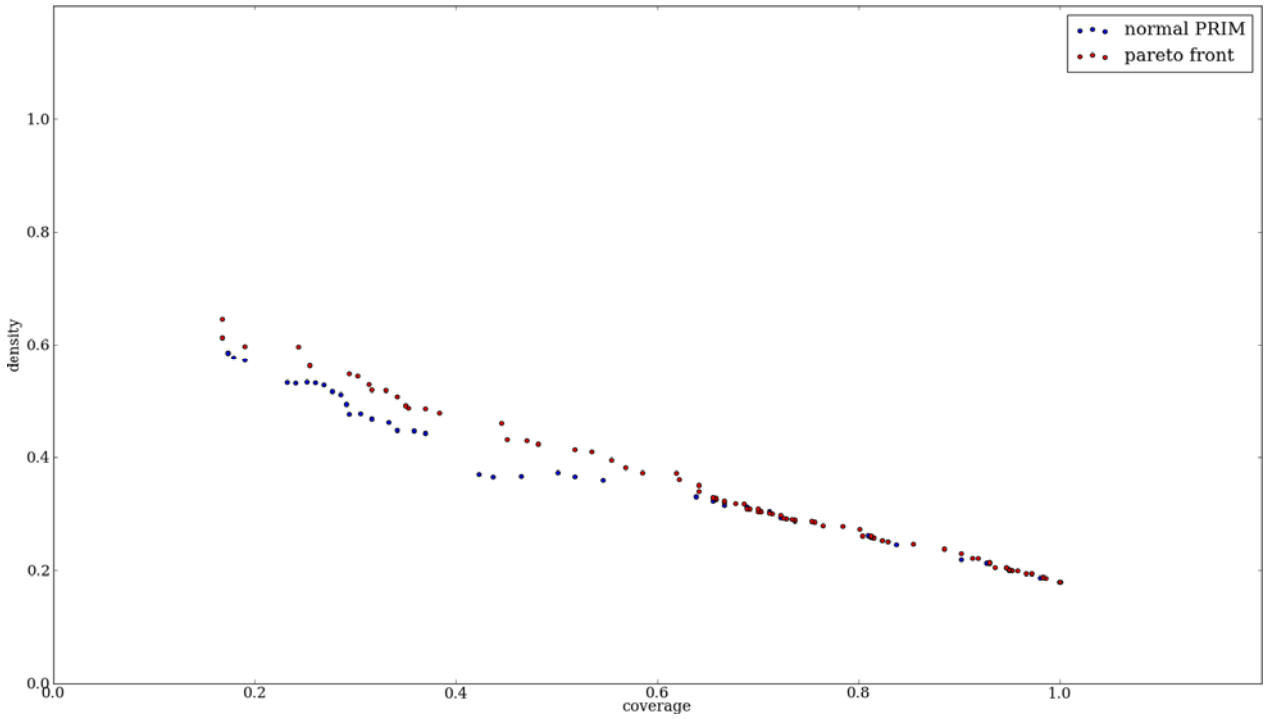


Fig. 4. Peeling trajectories.

Table 2 and Table 3 show the box definitions for the box with the highest density for the random boxes procedure and the normal PRIM procedure respectively. Both box definitions consist of 15 uncertainties. In case of the random boxes procedure, this is the maximum number possible. Out of these 15 uncertainties, 5 uncertainties occur in both (SWITCH economic growth, economic lifetime ngcc, investment proportionality constant, weight factor technological familiarity, and year), the other 10 are unique to either. Looking at the 5 shared uncertainties, we observe that the exact limits are slightly different, but the ranges overlap to a large degree. The substantial difference between the two boxes further supports the claim that the normal PRIM can get stuck in local optima that can be avoided by using the random boxes procedure.

TABLE 2. BOX DEFINITION OF HIGHEST DENSITY BOX FOUND THROUGH THE RANDOM BOXES PROCEDURE

| **uncertainty** | **bandwidth** |
|---|---|
| SWITCH TGC obligation curve | set([1, 3]) |
| SWITCH economic growth | set([1, 2, 5]) |
| SWITCH lookup curve TGC | set([1, 2, 3, 4]) |
| SWITCH low reserve margin price markup | set([1, 2, 3, 4]) |
| SWITCH storage for intermittent supply | set([1, 3, 5, 6, 7]) |
| economic lifetime biomass | 30.00 - 44.44 |
| economic lifetime hydro | 50.01 - 69.98 |
| economic lifetime ngcc | 25.00 - 40.00 |
| economic lifetime nuclear | 51.62 - 69.05 |
| investment proportionality constant | 0.70 - 3.45 |
| price volatility global resource markets | 0.10 - 0.20 |
| starting construction time | 0.10 - 3.00 |
| weight factor carbon abatement | 1.54 - 9.99 |
| weight factor technological familiarity | 1.84 - 9.56 |
| year | 0.92 - 1.10 |

TABLE 3. BOX DEFINITION OF HIGHEST DENSITY BOX FOUND THROUGH THE NORMAL PRIM PROCEDURE

| uncertainty | bandwidth |
|---|---|
| SWITCH economic growth | set([1, 2, 5]) |
| SWITCH electrification rate | set([1, 2, 5]) |
| economic lifetime ngcc | 25.00 - 39.42 |
| economic lifetime wind | 20.00 - 28.01 |
| investment proportionality constant | 0.50 - 3.82 |
| progress ratio coal | 0.90 - 1.04 |
| progress ratio gas | 0.85 - 0.99 |
| progress ratio nuclear | 0.90 - 1.04 |
| progress ratio pv | 0.76 - 0.90 |
| progress ratio wind | 0.88 - 1.00 |
| time of nuclear power plant ban | 2036.58 - 2099.99 |
| uncertainty initial gross fuel costs | 0.50 - 1.38 |
| weight factor marginal investment costs | 2.20 - 10.00 |
| weight factor technological familiarity | 1.40 - 10.00 |
| year | 0.91 - 1.10 |

Looking at Table 2 and Table 3 from a content perspective, it is noteworthy that the box found through the random boxes procedure contains various uncertainties related to the economic lifetime of the various technologies, while the normal procedure focusses on the progress ratios of these technologies instead. Both point to the same underlying technological dynamic. The lifetime determines when reinvestments are needed. Investments in technology drive the progress of the technology. We speculate that there is some interaction effects between these two sets of uncertainties: either set of uncertainties can be included in the box definition and help in explaining the cases of interest, but including one set with the other will not add much to the overall explanation.

Table 4 shows the feature scores for each of the individual uncertainties. 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. Feature scores give insight into the uncertainties that are most important in correctly classifying observations across the entire ensemble. The higher the score, the more important. The feature scores appear to follow a power law, so the first few features are the most important and the feature scores drops off quickly leaving a tail of unimportant features. This feature score can be used to identify spurious uncertainties that should be excluded from the box definition. Based on Table 4, the analyst can make a reasoned choice about which uncertainties should be used in training PRIM, and which not. It appears from this table that for this case the first five to ten uncertainties are the most important, and the rest can be ignored.

TABLE 4. FEATURE SCORES FOR THE 46 UNCERTAINTIES

| Uncertainty | Feature score | Uncertainty | Feature score |
|---|---|---|---|
| SWITCH economic growth | 0.201319 | progress ratio gas | 0.015899 |
| SWITCH electrification rate | 0.195479 | SWITCH lookup curve TGC | 0.015837 |
| SWITCH physical limits | 0.146025 | SWTICH preference carbon curve | 0.015262 |
| time of nuclear power plant ban | 0.12489 | weight factor technological familiarity | 0.013156 |
| progress ratio wind | 0.094162 | progress ratio coal | 0.012961 |
| economic lifetime wind | 0.082826 | economic lifetime hydro | 0.012089 |
| progress ratio nuclear | 0.058738 | SWITCH interconnection capacity expansion | 0.011704 |
| progress ratio biomass | 0.042281 | maximum no storage penetration rate pv | 0.011039 |
| SWITCH carbon cap | 0.039217 | economic lifetime igcc | 0.010754 |
| starting construction time | 0.032507 | SWITCH TGC obligation curve | 0.010553 |
| SWITCH storage for intermittent supply | 0.028046 | maximum battery storage uncertainty constant | 0.010231 |
| weight factor marginal investment costs | 0.026424 | investment proportionality constant | 0.010095 |
| uncertainty initial gross fuel costs | 0.02633 | progress ratio hydro | 0.009569 |
| economic lifetime nuclear | 0.025166 | economic lifetime ngcc | 0.009501 |
| investors desired excess capacity investment | 0.022952 | progress ratio igcc | 0.009137 |
| progress ratio ngcc | 0.021385 | price volatility global resource markets | 0.008174 |
| price demand elasticity factor | 0.019978 | economic lifetime coal | 0.007423 |
| economic lifetime biomass | 0.019493 | demand fuel price elasticity factor | 0.007096 |
| year | 0.01927 | economic lifetime pv | 0.006439 |
| SWITCH low reserve margin price markup | 0.019101 | weight factor technological growth potential | 0.005161 |
| economic lifetime gas | 0.018082 | SWITCH Market price determination | 0.004689 |
| SWITCH carbon price determination | 0.017075 | maximum no storage penetration rate wind | 0.003547 |
| weight factor carbon abatement | 0.016649 | | |
| progress ratio pv | 0.016577 | | |

## V. DISCUSSION

The boxes identified in this case are still quite far removed from the ideal of having both a coverage and density close to 1. This is most likely due to a complicating dependency between the various uncertainties. That is, the assumption that the cases of interest arise out of subspaces that can be described as hyper-rectangles in the model input space, might not hold [1]. This suggests that substantial gains in coverage and density could be achieved by preprocessing the data using Principal Components analysis [1]. The outlined random boxes procedure could then still be used after this pre-processing step.

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. Adaboost is an alternative to Random Forest. Like Random Forest, Adaboost is an ensemble method. In contrast to Random Forest, each ensemble member is generating 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 straight forward manner.

Both the improvement suggested above and the improvement explored in this paper keep the peeling and pasting procedure used in PRIM intact. However, 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 the coverage and density 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.

## VI. CONCLUSION

In this paper we explored a way of improving PRIM, which is the dominant algorithm currently used for scenario discovery. We drew inspiration from work on Random Forests. 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 features is considered. 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. The question we explored was whether random feature selection and bagging can be combined with PRIM and whether the resulting algorithm would outperform normal PRIM. From our analysis, we conclude that the resulting random boxes approach does indeed outperform PRIM. That is, we have shown that it is possible to improve on the results found through normal PRIM by adapting a random boxes approach.

In the case study, we used a System Dynamics model of the European energy system and we explored the impact of the emission trading system on the future evolution of the energy system in terms of carbon emissions and the share of renewables in the overall energy mix. We generated a training set of 5000 computational experiments and a test set of 2000 computational experiments that cover the space spanned by 46 uncertainties associated with the European energy system. We performed scenario discovery using both normal PRIM and the random boxes approach and compared the results. We found that the best boxes in terms of both coverage and density identified through the random boxes approach dominated the best boxes identified through normal PRIM. When we compared two candidate boxes, one from each, we observed that the boxes shared only a third of the uncertainties. This implies that the solution found by PRIM was a local optima, while the random boxes approach was able to find a better solution. This also confirmed our suspicion, voiced in the introduction, that the lenient hill climbing optimization procedure used by normal PRIM can suffer from the usual defects of such optimization procedures like local optima.

In the case study we also calculated feature scores using the ensemble of random boxes. These feature scores give insight into the relative importance of the different uncertainties in classifying the results. The higher the feature score, the more important. These feature scores can be used in selecting which uncertainties should be included and which uncertainties are spurious. Low scores uncertainties should not be included in the definition of the box.

In scenario discovery, there are two important issues. The first is the interpretability of the results. The presented random boxes approach does help in interpretability through both the feature scores and the identification of the Pareto front. The 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, helps in finding high coverage high density boxes. The second important issue in scenario discovery is the inclusion of spurious uncertainties in box definitions. The feature scores over an additional tool, in addition to quasi p-values, that analysts can use to mitigate this problem.

The analysis in this paper is based on only a single case. 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. Another direction for future work 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. Besides improving the random boxes procedure introduced in this paper, it could be of great value to explore other avenues for improving on normal PRIM. Noteworthy directions here include a more direct global optimization procedure instead of the lenient hill climbing currently used by PRIM, and adapting Adaboost to work with PRIM.

## REFERENCES


[1] S. Dalal, B. Han, R. Lempert, A. Jaycocks, and A. Hackbarth, "Improving Scenario Discovery using Orthogonal Rotations," *Environmental Modelling & Software,* vol. 48, pp. 49-64, 2013.

[2] S. C. Bankes, W. E. Walker, and J. H. Kwakkel, "Exploratory Modeling and Analysis," in *Encyclopedia of Operations Research and Management Science*, S. Gass and M. C. Fu, Eds., 3rd ed. Berlin, Germany: Springer, 2013.

[3] B. P. Bryant and R. J. Lempert, "Thinking Inside the Box: a participatory computer-assisted approach to scenario discovery," *Technological Forecasting and Social Change,* vol. 77, pp. 34-49, 2010.

[4] J. H. Kwakkel, W. L. Auping, and E. Pruyt, "Dynamic scenario discovery under deep uncertainty: the future of copper," *Technological Forecasting and Social Change,* vol. 80, pp. 789-800, 2013.

[5] National Research Council, *Informing decisions in a changing climate*: National Academy Press, 2009.

[6] S. A. van 't Klooster and M. B. A. van Asselt, "Practising the scenario-axes technique," *Futures,* vol. 38, pp. 15-30, 2006.

[7] J. Derbyshire and G. Wright, "Preparing for the future: Development of an 'antifagile' methodology that complements scenario planning by omitting causation," *Technological Forecasting and Social Change,* 2013.

[8] P. W. F. van Notten, A. M. Sleegers, and M. B. A. van Asselt, "The future shocks: on discontinuity and scenario development," *Technological Forecasting and Social Change,* vol. 72, pp. 175-194, 2005.

[9] T. J. B. M. Postma and F. Liebl, "How to improve scenario analysis as a strategic management tool?," *Technological Forecasting and Social Change,* vol. 72, pp. 161-173, 2005.

[10] R. J. Lempert, S. Popper, and S. Bankes, "Shaping the Next One Hundred Years: New Methods for Quantitative, Long Term Policy Analysis," RAND, Santa Monica, CA, USA Report MR-1626-RPC, 2003.

[11] S. Hallegatte, A. Shah, R. Lempert, C. Brown, and S. Gill, "Investment Decision Making Under Deep Uncertainty: Application to Climate Change," The World Bank Policy Research Working Paper 6193, 2012.

[12] J. H. Kwakkel, W. E. Walker, and V. A. W. J. Marchau, "Classifying and communicating uncertainties in model-based policy analysis," *International Journal of Technology, Policy and Management,* vol. 10, pp. 299-315, 2010.

[13] J. Rozenberg, C. Guivarch, R. J. Lempert, and S. Hallegatte, "Building SSPs for climate policy analysis: a scenario elicitation methodology to map the space of possible future challenges to mitigation and adaptation," *Climatic Change,* 2013.

[14] M. D. Gerst, P. Wang, and M. E. Borsuk, "Discovering plausible energy and economic futures under global change using multidimensional scenario discovery," *Environmental Modelling & Software,* 2013.

[15] R. J. Lempert, D. G. Groves, S. Popper, and S. Bankes, "A general analytic method for generating robust strategies and narrative scenarios," *Management Science,* vol. 52, pp. 541-528, 2006.

[16] R. J. Lempert and M. Collins, "Managing the Risk of Uncertain Threshold Response: Comparison of Robust, Optimum, and Precautionary Approaches," *Risk Analysis,* vol. 24, pp. 1009-1026, 2007.

[17] C. Hamarat, J. H. Kwakkel, and E. Pruyt, "Adaptive Robust Design under Deep Uncertainty," *Technological Forecasting and Social Change,* vol. 80, pp. 408-418, 2013.

[18] W. E. Walker, M. Haasnoot, and J. H. Kwakkel, "Adapt or perish: a review of planning approaches for adaptation under deep uncertainty," *Sustainability,* vol. 5, pp. 955-979, 2013.

[19] J. H. Friedman and N. I. Fisher, "Bump hunting in high-dimensional data," *Statistics and Computing,* vol. 9, pp. 123-143, 1999.

[20] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Monterey, CA: Wadsworth, 1984.

[21] R. J. Lempert, B. P. Bryant, and S. Bankes, "Comparing Algorithms for Scenario Discovery," RAND, Santa Monica, CA, USA WR-557-NSF, 2008.

[22] B. P. Bryant. (2012). *sdtoolkit: Scenario Discovery Tools to Support Robust Decision Making*. Available: http://cran.r-project.org/web/packages/sdtoolkit/index.html

[23] A. M. Parker, S. V. Srinivasan, R. J. Lempert, and S. H. Berry, "Evaluating simulation-derived scenarios for effective decision support," *Technological Forecasting and Social Change,* 2014.

[24] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, pp. 5-23, 2001.

[25] L. Breiman, "Bagging Predictors," *Machine Learning,* vol. 24, pp. 123-140, 1996.

[26] J. R. Quinlan, "Introduction to decision trees," *Machine Learning,* vol. 1, pp. 81-106, 1986.

[27] J. R. Quinlan, *C4.5: Programs for machine learning*: Morgan Kaufmann Publsihers, 1993.

[28] T. G. Dietterich, "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization," *Machine Learning,* vol. 40, pp. 139-157, 2000.

[29] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, pp. 832-844, 1998.

[30] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural Computation,* vol. 9, pp. 1545-1588, 1997.

[31] European Commission, "Europe 2020: A European Strategy for smart, sustainable and inclusive growth," *European Commission, COM (3.3. 2010),* 2010.

[32] J. W. Forrester, *Industrial Dynamics*. Cambridge: MIT Press, 1961.

[33] J. D. Sterman, *Business Dynamics: Systems Thinking and Modeling for a Complex World*: McGraw-Hill, 2000.

[34] E. Pruyt. (2013). *Small System Dynamics Models for Big Issues: Triple Jump towards Real-World Complexity*.

[35] Ventana Systems Inc., "Vensim Reference Manual," ed: Ventana System Inc., 2011.

[36] E. Loonen, "Exploring Carbon Pathways in the EU Power Sector, Using Exploratory System Dynamics Modelling and Analysis to assess energy policy regimes under deep uncertainty," Master Master, Faculty of Technology, Policy and Management, Delft University of Technology, Delft 2012.