Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes
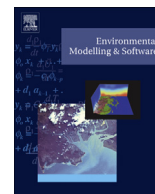
Kwakkel, JH; Jaxa-Rozen, M.

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

Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes☆

Jan H. Kwakkel[*], Marc Jaxa-Rozen

Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX, Delft, The Netherlands

ABSTRACT

Scenario discovery is a novel model-based approach to scenario development in the presence of deep uncertainty. Scenario discovery frequently relies on the Patient Rule Induction Method (PRIM). PRIM identifies regions in the model input space that are highly predictive of producing model outcomes that are of interest. To identify these, PRIM uses a lenient hill climbing optimization procedure. PRIM struggles when confronted with cases where the uncertain factors are a mix of data types, and can be used only for binary classifications. We compare two more lenient objective functions which both address the first problem, and an alternative objective function using Gini impurity which addresses the second problem. We assess the efficacy of the modification using previously published cases. Both modifications are effective. The more lenient objective functions produce better descriptions of the data, while the Gini impurity objective function allows PRIM to be used when handling multinomial classified data.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Software

This paper makes use of the Exploratory Modeling Workbench, available via https://github.com/quaquel/EMAworkbench. Section 3.4 relies on extensions to classes available in the workbench. These extensions are provided as supplementary material. The detailed code with rudimentary documentation is provided in the form of 3 pdf representations of the underlying IPython notebooks.

1. Introduction

Scenario discovery is a relatively novel approach aimed at addressing the challenges of characterizing and communicating deep uncertainty associated with simulation models (Dalal et al., 2013). The basic idea is that the consequences of the various deep uncertainties encountered in a model-based decision support exercise are systematically explored through series of computational experiments (Bankes et al., 2013). These computational experiments are designed to exhaustively sample the space spanned by the various deeply uncertain factors. The results of the set of computational experiments are analyzed to identify regions in the uncertainty space that are of interest (Bryant and Lempert, 2010; Kwakkel et al., 2013). These identified regions can subsequently be communicated as scenarios.

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 (Bryant and Lempert, 2010; van't Klooster and van Asselt, 2006). Rather than trying to achieve consensus or facilitate a process of joint sense-making to resolve the differences between worldviews, scenario discovery aims at making transparent which uncertain factors actually make a difference for the decision problem at hand. Another shortcoming identified in the evaluative literature is that scenario development processes have a tendency to overlook surprising developments and discontinuities (Derbyshire and Wright, 2014; van Notten et al., 2005). This might be at least partly due to the fact that many scenario approaches move from a large set of relevant uncertain factors to a smaller set of drivers or "megatrends". 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 scenario logic approaches.

Although scenario discovery can be applied on its own (Gerst et al., 2013; Kwakkel et al., 2013; Rozenberg et al., 2013), it is also a key step in Robust Decision Making (RDM) (Dalal et al., 2013; Hamarat et al., 2013; Lempert and Collins, 2007; Lempert et al., 2006). 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 case of scenario discovery suggests that it could be used also 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).

Currently, the main statistical rule induction algorithm that is used for scenario discovery is the Patient Rule Induction Method (PRIM) (Friedman and Fisher, 1999), although other algorithms such as Classification and Regression Trees (CART) (Breiman et al., 1984) are sometimes used (Gerst et al., 2013; Lempert et al., 2008). PRIM aims at finding combinations of values for the uncertain input variables that result in similar characteristic values for the outcome variables. Specifically, PRIM seeks a set of subspaces of the uncertainty space within which the value of a single output variable is considerably different from its average value over the entire domain. PRIM describes these subspaces in the form of hyper rectangular boxes of the uncertainty space. To identify these subspaces, PRIM uses a lenient or patient, as opposed to greedy, hill climbing optimization procedure. In the context of scenario discovery, the outcome variable is typically a binary variable denoting whether a given set of inputs is of interest or not. The hyper rectangular boxes identified by PRIM are not always the best description of the combination of input variables that produces similar characteristic values for the outcome variables. Sometimes, these characteristic values are grouped along another axes than the set of uncertain input variables. Preprocessing the data using principal components analysis can help to identify such axes and rotate the data (Dalal et al., 2013). 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 (Bryant, 2014). A Python implementation of PRIM, including support for the PCA preprocessing, is available as part of the Exploratory Modeling Workbench (Kwakkel and Pruyt, 2015).

There are two problems related to PRIM that are addressed in this paper. First, although originally presented as a regression based rule induction algorithm, in the context of scenario discovery PRIM is typically used on a binary classification of the data. In contrast to e.g. CART, PRIM cannot be used directly for handling the situation where the output data is classified using more than two classes (Gerst et al., 2013; Rozenberg et al., 2013). Second, when the uncertain factors are represented by integers or categories, the lenient hill climbing optimization procedure used in PRIM needs to account for this. Friedman and Fisher (1999) offer several suggestions for adapting the objective function used by PRIM to account for this. Both the scenario toolkit, and the Python implementation include these modified objective functions. However, to date the efficacy of these alternative objective functions has not been systematically evaluated in the context of scenario discovery. We address both problems in this paper because their solutions are both closely related with the lenient hill climbing optimization approach used in PRIM.

To address these two problems, we first outline in Section 2 in more detail the PRIM algorithm. We will discuss the suggestions of

Friedman and Fisher (1999) for handling integer and categorical data in evaluating the next possible steps of the algorithm. To address the problem of multinomial classified data, we draw on the way in which CART handles this and show how by adapting the objective function used by PRIM, it can be made applicable also to problems where the data is classified using multinomial classification. The resulting modifications to PRIM do not affect the efficacy of preprocessing steps such as employed in PCA-PRIM (Dalal et al., 2013). We provide an open source implementation in Python for this modified version of PRIM.

In Section 3, we assess the efficacy of alternative ways of accounting for categorical and discrete data in the objective function used by PRIM. In particular, we apply it to the same data as used in the original paper of Bryant and Lempert (2010), 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. In Section 4, we explore the objective function for handling multinomial classified data and compare it to both CART and a sequential PRIM approach. For this we use the case study of Rozenberg et al. (2013). A discussion of the results is presented in Section 5 and the conclusions are presented in Section 6.

2. Method

2.1. PRIM

Fig. 1 offers a visual explanation of the PRIM algorithm. In the top left corner we see the dataset. The dataset consists of 110 computational experiments, 30 of which are of interest. Each experiment is described by two variables. The first variable, U_1, is a categorical variable and the possible values are $\{a,b,c\}$. The second variable, U_2, is a continuous variable ranging between 0 and 2. Together, U_1 and U_2 span the uncertainty space. We use PRIM to find an orthogonal subspace, or box, within the uncertainty space that has a high concentration of experiments of interest.

PRIM starts with an initial box B_1 that covers all of the data. Next, the size of this box is recursively reduced. Reducing the size of the box is done by removing a small slice of data along one of the dimensions. To find the best slice of data to remove, the algorithm first enumerates all possible slices, b_j, that can be removed, and next uses an objective function to determine the best possible slice to remove. This results in a new box B_l. The series of boxes resulting from this recursive peeling is also known as the peeling trajectory.

How does PRIM enumerate all the possible slices of data that can be removed? PRIM will only remove data along a single dimension. So, for each dimension, PRIM enumerates all the possibilities. The exact possibilities depend on the data type of the dimension. In the example given in Fig. 1, we have two different data types. U_1 is a categorical variable. In this case, PRIM will consider the removal of each of the individual categories.[1] In our example, this means that there are three alternative slices of data that PRIM considers for removal for this dimension. U_2 is a continuous variable. In this case, PRIM will consider the removal of a small slice from the top and a small slice from the bottom. Continuous variables will thus contribute two alternative slices of data that PRIM will consider for removal. The same is true for integer data.

---

[1] The Python implementation of PRIM, available as part of the EMA workbench, follows the outlined approach for the generation of candidate boxes. Categorical variables are not presently supported by the R implementation in the Scenario discovery toolkit. For more details on the consequences of this, see the supplementary material.
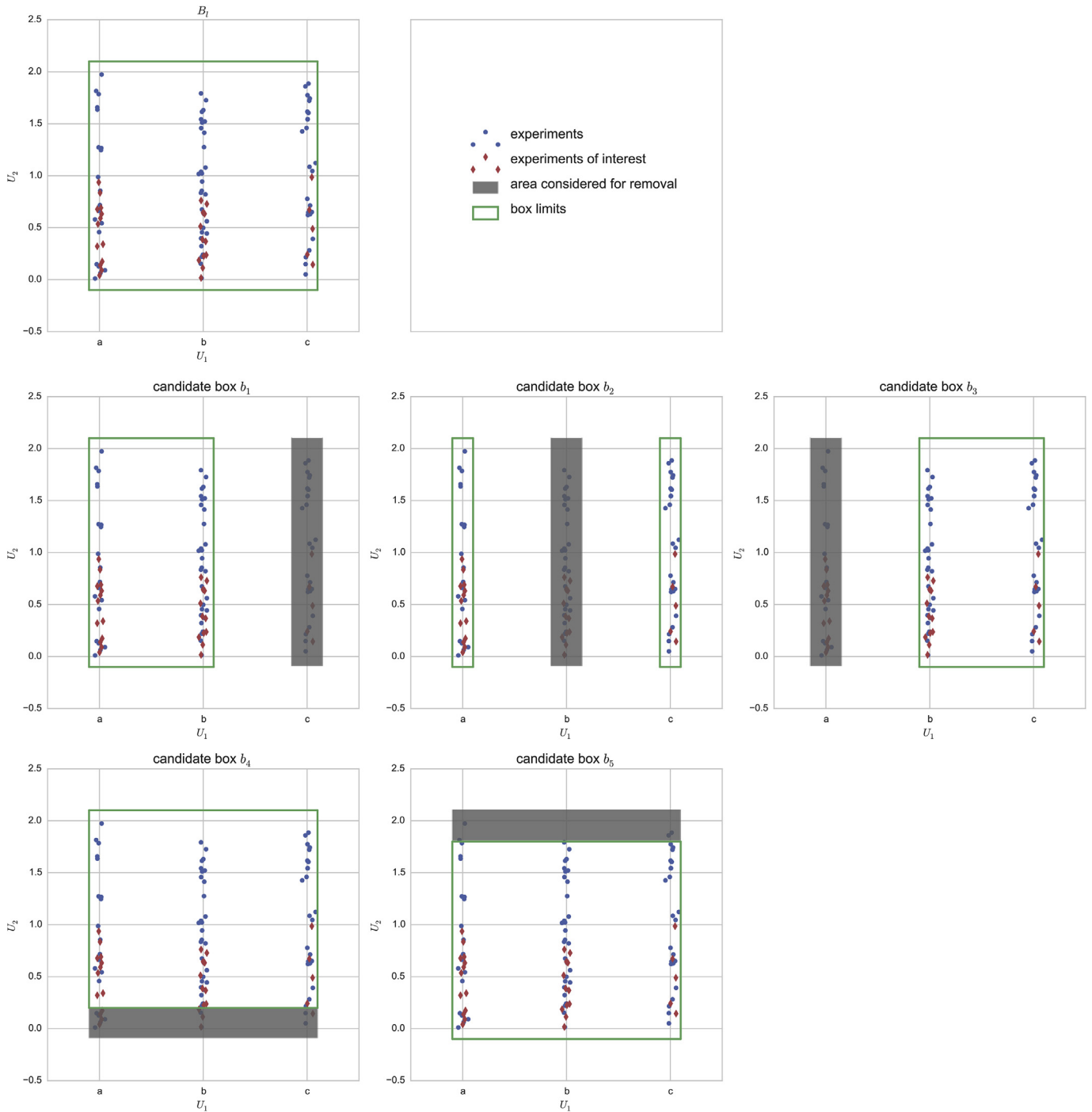
Fig. 1. Visual explanation of PRIM. For visualization purposes, the data has been given a small Gaussian offset on the U_2 axis.

Once all possible slices b_j have been enumerated, the next step of the algorithm is to determine the best one. For this PRIM uses an objective function. Different criteria can be used for selecting the best slice b^* from the all candidate slices b_j that are eligible for removal. Following (Friedman and Fisher, 1999), the simplest criterion is to choose the slice b^* that has the largest output mean value for the new box resulting from removing b from B_l.

$$b^* = \arg \max_{b \in C(b)} \bar{y}_{B_l - b} \qquad (1)$$

However, in case of heterogeneously typed data, this simple

criterion is flawed. The mean value for a new box resulting from the removal of slice b_j in case of a continuous variable will typically be based on more data points than for a categorical variable. As a result, the categorical variables will dominate the search process, potentially producing inferior boxes. To correct for this, the number of data points has to be taken into consideration as well. This can be done in various ways. Friedman and Fisher (1999) suggest two more lenient criteria. A first more lenient criterion is

$$b^* = \arg \max_{b \in C(b)} \frac{\overline{y}_{B_l - b} - \overline{y}_{B_l}}{\beta_{B_l} - \beta_{B_l - b}} \tag{2}$$

where β_{B_l} is the number of data points in box B_l, $\beta_{B_l - b}$ is the number of data points in the box resulting from removing b from B_l, and $\overline{y}_{B_l}$ is the average of the data in box B_l. This criterion measures the increase in the mean divided by the amount of data removed. A second, even more lenient criterion also suggested by Friedman and Fisher (1999) is

$$b^* = \arg \max_{b \in C(b)} \frac{\beta_{B_l - b} * (\overline{y}_{B_l - b} - \overline{y}_{B_l})}{\beta_{B_l} - \beta_{B_l - b}} \tag{3}$$

That is, the change in mean times the number of data points in the box resulting from removing b from B_l, divided by the change in the number of data points. This even more lenient criterion not only accounts for the increase in the mean per removed number of data points, but also takes into account the number of data points remaining.

Returning to our example, there are two candidate boxes that stand out as being potentially good boxes to select. These are b_1 and b_5. Fig. 2 shows both boxes, as well as the calculation of the scores for objective functions 1 and 2. If we were using the first objective function, the algorithm would choose box b_1. This would entail removing a substantial amount of the data, including 5 experiments of interest. In contrast, if we were using the second, more lenient objective function, the algorithm would choose box b_5 instead. The same is true for the third even more lenient objective function. As this example demonstrates, in case of the simplest objective function, categorical variables can dominate the search. In Section 3, we will compare the performance of PRIM using the first and second objective functions for three distinct datasets and compare the results. For the third dataset, we will also include the third even more lenient objective function. Given the characteristics of the three cases, it is expected that the third objective function will only make a difference for the third case.

The PRIM algorithm recursively removes small slices of data, resulting in a peeling trajectory. In the context of scenario discovery, the analyst selects an appropriate box using 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 the box. Density is the fraction of cases within the box that is of interest. Interpretability is a more difficult objective. Typically, the number of uncertain factors that make up the box definition is used as a proxy for interpretability. Below, we will use a trade-off curve between coverage and density for the peeling trajectory for comparing the original objective function with the more lenient objective function. We are thus not explicitly considering the interpretability objective in the comparison, although the information is available in the supplementary material.

2.2. Handling multinomial classified outcomes

A separate problem is that PRIM cannot directly handle the situation where the output variable contains multiple classes. Scenario discovery was initially proposed to identify the conditions under which a proposed policy fails to meet its goals (Bryant and Lempert, 2010). This implies a binary classification on the basis of model outcomes. Increasingly, scenario discovery is used more broadly for supporting the model-based bottom up development of scenarios (Gerst et al., 2013; Kwakkel et al., 2013; Rozenberg et al., 2013). In such a context, it can happen that the outputs of the series of computational experiments are clustered into more than two classes. In order to perform scenario discovery on such a data set, the current practice is to turn this data set into a binary classification for each class and apply PRIM to each data set separately (Gerst et al., 2013; Rozenberg et al., 2013). This is a pragmatic solution that is easy to carry out.

There are two downsides to this approach. First, the pragmatic approach can produce results where the boxes overlap. The box explaining the occurrence of class A might have substantial overlap with the box found for explaining the occurrence of class B. So,
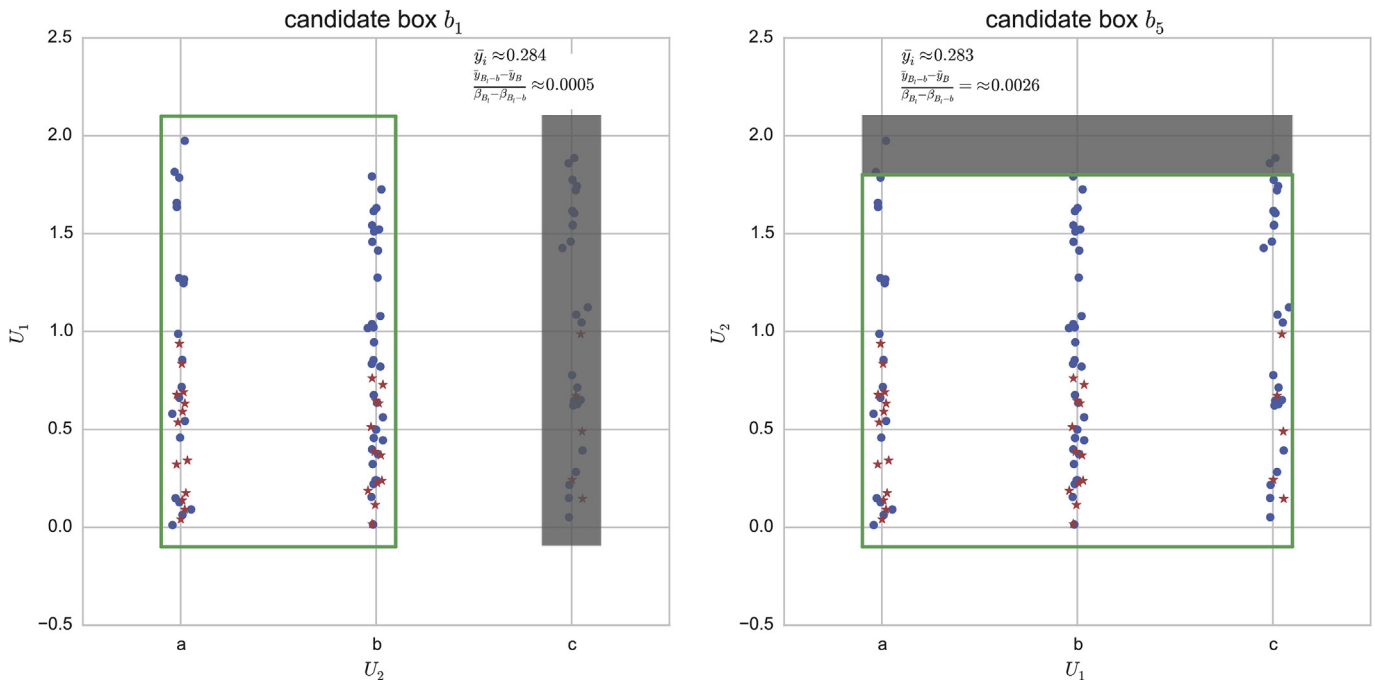

Fig. 2. The two alternative objective functions applied to our example.

rather than having two distinct explanations for the occurrence of class A or class B, one would end up with a partially overlapping explanation. For this overlapping area, we cannot tell whether a given case belongs to either class A or B. Moreover, it is not trivial to describe the overlapping area, or at least, this is not directly supported by any of the currently available software for scenario discovery. The severity of this drawback depends on the analysis. In a policy analytic context where the purpose is to develop robust policies, this might not be a severe drawback. In contrast, if the purpose is to develop narrative scenarios from the boxes identified by PRIM, this might become an issue. Because PRIM is used for each class separately, the algorithm is not forced to find boxes of e.g. class A, while simultaneously accounting for all the other classes as well. A second downside is that the pragmatic approach requires performing several PRIM analyses, one for each category. If PRIM can be used for multinomial classified data, a single PRIM analysis can be used instead.

In contrast to PRIM, the CART algorithm can be used for both regression as well as classification. In case of classification, CART uses Gini impurity to determine the best split at every level in the classification tree.

$$I_G(f) = 1 - \sum_{i=1}^{m} f_i^2 \qquad (4)$$

where i is a value from $\{1,2,\ldots,m\}$ and f_i is the fraction of items in the set labeled with i. Gini impurity is zero if all items have the same value i. The more heterogeneous the set is, the closer to 1 the Gini impurity will be. We can adapt Gini impurity for use in PRIM. As an analogue to the normal objective function shown in equation (1), we could pick the box with the lowest Gini impurity. However, this would suffer from the same drawback in case of categorical variables. Therefore, a better option is to take the reduction in impurity divided by the number of data points removed

$$b^* = \arg \max_{b \in C(b)} \frac{I_{B_l} - I_{B_l-b}}{\beta_{B_l} - \beta_{B_l-b}} \qquad (5)$$

where I_{B_l} is the Gini impurity of box B_l, and I_{B_l-b} is the Gini impurity of the box resulting from removing b from B_l.

3. Application of the alternative objective functions

3.1. Case 1: continuous uncertain factors

Bryant and Lempert (2010) demonstrated scenario discovery by investigating the potential impact of a policy which aims at 25% renewable sources for electricity and motor fuels by 2025 in the United States, on greenhouse gas emissions and economic costs. 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. They used an existing simulation model (Toman et al., 2008) which calculates the economic costs and greenhouse gas emissions given assumptions about 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. All 9 uncertain factors are continuous variables. The resulting dataset as used both here and in the original work of Bryant and Lempert (2010) contains 882 cases.

To compare the original objective function with the more lenient objective function (2), we apply PRIM to the dataset using both objective functions. Next, we compare the peeling trajectories resulting from both objective functions. Given that this case uses

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 gallons per ton
Oil supply elasticity	0.2—06
Transportation demand elasticity	−0.2 to −0.8
Electricity co-product	0—2 kW h per gallon
Shift in oil supply curve	−10 to 10% change
Biomass backstop price	90—200 $ per ton

continuous uncertain factors, we expect that the change of objective function will not yield different results. As can be seen in Fig. 3, this is indeed the case. The coverage and density for both objective functions are identical for a very large part of the peeling trajectory. The only difference is when the density reaches 1. The lenient objective function in this case stops peeling, while the original criterion continues. For practical applications of PRIM in the context of scenario discovery this is immaterial.

3.2. Case 2: integer uncertain factors

Rozenberg et al. (2013) use scenario discovery for the bottom up development of Shared Socio-economic Pathways (SSPs). The SSP's have been developed over the last few years following a scenario logic approach. Rozenberg et al. (2013) argue that such a top down approach to scenario development might miss important driving forces for the different SSPs. Hence, they suggest a bottom up 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) used the IMACLIM-R model (Sassi et al., 2010), which 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. IMACLIM-R has
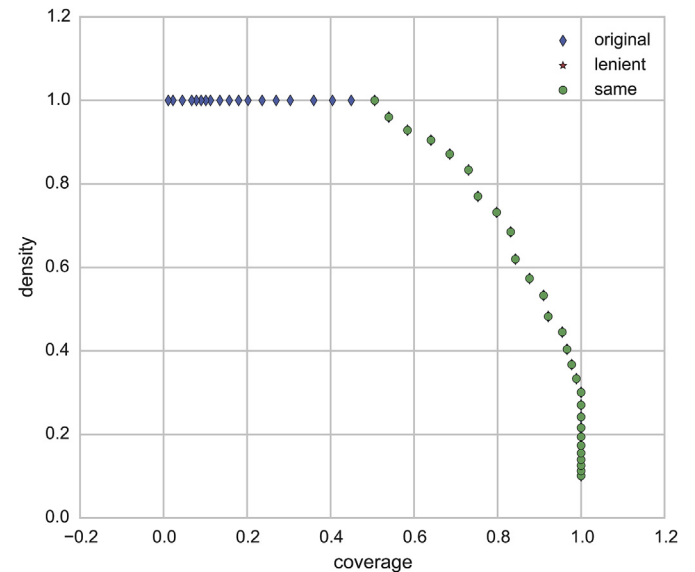

Fig. 3. Peeling trajectories for the Bryant and Lempert case for both objective functions.

many exogenous parameters and trends that can be varied simultaneously. To address the potential combinatoric explosion arising from this, Rozenberg et al. (2013) created four groups of related parameters: globalization, environmental stress, carbon dependence, and equity. For each of the groups, several internally consistent sets of alternative assumptions about parameter values are formulated, and a full factorial analysis on these sets is performed resulting in 286 cases. IMACLIM-R is run for each of these cases and the outcomes are scored in terms of challenges to mitigation and challenges to adaptation. To allow a comparison with the existing SSP's, a rule-based clustering of the cases is used such that a large majority of cases can uniquely be assigned to represent one of five existing SSPs. Those cases that cannot be uniquely assigned to any of the five existing SSPs are allocated to a sixth class instead. Next, PRIM is used to discover the key drivers for each of the five clusters.

In order to compare the original objective function and the first more lenient alternative (2), we apply PRIM to each of the five classes using both objective functions. The resulting peeling trajectory for each of the five classes is shown in Fig. 4. Similar to the previous case, we do not expect major differences between both objective functions. The analysis confirms this. We only see minor differences for classes 1, 3, and 5. Classes 2 and 4 are identical, until density reaches 1. Just like with the continuous uncertain factors, the lenient criterion stops earlier than the original criterion.

Looking at the peeling trajectories for classes 1, 3, and 5, we can observe the difference between both objective functions. For example, in the lower right corner of the peeling trajectory for class 5, we see that the lenient criterion declines more slowly in coverage than the original criterion. In contrast, for class 3 we see that the original criterion is greedier in increasing density at the expense of coverage. Which is better depends on the specific application, but if we look at the peeling trajectory for class 1, we see that the lenient criterion finds a candidate box with coverage just below 0.6 and density just below 0.8. The nearest alternative, as found by the original criterion, trades around 10% in coverage for substantially less than 10% increase in density.

3.3. Case 3: continuous, integer, and categorical uncertain factors

Hamarat et al. (2014) use scenario discovery for the development of an adaptive policy for steering the European energy system towards more sustainable functioning. The adaptive policy is built on the current emission-trading scheme (ETS). A System Dynamics (Forrester, 1961; Sterman, 2000) model representing the EU energy system including interconnections and congestion is used. In this model, the EU is split in seven regions and nine power generation technologies are included, ranging from conventional coal to wind and solar. 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_2 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 2. The uncertain factors are a mix of continuous variables, integer variables, and categorical variables. For the purpose of this paper, we created a new set of 20,000 cases using Latin hypercube sampling. Next, scenario discovery is used to reveal combinations of uncertain factors under which ETS performs poorly.

Fig. 5 shows the peeling trajectories for all three objective functions. Only the starting point, not surprisingly, is identical. The remainder of the peeling trajectories is quite different. First, the original objective function contains fewer points, with wider gaps on the coverage axis. This implies that the original objective function prefers to peel along categorical uncertain factors. This confirms the claim that the simple objective function can result in

categorical variables dominating the search. Peeling along these categorical variables offers most gain, and this objective function ignores the loss of data points remaining in the box required to achieve this gain. In contrast, both lenient criteria tend to peel along continuous variables in a lenient fashion, with a few exceptions. Both lenient criteria will peel along a categorical uncertain factor only when the loss in data points is worth it in light of the gain in density. For this particular dataset, both lenient objective functions produce a peeling trajectory that scores better in terms of both coverage and density than the peeling trajectory resulting from the original criterion. That is, either lenient criterion is better than the original objective function. If we compare the two lenient criteria, there are some minute differences. Where objective function number 2 (lenient 1 in the figure) peels along a categorical dimension, the more lenient objective function 3 (lenient 2 in the figure) holds on to this for a while longer. Based on this one case, however, there is no clearly superior lenient criterion to use. Friedman and Fisher (1999) suggest combining the results of both criteria by including only the dominant points, and allowing the analyst to make his or her own trade off.

4. Gini-based objective function for multinomial classified output data

In order to assess the efficacy of the Gini impurity objective function for handling situations where one encounters multinomial classified data with PRIM, we use the same case as in Section 3.2. For comparative purposes, we apply PRIM sequentially to each class as shown in Section 3.2, PRIM with the Gini-based objective function, and CART. Given that coverage and density cannot properly be defined in case of multinomial classification, we used the tradeoff between impurity and mass (i.e. the fraction of cases remaining inside a box) as a basis for selecting appropriate boxes in the case of the Gini-based objective function (see the online supplementary material). We have parameterized CART such that the terminal leaves are at least equal in terms of the number of data points they contain as PRIM.

The results for sequential PRIM are shown in Table 3, Table 4 shows the results for the Gini objective function in PRIM, and Table 5 shows the results for CART. These tables show the boxes as rows. For each box, in gray the table specifies the uncertain factors that are used to define that box. If no gray is used for a given uncertain factor, this means that the box found does not restrict this uncertain factor. So, in Table 3, the box for class 1 is specified by limits on inequalities, population, and behaviors. We also show how many cases of each class are present in each box (the box composition), and the associated Gini impurity of the box. Note that the case contains 5 classes of interest number 1 to 5, and a 6th class for the remaining cases.

If we look at each of the analyses in isolation, we observe that sequential PRIM is able to offer succinct explanations for each of the five classes of interest. Moreover, there is no evidence of box overlap. For example, the boxes for class 1 and 2 are identical with respect to the limits on inequalities and population, but different with respect to behaviors. Similarly, classes 3 and 4 are identical on inequalities, overlap on convergence and population, but again are distinct on behaviors. Class 3 and 5 are different on inequalities. If we look at the quality of the boxes in terms of how many cases of interest they contain, and their Gini impurity, we see that most boxes contain contamination in the form of class 6. Only for class 4 do we have a really excellent box that contains virtually all cases of interest. In contrast, for class 2, we have a box than only contains 13 of the 28 cases of interest. Gini PRIM produces some good explanations, most notably for class 3 (box 1) and class 1 (box 3). It fails to provide an explanation for class 4, however, and a clear
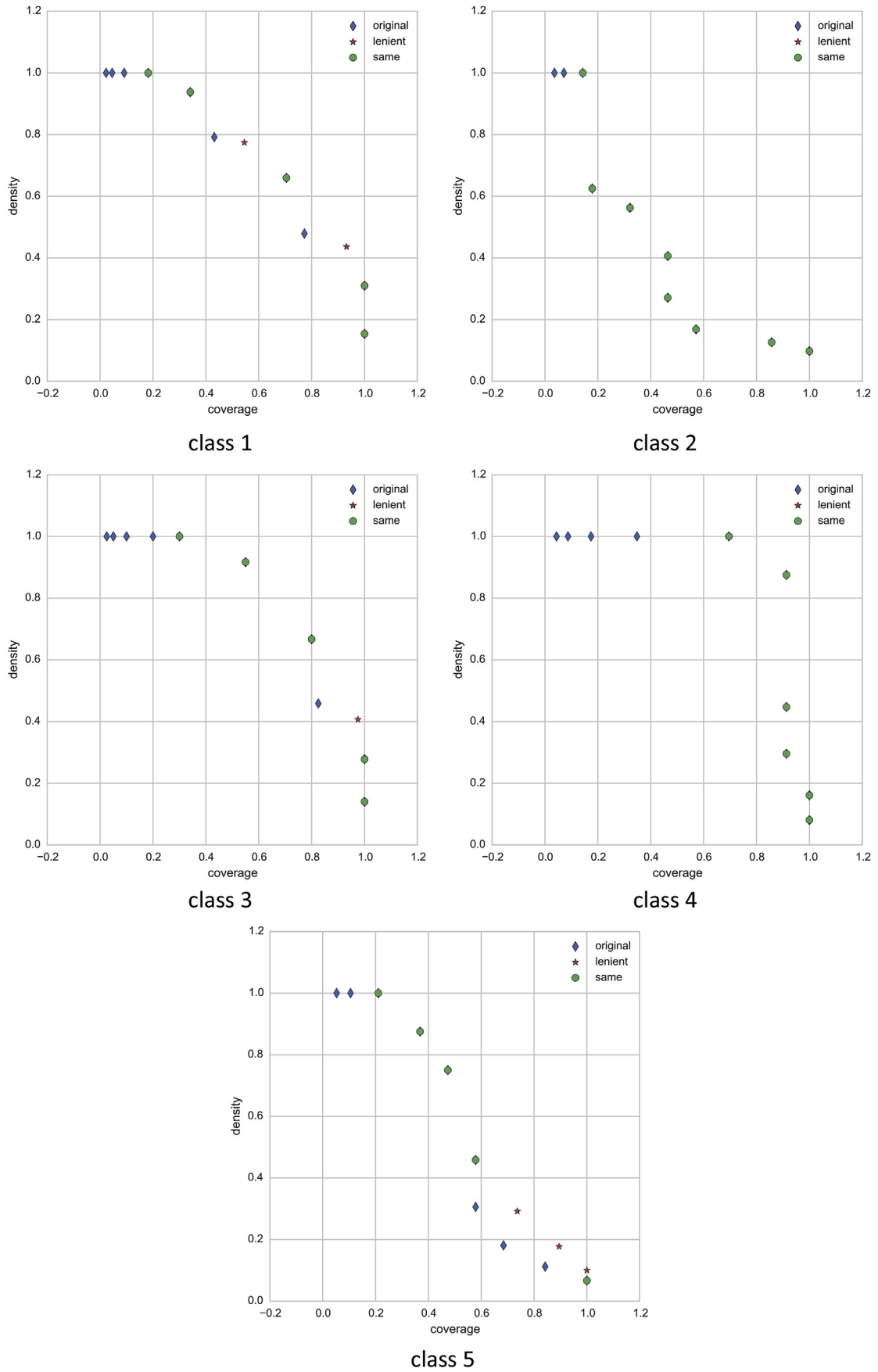
Fig. 4. Comparison of peeling trajectories for all five classes, with the classes numbered from left to right and top to bottom.

Table 2
High level summary of uncertainties to be explored.

| Name	Description	Variable type
Economic lifetime	The average lifetime of nine energy generation technologies	Continuous
Learning curve	The future performance increase of nine energy generation technologies	Continuous
Economic growth	Six alternative economic growth scenarios are considered	Categorical
Electrification rate	The rate of electrification of the economy is explored by means of six different electrification trends	Categorical
Physical limits	The effect of physical limits on the penetration rate of a technology is unknown. Two different behaviors are considered for each of the nine technologies	Categorical
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	Continuous
(Battery) storage	Storage of energy mainly in relation to sustainable energy generation technologies	Continuous
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.	Integer
Price − demand elasticity	The elasticity of demand to price changes	Continuous
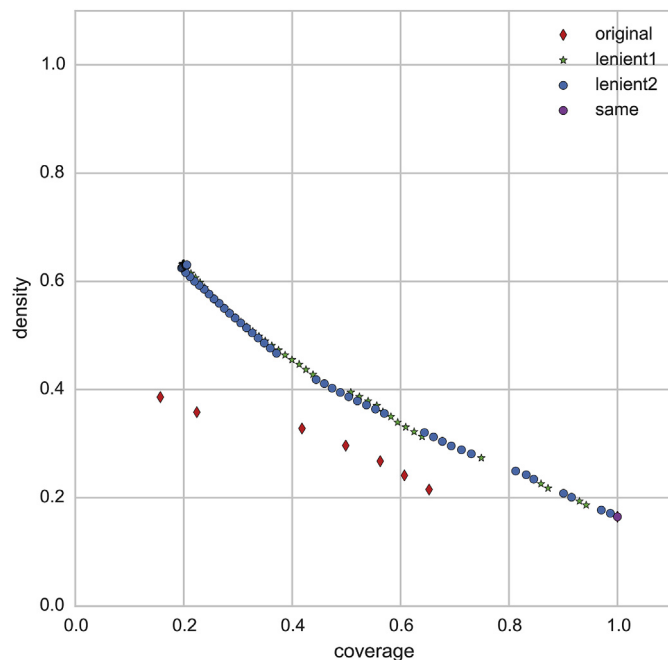
Fig. 5. Comparison of peeling trajectories for the original objective function and the lenient objective function for the Hamarat et al. dataset.

explanation is also absent for class 2 and 5. The same is true for CART. CART is able to offer interesting explanations for class 3 (box 3), class 1 (box 11, 12 and 13), and class 4 (box 8). Another interesting insight that follows from CART is that the behavior parameter largely determines whether a given experiment belongs to either class 1 or 2, or to classes 3 and 4. Class 5 is possible for both values, but occurs more frequently if the value of behaviors is 0.

Comparing the three analyses we see some interesting similarities. Gini PRIM and CART produce the exact same box definition for class 3. Sequential PRIM and CART produce the exact same box definition for class 4. If we compare the boxes identified through sequential PRIM, Gini PRIM, and CART, we observe no major difference in terms of quality as measured by Gini impurity. The three approaches find both relatively pure boxes, but also several relatively impure boxes. Looking at the boxes with low Gini impurity as found by CART, we see that most of those contain only a very limited number of data points. For example, box 13 has a Gini impurity of 0.12. However, this box only contains 16 of the 286 data points. An advantage of the Gini based PRIM analysis, in contrast to

CART, is the ease with which an analyst can make a decision on the tradeoff between a decrease of Gini impurity and a loss of mass. At the same time, this is a drawback. In this example, the user will have to produce several boxes to adequately describe the dataset. In contrast CART produces an exhaustive partitioning of the dataset in a single run of the algorithm.

The results from Gini PRIM demonstrate that it is possible to extend PRIM so that it can be used for multinomial classified data. It is able to offer a good explanation for at least some of the classes. At the same time, the comparison with both sequential PRIM and CART has brought out some of the downsides as well. Sequential PRIM can find explanations for each of the classes, although of differing quality. CART is easy to use. It will generate an exhaustive partitioning of the data and can produce good explanations for several classes. Together, this suggests that the choice between sequential PRIM, Gini PRIM, and CART will at least partly come down to personal preferences of the analyst and the specifics of the case. Rather than picking one or the other algorithm, a better approach might be to use all three approaches simultaneously and combine the insights resulting from both.

5. Discussion

We have compared the performance of the three objective functions in PRIM in the context of scenario discovery on binary classified data. In the first two cases, we compared the standard objective function − which focuses on the mean only − with a more lenient objective function, which accounts for the increase in the mean offset by the loss of the number of data points inside the box. In the third case, we included a third even more lenient objective function in the comparison. Looking at the three cases for which we have compared the original objective function with the more lenient functions, we observe that the more lenient objective functions produces equal or better results. Thus, the lenient objective functions weakly, and in the presence of categorical uncertain variables, strongly, dominate the original objective function. That is, a more lenient objective function is a better default to use than the standard objective function.

We also analyzed a fourth distinct objective function that makes PRIM applicable to problems where the data is classified using a multinomial classification. Here, we compared sequential PRIM, Gini PRIM and CART. This comparison sheds some new light on an ongoing discussion on the use of CART and PRIM in the context of scenario discovery. Lempert et al. (2008) first compared the two. They concluded that neither is superior. PRIM can span too many dimensions and inadvertently combine otherwise disjoint regions in the same box. Still, PRIM allows for user interaction, which can

Table 3
Summary overview of separate PRIM analysis on each class.

	inequalities		convergence			population			behaviors		technologies		fossils		capital markets		Box composition						Gini impurity
	0	1	0	1	2	0	1	2	0	1	0	1	0	1	0	1	1	2	3	4	5	6	
Class 1		■								■							31					16	0.45
Class 2		■	■				■		■									13	1		2	16	0.58
Class 3	■						■												32			16	0.44
Class 4				■						■										21		3	0.22
Class 5		■				■				■					■						9	3	0.38
rest																	13	15	17	2	7	78	0.61

Table 4
Summary overview of PRIM with Gini objective function.

	inequalities		convergence			population			behaviors		technologies		fossils		capital markets		Box composition						Gini impurity
	0	1	0	1	2	0	1	2	0	1	0	1	0	1	0	1	1	2	3	4	5	6	
Box 1	■						■				■								22			2	0.15
Box 2							■	■					■					9		1	3	29	0.47
Box 3		■		■			■			■							24					7	0.35
Box 4				■	■		■										3	6				22	0.45
Box 5		■				■							■		■			2			1	17	0.27
Box 6	■			■		■			■										10			8	0.49
Box 7							■						■				1				9	14	0.52
rest																	16	11	8	22	6	33	0.78

help in addressing these downsides. CART can produce too many boxes for an analyst to interpret and can produce inappropriate asymmetric boxes. Hadka et al. (2015) rightly point out, however, that CART minimizes type 1 error (i.e. the number of false positives). Until now, CART had the additional advantage that it could be used for multinomial classified data directly (Gerst et al., 2013). As evidenced by the results presented here, it is possible to extend PRIM so that it can be used for multinomial classified data directly. As also seen in the comparison here, the other arguments in favor and against both CART and PRIM apply also in this case.

The results of Gini PRIM are inconclusive. Based on the case analyzed here, and the ongoing comparisons between PRIM and CART for scenario discovery, there is currently no clear superior approach. Further work is needed in this area. In particular, attention is needed for the use of PRIM for scenario discovery when it is difficult to find a single box that adequately explains the set of cases of interest. The approach suggested by Guivarch et al. (under review) is one interesting direction that could be combined with the Gini objective function presented here. There is also a clear need for applying sequential PRIM, CART, and Gini PRIM to more cases in order to assess whether indeed case specific concerns are the most relevant, or whether some more specific guidance on algorithm selection is possible.

In this paper, we have looked at modifications to the objective function used in PRIM. We have kept the lenient hill climbing optimization approach as used by 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. In the

Table 5
Summary overview of CART analysis.

	inequalities		convergence			population			behaviors		technologies		fossils		capital markets		Box composition						Gini impurity
	0	1	0	1	2	0	1	2	0	1	0	1	0	1	0	1	1	2	3	4	5	6	
Box 1	■					■			■									5	1	2		16	0.50
Box 2		■							■		■							9			5	10	0.64
Box 3	■						■				■								22			2	0.15
Box 4	■						■					■							10			14	0.49
Box 5			■	■			■				■								7			9	0.49
Box 6			■	■			■					■							4		1	11	0.46
Box 7					■		■														7	9	0.49
Box 8	■			■						■										21		3	0.22
Box 9		■		■						■							7	3				14	0.56
Box 10	■			■				■		■								1				14	0.12
Box 11		■		■				■									10					6	0.47
Box 12			■					■			■						9					6	0.48
Box 13												■					15					1	0.12
Box 14				■			■			■	■							5			6	5	0.66
Box 15				■				■	■			■					3	1				12	0.40

context of scenario discovery, PRIM is used to identify boxes that have a good coverage and density, and are relatively easy to interpret as measured by the number of dimensions used for describing the box (Lempert et al., 2008). PRIM is adapted to this purpose, but not designed for this multi-objective optimization problem. An interesting avenue for future research is to approach the problem of scenario discovery directly as a multi-objective optimization problem. That is, one could perform scenario discovery by optimizing jointly the coverage, density, and number of uncertain factors used in describing the box limits. These box limits would be the decision variables used in the optimization. This idea can be implemented relatively straightforwardly using either simulated annealing or a genetic algorithm.

6. Conclusions

Scenario discovery is a novel quantitative approach for characterizing and communicating uncertainties. Scenario discovery starts with the design and execution of series of computational experiments. These experiments cover the space of uncertain factors. Next, the results from the computational experiments are analyzed using a rule induction algorithm in order to find the combination of uncertain factors that explains the subset of cases that is of interest. PRIM is the most frequently employed algorithm for scenario discovery; however, its performance in the presence of heterogeneous uncertain factors was not yet analyzed. Moreover, PRIM could be used only for binary classifications. In this paper we

have addressed both problems.

To handle heterogeneously typed uncertain factors, we used data type specific modifications to the lenient hill climbing optimization procedure used by PRIM. More importantly, we analyzed alternative objective functions that are more suitable in the context of heterogeneously typed uncertain factors. These alternative objective functions account for both the gain in the mean and the decrease in the number of data points in the box. We compared the performance of the modified algorithm for three cases. In each case, the alternative objective function performed equally or better than the default objective function. The results are most striking in the third case that contained continuous, integer, and categorical uncertain factors. This strongly suggests that one should use one of the more lenient, objective function instead of the default.

To address the fact that PRIM was limited to binary classifications, we presented another objective function based on Gini impurity. Gini impurity is a measure of the purity of a dataset. By minimizing the impurity, PRIM is forced to find boxes that have a high concentration of cases of one or a few of classes. In this way, PRIM can be extended beyond binary classified data. When we compared Gini PRIM with CART, the *de facto* alternative in case of multinomial classified data, we observed that the relative merits of both PRIM and CART that have been discussed before in the literature (Hadka et al., 2015; Lempert et al., 2008), also apply in this case.

Acknowledgments

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.envsoft.2015.11.020.

References

Bankes, S.C., Walker, W.E., Kwakkel, J.H., 2013. Exploratory modeling and analysis. In: Gass, S., Fu, M.C. (Eds.), Encyclopedia of Operations Research and Management Science, third ed. Springer, Berlin, Germany.

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

Bryant, B.P., 2014. Sdtoolkit: Scenario Discovery Tools to Support Robust Decision Making (v2.33-1).

Bryant, B.P., Lempert, R.J., 2010. Thinking inside the Box: a participatory computer-assisted approach to scenario discovery. Technol. Forecast. Soc. Change 77 (1), 34–49.

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. Change 215–225. http://dx.doi.org/10.1016/j.techfore.2013.07.001.

Forrester, J.W., 1961. Industrial Dynamics. MIT Press, Cambridge.

Friedman, J.H., Fisher, N.I., 1999. Bump hunting in high-dimensional data. Statistics 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. http://dx.doi.org/10.1016/j.envsoft.2012.09.001.

Guivarch, C., Rozenberg, J., Schweizer, V., 2015. The diversity of socio-economic pathways and CO$_2$ emissions scenarios: insights from the investigation of a scenarios database. submitted for publication Environ. Model. Softw.

Hadka, D., Herman, J.D., Reed, P.M., Keller, K., 2015. OpenMORDM: an open source framework for many objective robust decision making. Environ. Model. Softw. 74, 114–129.

Hamarat, C., Kwakkel, J.H., Pruyt, E., 2013. Adaptive robust design under deep uncertainty. Technol. Forecast. Soc. Change 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.

Kwakkel, J.H., Auping, W.L., Pruyt, E., 2013. Dynamic scenario discovery under deep uncertainty: the future of copper. Technol. Forecast. Soc. Change 80 (4), 789–800.

Kwakkel, J.H., Pruyt, E., 2015. Using system dynamics for grand challenges: the ESDMA approach. Syst. Res. Behav. Sci. 32 (3), 358–375. http://dx.doi.org/10.1002/sres.2225.

Lempert, R.J., Bryant, B.P., Bankes, S., 2008. Comparing Algorithms for Scenario Discovery. RAND, Santa Monica, CA, USA.

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., Groves, D.G., Popper, S., Bankes, S., 2006. A general analytic method for generating robust strategies and narrative scenarios. Manag. Sci. 52 (4), 528–541.

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. Change 122 (3), 509–522. http://dx.doi.org/10.1007/s10584-013-0904-3.

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.

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

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.

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., Sleegers, A.M., van Asselt, M.B.A., 2005. The future shocks: on discontinuity and scenario development. Technol. Forecast. Soc. Change 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, 955–979.