Environmental Modelling and Software 145 (2021) 105177

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

A novel concurrent approach for multiclass scenario discovery using Multivariate Regression Trees: Exploring spatial inequality patterns in the Vietnam Mekong Delta under uncertainty

Bramka Arga Jafino *, Jan H. Kwakkel

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

A R T I C L E I N F O

Keywords:
Model-based decision support
Fairness
Adaptation
Deep uncertainty
Equity

A B S T R A C T

To support equitable planning, model-based analyses can be used to explore inequality patterns arising from different scenarios. Scenario discovery is increasingly used to extract insights from ensembles of simulation. Here, we apply two scenario discovery approaches for unraveling inequality patterns and their drivers, with an application to spatial inequality of farms profitability in the Vietnam Mekong Delta. First, we follow an established sequential approach where we begin with clustering the inequality patterns from the simulation results and next identify model input subspaces that best explain each cluster. Second, we propose a novel concurrent approach using Multivariate Regression Trees to simultaneously classify inequality patterns and identify their corresponding input subspaces. Both approaches have comparable output space separability performance. The concurrent approach yields significantly better input space separability, but this comes at the expense of having a larger number of subspaces, requiring analysts to make extra effort to distill policy-relevant insights.

1. Introduction

Recent model-based studies for supporting climate planning have advocated for assessing distributional outcomes of alternative policies (see e.g., Gourevitch et al., 2020; Kind et al., 2017; Rao, 2013). This is because evaluating policies using aggregate metrics can be misleading, as a policy that is optimal from an aggregate point of view might actually benefit some people at the expense of the others (Hansson, 2007; Rao et al., 2017; Sayers et al., 2018). Looking only from an aggregate point of view can introduce, or even exacerbate inequalities. Furthermore, there exists uncertainty in not only the magnitude and the spatial distribution of climate change, but also in the differential exposure, vulnerability, and adaptive capacity of the people and how these factors evolve over time (Green, 2016; O'Neill et al., 2017; Thomas et al., 2019). This makes it even more crucial to assess *ex ante* the distributional consequences of adaptation and mitigation policies.

There are two types of analyses for assessing distributional outcomes. The first one is normative analysis. Here, the aim is to identify a policy that best satisfies a moral principle. For instance, in climate change mitigation, the polluters pay principle and the equal per capita entitlements are two often used imperatives for allocating mitigation responsibility (Gardiner, 2010; Okereke, 2010). In adaptation, the use of differentiated historical responsibility has been proposed for determining funding responsibility (Grasso, 2007), whereas 'putting the most vulnerable first' has been proposed for distributing benefits (Paavola and Adger, 2006). These principles can be operationalized for use in quantitative model-based studies. For example, Adler et al. (2017) operationalize the prioritarian principle (giving higher weights to outcomes experienced by worse-off people) for calculating the social cost of carbon.

The second type is explorative analysis. Rather than putting value judgements on whether the distribution of outcomes is morally acceptable, explorative analysis aims to identify groups who become better-off and worse-off because of the implementation of policies. There are various ways to define population subgroups. For example, Ciullo et al. (2020) look at the distribution of flood risk reduction benefits across people living in different locations (i.e., dike rings). By identifying potential 'winners' and 'losers', explorative analysis can help planners in anticipating unintended distributive consequences and ameliorating potential injustices, for instance by preparing compensation measures to worse-off actors.

When performing explorative analysis, the analyst faces an interpretation problem arising out of two concerns (Jafino et al., 2021b). First, identifying inequality patterns requires calculating the outcomes

* Corresponding author.
E-mail address: B.A.Jafino@tudelft.nl (B.A. Jafino).

https://doi.org/10.1016/j.envsoft.2021.105177
Accepted 17 August 2021
Available online 31 August 2021

experienced by individual actors, leading to a larger number of performance indicators. This sometimes requires a modification to the model structure (Rao et al., 2017), and how model outputs are treated (Franssen, 2005; Kasprzyk et al., 2016). Second, the fact that distributional outcomes can vary substantially under different futures necessitates the exploration of inequality pattern across a large ensemble of scenarios (Schweizer, 2018; Taconet et al., 2020). Taken together, the large ensemble of scenarios and the high dimensionality of the output space make it hard to distill policy-relevant insights about the different plausible modes of inequality patterns, and the associated policies and uncertainties under which the different modes arise.

Scenario discovery is an approach for deriving policy-relevant insights from large ensembles of simulation results (Bryant and Lempert, 2010; Groves and Lempert, 2007). Scenario discovery process begins with generating simulation results database through running the model under a large number of scenarios (Bankes, 1993; Moallemi et al., 2020), and proceeds with identifying combinations of driving forces that lead to a certain pattern of model outcomes. Scenario discovery answers the question 'under which conditions or scenarios do the model outcomes behave in a certain way?'. Scenario discovery by now is a recognized approach to deal with deep uncertainty in model-based planning for climate change and to make sense of large-scale computational experiment (see e.g., Guivarch et al., 2016; Herman et al., 2015; Knox et al., 2018; Lamontagne et al., 2018; Moallemi et al., 2017; Rozenberg et al., 2014; Weaver et al., 2013).

Traditional applications of scenario discovery include policy stress testing and vulnerability analysis (e.g., Eker and van Daalen, 2015; Halim et al., 2015; Hidayatno et al., 2020; Shortridge and Zaitchik, 2018) as part of (Many Objective) Robust Decision Making (Bartholomew and Kwakkel, 2020; Kasprzyk et al., 2013). The main objective here is identifying conditions under which a policy fails to meet its objectives. This requires users to set a threshold for classifying policy success. If the performance of the policy exceeds (or goes below, in case of a maximization problem) the threshold, the policy is deemed to fail in reaching its objectives. In this established application of scenario discovery, one applies a binary classification to the model output space (from the simulation results database) by dividing the output space into a region where the policy performance meets the minimal requirement and a region where it fails to do so. A rule induction algorithm is then applied to identify combinations of input parameters that lead to the vulnerable region in the output space.

In this study, we investigate the merits of using multiclass scenario discovery, an extension of the standard binary-class scenario discovery, for performing explorative analysis of distributional outcomes. In multiclass scenario discovery the model output space is partitioned into multiple clusters, and the input subspaces for each cluster are then identified. Multiclass scenario discovery is appropriate for explorative analysis of distributional outcomes. As there might be numerous modes of inequality in the future, we cannot simply impose a binary classification on the distributional outcomes. Distinctive inequality patterns might emerge, but due to system complexity and non-linearity, similar patterns might arise from completely distinct uncertainty and policy scenarios (Jafino et al., 2021a).

We explore two alternative approaches to multiclass scenario discovery. First, we adapt the cluster-then-identify approach as has been used in previous multiclass scenario discovery studies (Gerst et al., 2013; Rozenberg et al., 2014; Steinmann et al., 2020). In this approach, the clustering of the model output space is performed first, followed by the identification of input subspaces for each cluster separately. This can negatively affect interpretability because different clusters in the output space might be linked to overlapping subspaces of the input space. To address this, we propose and test the use of Multivariate Regression Tree (MRT) for multiclass scenario discovery. In this second approach, the output space clustering and input subspace identification are solved concurrently through the MRT algorithm. We apply both the established sequential and the novel concurrent approach for multiclass scenario discovery to an agriculture adaptation planning problem for the upper Vietnam Mekong Delta (VMD). We explore spatial inequality of district-level farms profitability resulting from different realizations of uncertainties and implementation of adaptation measures.

The rest of the paper is structured as follows. In section 2, we describe the two approaches of multiclass scenario discovery and explain further the concept of input and output space separability. In section 3, we provide the background of the case study and introduce the model that is being used. In Section 4, we present the results of the two approaches. In Section 5, we discuss the merits of each approach, i.e., their performance in terms of input and output space separability as well as the resulting scenario narratives identified by each approach. In Section 6, we summarize our main findings and insights.

2. Methods

2.1. Multiclass scenario discovery

There are a number of scenario discovery applications that extend the output space partitioning from binary classification to multiclass classification (Gerst et al., 2013; Kwakkel and Jaxa-Rozen, 2016; Rozenberg et al., 2014; Steinmann et al., 2020). A major difference between traditional scenario discovery and multiclass scenario discovery lies in the characterization of the output space. In traditional scenario discovery, the output space is partitioned into only two classes: those which are of interest and those which are not (Kwakkel et al., 2013). In contrast, in multiclass scenario discovery the output space is partitioned into more than two classes. Multiclass scenario discovery involves two tasks: the output space has to be partitioned into multiple distinct classes, and for each class input subspaces which are highly predictive for it have to be identified. The highly predictive input subspaces form the narrative behind each class in the output space.

For the first task (partitioning the output space), various approaches for specifying the classes have been used. Classification can be performed by either manually imposing a threshold on the outcome variables (e.g., Guivarch et al., 2016; Rozenberg et al., 2014), or by using a clustering algorithm to automatically identify the classes (e.g., Berntsen and Trutnevyte, 2017; Gerst et al., 2013; Moallemi et al., 2017; Steinmann et al., 2020). In the manual threshold approach, the analyst has full control over how the output space is partitioned, thus enhancing the interpretability of the resulting classes. However, the task becomes increasingly complex with increasing number of outcome variables. In contrast, clustering algorithms can handle a larger set of outcome variables but at the expense of worsening interpretability. For the second task (identifying highly predictive input subspaces), both the Patient Rule Induction Method (PRIM) (Friedman and Fisher, 1999) and Classification and Regression Tree (CART) algorithms (Breiman et al., 1984) have been widely used. For multiclass scenario discovery, PRIM is iteratively and independently applied to each cluster of the output space (see e.g. Rozenberg et al., 2014). In contrast, CART can identify highly predictive input subspaces for multiple clusters of the output space simultaneously, by predicting the membership of each scenario in one of the identified clusters (see e.g. Gerst et al., 2013).

The partitioning of the output space and the identification of highly predictive input subspaces are traditionally performed sequentially. In this study, we propose the use of Multivariate Regression Tree (MRT) for multiclass scenario discovery to concurrently perform these two tasks. MRT is an extension of CART where multiple dependent variables are being used to characterize the impurity of a decision node (De'ath, 2002). MRT has previously been used for model-based analysis, such as for unraveling tradeoffs and synergies between management objectives (Ndong et al., 2020; Smith et al., 2019). For multiclass scenario discovery, the input parameters of the simulation model become the independent variables of the MRT, while the outcome variables of interest become the dependent variables. The leaves resulting from the regression tree then act as the clusters of the output space. The variables being

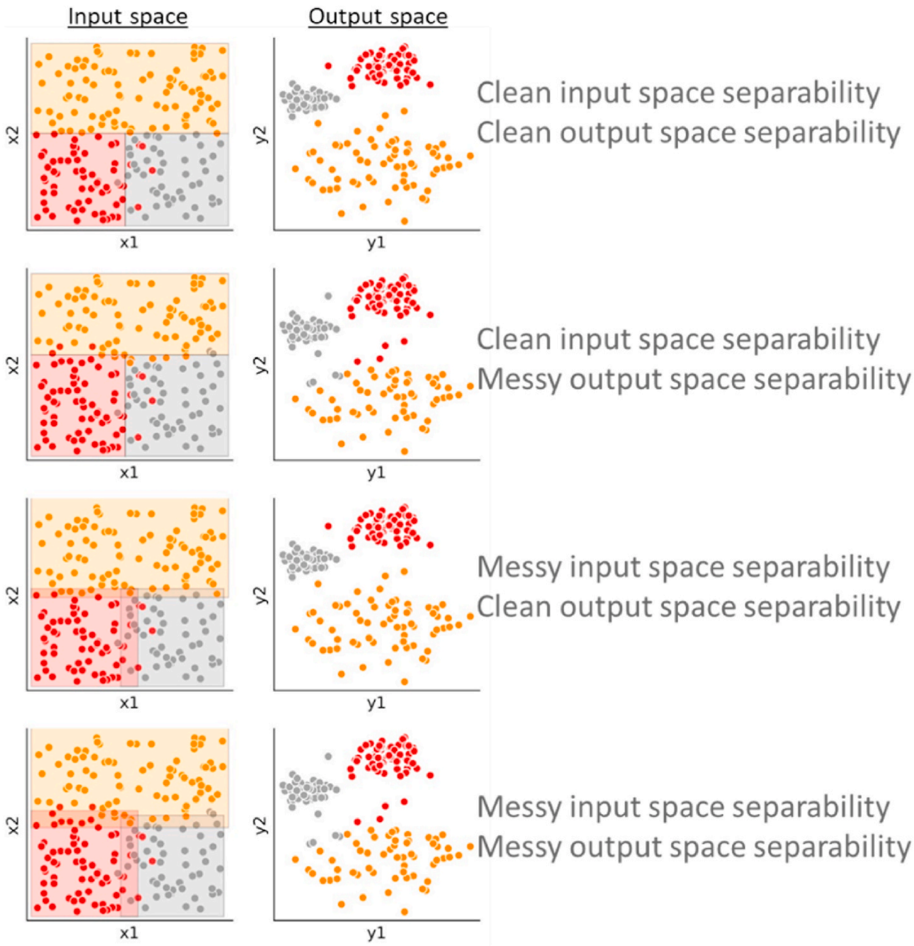

Fig. 1. Illustration of input and output space separability. The color of the data points is based on the identified clusters in the output space. The shaded regions are the identified input subspaces from the rule induction algorithm. The output space separation in the second and fourth row misclassifies some of the orange cluster into the grey and red clusters. The input space separation in the third and fourth row is not clean as there is too much overlap between the induced subspaces.

used in each decision node and their corresponding splitting values form the narrative behind each cluster of output space.

Scenario discovery enables the extraction of policy-relevant insights (e.g., exploring plausible modes of inequality patterns) from large-scale computational experiments by making the large ensemble of simulation results interpretable. The interpretability of multiclass scenario discovery can be evaluated using three criteria. The first criterion is output space separability, which is similar to the objective of clustering algorithms (Hastie et al., 2009; Jain, 2010). After clustering the output space, members within the same cluster should have similar outcome characteristics (e.g., spatial inequality patterns), while members from different clusters should be dissimilar. The second criterion is input space separability (Steinmann et al., 2020), which focuses on the rule induction part of scenario discovery. Each class of outcome should originate from distinct and non-overlapping subspaces in the input space. As illustrated in Fig. 1, scenario discovery results are ideal if the identified input and output subspaces are completely separable, i.e., if each cluster in the output space is distinctive from the other clusters and is driven by distinctive subspaces in the input space. The third criterion is the resulting number of scenario narratives. Having a larger number of clusters generally leads to better output space separability (Hastie et al., 2009), but it comes at the expense of having more complicated narratives to be communicated to decision makers.

2.2. *Sequential approach: cluster-then-identify*

2.2.1. *Clustering phase*

The clustering phase aims to find distinctive patterns of outcomes within the simulation results. Clustering performance is evaluated by the explained variance:

$$EV_K = 1 - \frac{\sum_{k=1}^{K} SSE_k}{SSE_{all}}$$

where EV_K is the explained variance of the algorithm with K clusters, SSE_k is the sum of squared error of members in cluster k, and SSE_{all} is the sum of squared error of the entire dataset. Explained variance generally increases with the number of clusters. The more clusters are used, the smaller the differences between members within each cluster will be. We use the elbow method to select the optimal number of clusters (Ketchen and Shook, 1996). Here, we calculate the difference of the explained variance:

$$\Delta EV_K = EV_K - EV_{K-1}$$

We can then set a threshold T and determine the number of clusters where an additional cluster would yield $\Delta EV_K < T$ as an optimal number of clusters for the particular algorithm.

We consider five clustering algorithms that are commonly used in model-based analysis (Bandaru et al., 2017; Bárcena et al., 2015; Moallemi et al., 2018; Rohmer et al., 2018; Szekely and Rizzo, 2005): k-means clustering, k-medoids clustering, Gaussian mixture model,

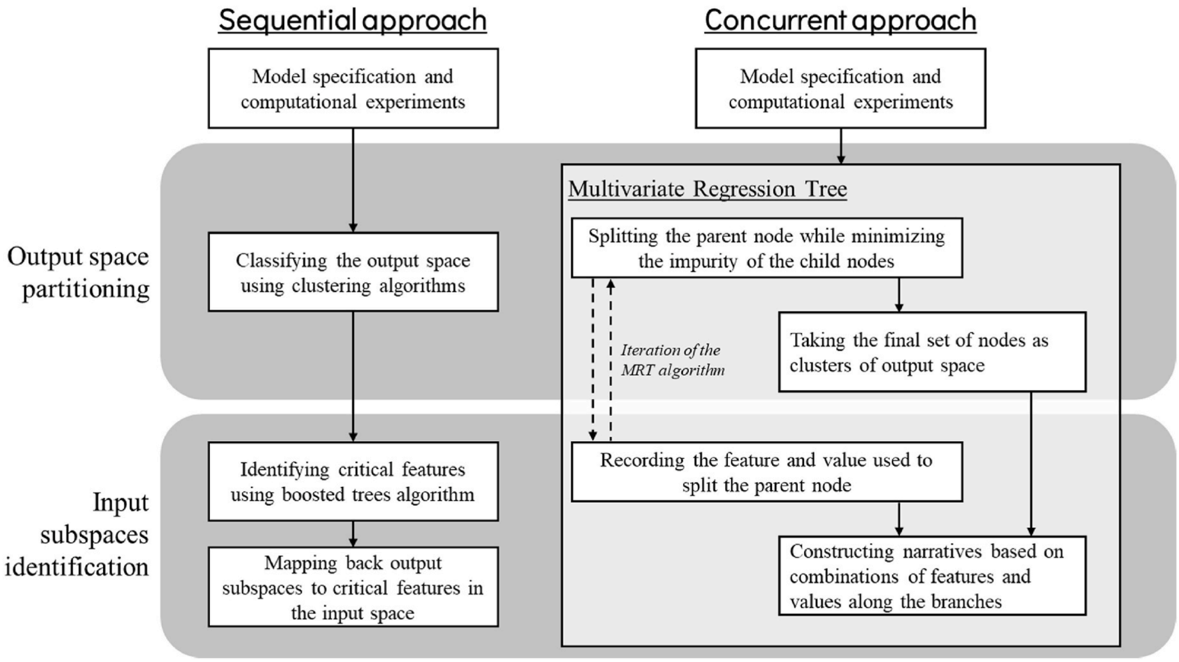

Fig. 2. Schematic comparison of the sequential and the concurrent approach.

agglomerative clustering with complete linkage, and agglomerative clustering with average linkage. The combination of clustering algorithm and corresponding optimal number of clusters that yields the highest explained variance is selected for further analysis.

2.2.2. Input subspace identification phase

We adopt the boosted trees algorithm to induce subspaces conditional on each class of the output space (Trindade et al., 2019). Boosted trees build upon CART by generating an ensemble of classification trees, where each tree tries to minimize the impurity in the dataset by iteratively splitting the dataset into leaves (De'ath, 2007; Hastie et al., 2009; Schapire and Freund, 2012). A leaf is impure if it contains mixes of data points from different classes, or, in our case, simulation results from different clusters. We use the Gini impurity criterion:

$$I_G(m) = \sum_{k=1}^{K} p_{mk}(1 - p_{mk})$$

where $I_G(m)$ is the Gini impurity of leaf or node m, K is the total number of classes of the output space, and p_{mk} is the proportion of scenarios with class k in node m. In each iteration, a classification tree looks for all possible splits across the input features and selects the one that yields the highest reduction in impurity. Boosted trees employ an ensemble of weak classification trees through multiple boosting iterations. In each boosting iteration, the algorithm readjusts the weights of misclassified data that are to be inputted to the weak classifier in the successive iterations (Freund and Schapire, 1997; Hastie et al., 2009). Users control the algorithm by setting the maximum number of boosting iteration and limiting the complexity of individual trees (Pedregosa et al., 2011; Zhu et al., 2009).

The setup of boosted trees allows for calculating the relative importance of each input feature. In each splitting iteration, a classification tree uses one input feature to separate a parent node into two child nodes. The importance of an input feature can be estimated as a function of how often a given feature is selected as the splitting variable and how much impurity reduction is achieved. Specifically, the importance is measured by the normalized percentage of total impurity loss across all trees due to splits using the input feature. Finally, for scenario discovery, the most influential input features are mapped back to the identified clusters of the output space – a technique often coined factor mapping (Trindade et al., 2019). The factor maps can be used to visually construct rules or scenario narratives (i.e., combinations of input parameters) for each cluster of output space.

2.3. Concurrent approach: Multivariate Regression Trees

MRTs are an extension of univariate regression tree where multiple response variables are being used simultaneously to find candidate splits in each decision node (De'ath, 2002). In each iteration, MRT looks for the best split in the input features that leads to the largest reduction of impurity in the child nodes. For regression problems, the impurity of a node in terms of a single response variable is calculated as the summed Euclidean distance between each data point to the mean of the response variable. Accordingly, in MRT, the total impurity of a node (also termed the error of the node for regression trees) is calculated as the summation of the impurity of each response variable:

$$E_m = \sum_{n=1}^{N_m} \sum_{j=1}^{J} \left(y_{ij} - \overline{y}_{j(m)}\right)^2$$

where E_m is the error or impurity of node m, N_m is the total number of data points in node m, J is the total number of response variables, y_{ij} is the value of response variable j from data point i, and $\overline{y}_{j(m)}$ is the mean of response variable j across data points in node m. The algorithm looks for the optimal split in the input space that yields the lowest sum of errors from the two child nodes.

In our application, the leaves from the tree will directly turn into the clusters of inequality patterns. This is because the splitting criterion in MRT is intended to minimize the similarity of outcome variables between the child nodes while maximizing the similarity within the child nodes. To maintain interpretability, it is important to balance the size of the tree with the purity of the tree. The size of the tree (the tree 'depth') in an MRT is externally determined by the user by specifying a stopping criterion, such as the maximum number of leaves, or the minimum impurity of the leaves (Breiman et al., 1984; De'ath, 2002; Pedregosa et al., 2011). We use a 10-fold cross validation technique to decide the appropriate depth of the tree (Larsen and Speckman, 2004). In each fold, the algorithm is trained on 90% of the data and the accuracy of the

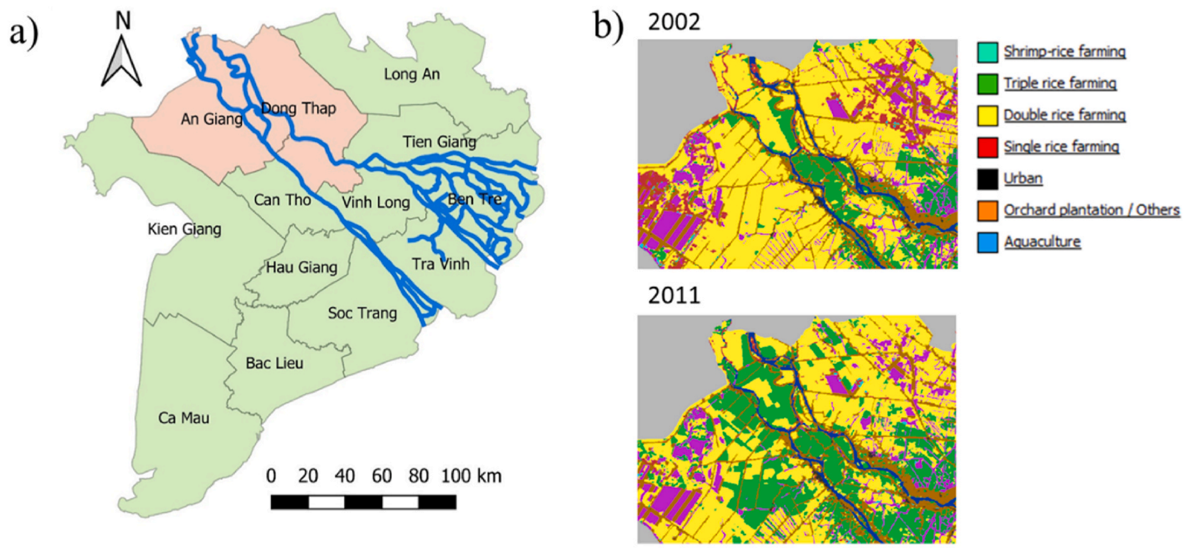

Fig. 3. a) Provinces in the VMD, with the case study area highlighted in red and the Mekong river branches represented by blue lines; b) Land-use changes in An Giang and Dong Thap between 2002 and 2011 (GAEN-View, 2013; Sakamoto et al., 2009). The expansion of triple-rice farming indicates the construction of high dikes, as triple-rice farming systems are only possible in high dikes enclosed areas.

resulting tree is tested on the rest 10% of the data. The accuracy is indicated by the coefficient of determination score:

$$R^2 = 1 - \frac{SSE_{res}}{SSE_{all}}$$

where SSE_{res} is the sum of squared error between the predicted values and the actual values of the response variables, while SSE_{all} is the sum of squared error between the actual values and the mean values of the response variables in the entire dataset. The accuracy of an MRT will increase with the depth of the tree. Hence, we also calculate the changes in accuracy and attempt to balance this with the complexity of the tree (Ndong et al., 2020; Smith et al., 2019). The selected tree depth is the one that has changes in accuracy smaller than a specified threshold T.

The resulting decision tree can be analyzed and visually inspected starting from either the leaves or the root (Smith et al., 2019). In leaves-first analysis, users begin with looking for the leaf that contains certain patterns of interest. The analysis then goes up the decision tree to understand conditions (i.e., combinations of input parameters and their values) that lead to the leaf of interest. In root-first analysis, users start from the very first decision node at the top of the tree, and go down the tree to explore a specific scenario. Leaves-first analysis is a bottom-up approach to reading a decision tree while root-first analysis is a top-down approach. Note that leaves-first and root-first analyses are concerned with how we read the MRT results. Hence, the choice between these two does not alter the results of the algorithm itself.

Fig. 2 summarizes how the two main steps in multiclass scenario discovery (i.e., output space partitioning and input subspaces identification) are carried out in the sequential and the concurrent approach. Through iteratively minimizing the impurity of the child nodes, the MRT partitions the output space to find distinctive patterns of outcomes. At the same time, the input features used to split each parent node as well as the splitting value of these features are used to construct narratives behind each final child nodes of the tree.

3. Case study

3.1. Adaptation planning in the upper Vietnam Mekong Delta (VMD)

The VMD, located in the southern part of the country, is one of the largest deltas in the world. The delta supplies 55% of total rice production and 85% of total rice export of Vietnam (GSO, 2019; Toan, 2014). The upstream part of the delta – including An Giang and Dong Thap provinces (see Fig. 3a) – is subject to annual monsoon flooding which could be worsened by climate change (Hoang et al., 2019; Triet et al., 2020). Flood risks are further exacerbated by land subsidence of which 7–17 mm/year has been attributed to agricultural activities (Minderhoud et al., 2018). Sediment starvation puts another pressure on the delta. Further development of hydropower dams in Cambodia, which is located upstream of the VMD, reduces sediment concentration in the river, which has been one of the main sources of free nutrients for farmers in the VMD (Lauri et al., 2012; Manh et al., 2015).

The agricultural sector in the VMD has experienced several transitions in the past decades. The construction of water resources infrastructure allowed farmers to harvest twice a year (double-rice cropping): the winter-spring crop between December and March and the summer-autumn crop between April and July (Ngan et al., 2018; Son et al., 2013). Dikes of around 2 m high were initially constructed, but they do not protect the paddy fields against flooding during the annual peak discharges in the monsoon season. To facilitate further intensification of the agriculture sector, the government has been constructing high dikes of 4.5 m since the early 2000s, protecting the fields against monsoon flooding and thus enabling farmers to have a third cropping season (triple-rice cropping, see Fig. 3b). Recently, it was found that the high dikes expansion policy has unintended consequences for environmental sustainability (Garschagen et al., 2012; Tran et al., 2018) and for inequality between richer and poorer farmers (Chapman et al., 2016).

In this study, we evaluate the spatial inequality of farm profitability. Specifically, we look at how different spatial inequality patterns at a district level emerge from different combinations of anthropogenic pressure, climatic change, and implementation of alternative adaptation policies. This allows us to provide spatially explicit policy advice and administrative area-based recommendations for local decision makers. Our study complements previous inequality studies in the region that focus on the distributional outcomes from a household point of view (i.e., comparing poor and rich farmers at an individual household level) (Chapman and Darby, 2016; Chapman et al., 2016).

3.2. Integrated assessment metamodel

We used a spatially explicit integrated assessment metamodel to simulate the profitability of the farmers in An Giang and Dong Thap provinces, combining previously established complex models (Jafino

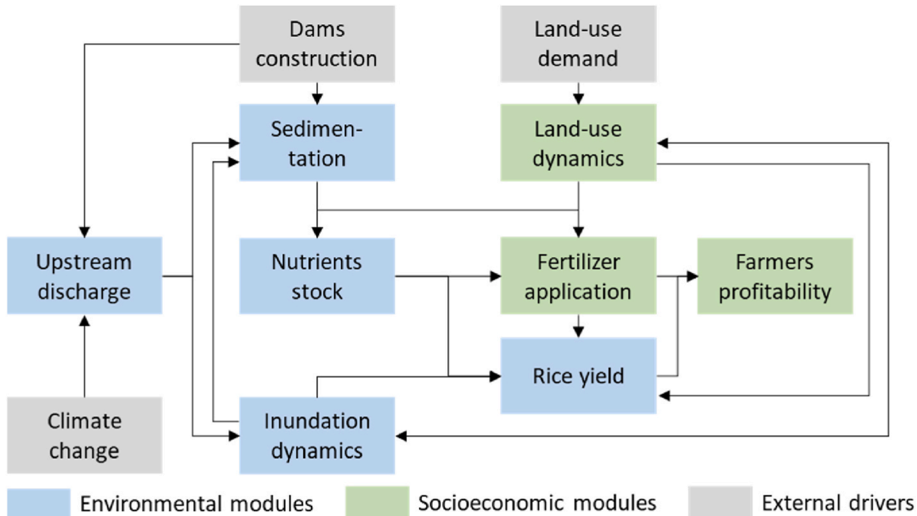

Fig. 4. Conceptual diagram of the spatially-explicit integrated assessment metamodel, taken from Jafino et al. (2021a).

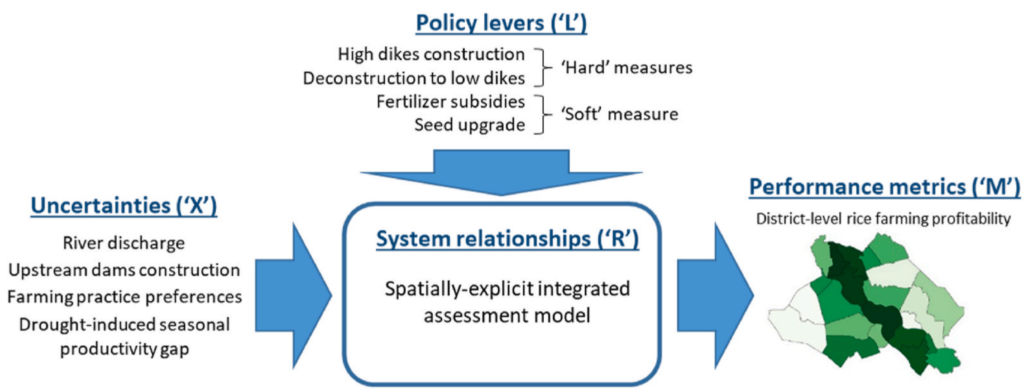

Fig. 5. XLRM overview of the case study.

et al., 2021a). Fig. 4 shows the general conceptualization of the model. In short, the model operates with a spatial resolution of 200m where each cell is represented by a particular land-use function (e.g., single rice, double-rice, triple-rice, orchard plantation, or aquaculture). Profitability is then calculated for each cell. based on income from selling rice and cost of purchasing fertilizer. We assume that nutrients are the limiting factors of rice yield, which is the case in most Southeast Asian countries (Sattari et al., 2014; Witt et al., 1999).

Inundation plays two opposing roles. On the one hand, unintentional inundation, for instance due to extremely high discharge in monsoon seasons, reduces the total annual rice yield. On the other hand, inundation supplies free nutrients through floodplain sedimentation. Flood risks are reduced in areas with higher dikes and are increased by land subsidence, which in turn is dependent on the land-use dynamics. Finally, the land-use module simulates farmers' behavior of changing cropping practices especially between double-rice cropping, triple-rice cropping, orchard plantation, and aquaculture. Future distributional outcomes are evaluated at a district level. Hence, the cell-level profitability is aggregated for each of the 23 districts in An Giang and Dong Thap provinces. The model is run with an annual time step from 2012 to 2050. The detailed model description, validation, and fit for purpose assessment are described in Jafino et al. (2021a).

3.2.1. Adaptation measures

We tested both 'hard' infrastructural and 'soft' non-infrastructural policies that affect the different modules within the model. The infrastructural policies are related to dike (de)construction. These policies are drawn from the recent flood control debates in the region: either further expansion of high dikes or deconstructing all established high dikes into low dikes (Käkönen, 2008; Tran et al., 2018; Triet et al., 2020). These policies are applied in An Giang and Dong Thap independently, resulting in a total of four alternative policies. The first soft policy is a seeds upgrade policy. We assume that by using a better seed variety the crops become more resilient to floods. We model this by reducing the steepness of the stage-damage curve (Dutta et al., 2003; Triet et al., 2018), so that the same level of inundation results in a lower yield reduction. The second policy is fertilizer subsidies where 50 kg of free fertilizer are distributed to farmers in each cropping season. Free fertilizer is given to farmers located far from the river, as they get a significantly lower nutrients concentration from floodplain sedimentation (Manh et al., 2013, 2014).

3.2.2. Uncertainties

There are five key uncertain factors affecting the productivity of the agricultural sector in the upper VMD. The first uncertain factor is future annual peak discharge that affects flood risk. We use synthetic future hydrographs of the Mekong River, generated by a global hydrological model driven by climatic data from two scenarios (RCP 4.5 and 8.5) (Sutanudjaja et al., 2018; Winsemius et al., 2013). The second uncertain factor is the hydropower dam development upstream in Cambodia. This factor affects the annual peak discharge and reduces total sediment supply to the VMD as the dams trap the sediment upstream. We use five dam development scenarios as worked out by Lauri et al. (2012) and Manh et al. (2015).

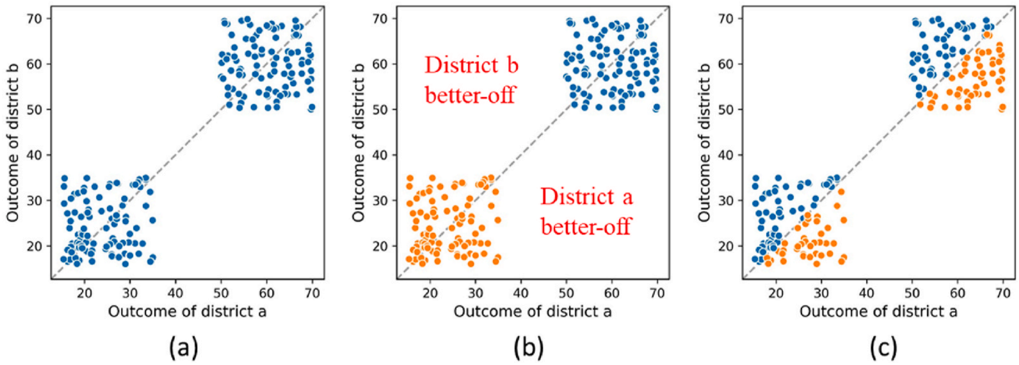

Fig. 6. An illustration of the implication of normalizing across scenarios and within scenarios. If district a and district b have the same outcomes, the data point will fall exactly on the diagonal dashed line. (a) Outcomes for the two districts across 500 scenarios. (b) Clustering results when the outcomes are normalized across scenarios. Here we see that in each cluster we have scenarios where both district b and district a are better-off. (c) Clustering results when the outcomes are normalized within each individual scenario. Here the resulting clusters have distinctive inequality pattern (district a is better-off in the orange cluster, while district b is better-off in the blue cluster).

The next two factors are the productivity gap among the three seasons. The winter-spring season that starts in December, just after the wet monsoon season, is the most productive season. The summer-autumn season and the autumn-winter season are less productive due to the limited water content in the soil in the former and the high degree of precipitation in the latter. In 2002, the summer-autumn season and the autumn-winter season in Dong Thap produced 38% and 50% fewer yield per hectare, respectively. In 2016, the productivity gap has been reduced to only 26% and 35% for the summer-autumn and autumn-winter season, respectively. In this study, we consider a wide range of plausible future productivity gap between 15 and 45%.

The last uncertain factor is the society's preference toward the different rice cropping system and the spatial plan for the region. This factor affects future land-use demand, which in turn is spatially allocated by the land-use change module. We consider four scenarios based on the competing narratives of agriculture intensification in the VMD as well as based on the Mekong Delta Plan (Mekong Delta Plan Consortium, 2013; Tran et al., 2018; Triet et al., 2018): continuing intensification (higher triple-rice cropping demand and lower double-rice cropping demand), reverting to double rice (the opposite of the first scenario), rising non-rice preferences (higher demand for alternative livelihoods such as orchard plantation, aquaculture, and shrimp-rice farming), and increasing urbanization (higher demand for residential area).

3.2.3. Experimental setup

The setup of the case study is summarized using the XLRM framework (Lempert et al., 2003) in Fig. 5. To allow for an exhaustive exploration of plausible combinations of uncertainties and policies, we apply full factorial sampling to input factors that are categorical and ordinal, i.e., we sample all possible combinations of categorical and ordinal input factors. These factors include the six policy variables and some of the uncertain variables (i.e., river discharge and farming practice preference). We combine the full factorial sample with a Latin Hypercube Sampling of the productivity gap uncertainties, as the values for these uncertainties take a continuous range. This experimental setup results in a total of 43200 computational experiments. The exploratory modelling workbench (Kwakkel, 2017) is used to perform these experiments.

3.3. Post-processing of simulation results

The clustering phase in the sequential approach and the calculation of error in the concurrent approach require the computation of 'distance' between the outcomes of each scenario in the simulation results database. To avoid having one outcome variable dictating the distance calculation, the values of each outcome variable are usually normalized to 0–1 across the scenarios (e.g., Giudici et al., 2020; Smith et al., 2019). Normalization of each outcome variable across the entire scenarios when doing explorative analysis of distributional outcomes is problematic. The outcome variables are the outcomes for each district. By doing a normalization we lose sights of the relative performance of each district compared to all other districts within each scenario (see Fig. 6a and b). Hence, we calculate instead the 'relative profitability' of each district, i.e., the 0–1 normalization is applied between the performance of each district within each scenario, instead of across scenarios (see Fig. 6c). In this way we maintain the information regarding the relative 'winners' and 'losers' in each scenario. As a result, the clustering algorithm is forced to look for distinctive inequality patterns.

4. Results

4.1. Sequential approach

The first step in the sequential approach is clustering the output space into a number of representative inequality patterns. We test five alternative clustering algorithms while varying the number of clusters (see Appendix A for details). We find that the k-Means algorithm with seven clusters yields the most satisfactory performance which balances the explained variance and the number of final clusters. The remainder of the sequential approach is thus based on the results from this clustering setup.

Fig. 7a shows the seven representative inequality patterns from each cluster of the output space. The representative scenario is taken from the medoid of the corresponding cluster, that is, the scenario which outcomes have the smallest Euclidean distance to all other scenarios in the cluster. At a glance, we can see that cluster 2, 3, 6 and 7 have similar inequality patterns where three districts located around the mid-northeastern part of the region have a higher relative profitability of higher than 0.7. The patterns are different once we inspect them in more detail. For example, in cluster 3, the district located in the top north-western part of the region is not relatively better-off. In cluster 6, this district is significantly better-off compared to the others (relative profitability = 1).

Next, we use the boosted trees algorithm to first identify the most critical input features that best explain the seven clusters of inequality patterns. Fig. 7b shows the results of the input feature scoring. The most important input feature is the degree of upstream dam development, followed by three dikes construction policies: expansion of high dikes in An Giang, in Dong Thap, and reverting back to low dikes in An Giang. The other input features have substantially lower importance scores.

We use the four most important input features to map back the input space to the seven clusters of output space. The importance scores of these four features add up to 0.705, implying that these features contribute to 70.5% of the total impurity reduction in the entire ensemble of trees. Fig. 7c shows the factor map for each cluster, where the cluster numbering corresponds to the seven inequality patterns in

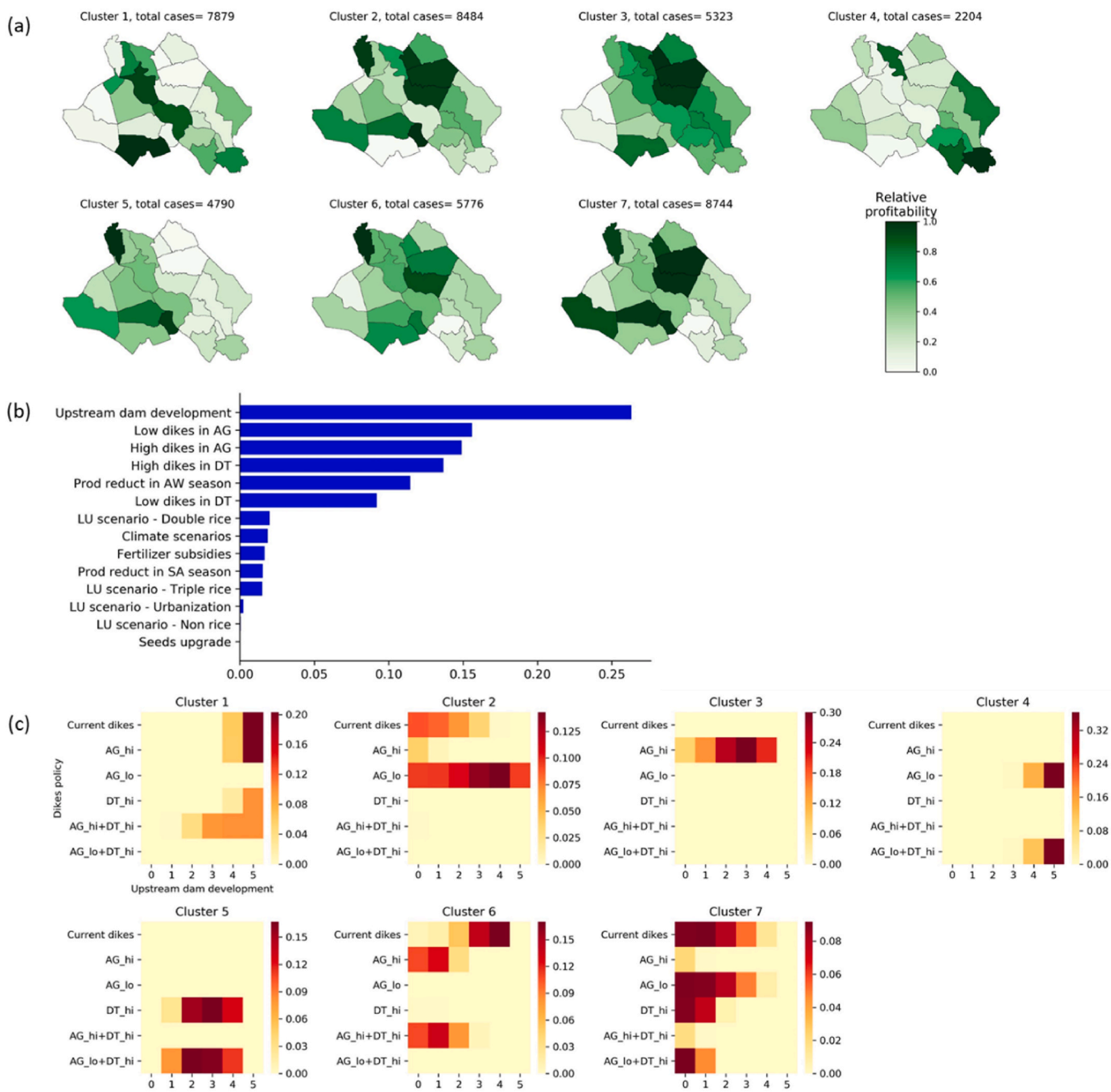

Fig. 7. Results from the sequential approach: (a) representative inequality pattern from each cluster of output space and the number of scenarios in each cluster; (b) relative importance of the input parameters as identified by the boosted trees algorithm; (c) Factor maps of the identified clusters of output space. AG_hi and DT_hi refer to further construction of high dikes in An Giang and Dong Thap, respectively. AG_lo refers to deconstruction of high dikes into low dikes in An Giang. The colorbar in each sub-figure refers to the fraction of total scenarios within that particular cluster.

Fig. 7a. Since three of the four most important features are related to dike construction policies, we combine them into a single axis (i.e., the vertical axis on Fig. 7c). The numbers underlying the heatmap correspond to the fraction of scenarios in that particular cluster. For example, 20% of the 7879 scenarios in cluster 1 are scenarios with high upstream dam development while maintaining the current dikes configuration in the VMD. Another 20% of the scenarios have a combination of high upstream dam development and low dikes policy in An Giang.

Fig. 7c shows that cluster 1, which has one of the more distinctive inequality patterns (see Fig. 7a), is primarily induced through a combination of high sediment trapping due to upstream dam development, and, either expansion of high dikes in An Giang, or the preservation of current dikes. Inequality pattern as exemplified by cluster 4 is caused by the transformation of high dikes back into low dikes in An Giang together with a high degree of upstream dam development. Cluster 2 and 7, which have similar inequality patterns, emerge if upstream dam development is relatively low and either the low dikes policy in An Giang is enacted or the current dikes system is maintained. Further construction of high dikes in An Giang in combination with relatively low upstream dam development would lead to inequality patterns as depicted either in cluster 3 or 6.

What do these results imply for adaptation planning in the VMD? The most important insight is that the interaction between what the VMD government does (in terms of dikes (de)construction) and what the

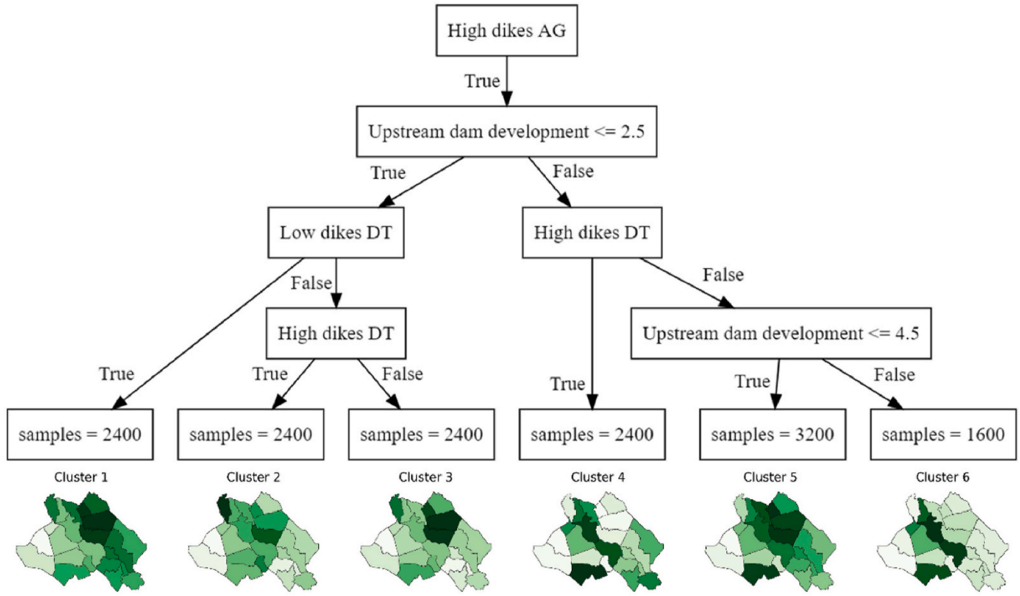

Fig. 8. Left branch of the multivariate regression tree and the corresponding representative inequality pattern.

Cambodian government does (in terms of hydropower dams development) has non-linear effects on the emerging spatial distribution of farms profitability. For instance, a relatively small degree of upstream dam development would make provinces that expand their high dikes worse-off. This follows from comparing cluster 3 and 6 with cluster 5. In cluster 3 and 6, the high dikes policy in An Giang is enacted and this makes districts within An Giang relatively worse-off. In cluster 5, the high dikes policy in Dong Thap is also enacted and this leads to districts in Dong Thap becoming worse-off. This stresses the importance of having transboundary basin management in order to ensure equitable future for the VMD farmers.

4.2. Concurrent approach

The first step in the concurrent approach is growing the regression tree and selecting an appropriate tree size. We iteratively grow the tree from three to 40 leaves and observe the evolution of the cross-validation scores (see Appendix B for details). We find that the tree with 18 leaves yields the most satisfactory cross validation score and proceed with this tree size in the remainder of the concurrent approach. For visualization purpose, we separate the entire regression tree into two figures: Fig. 8 and Fig. 9 together make up the entirety of the regression tree.

The first splitting variable identified by the MRT is the expansion of high dikes in An Giang. Fig. 8 shows the left branch of the tree (high dikes in An Giang is expanded) while Fig. 9 shows the right branch of the tree (high dikes in An Giang is not expanded). The number of scenarios and the representative inequality pattern from all scenarios in each leaf are provided at the bottom of the figures. Similar to the sequential approach before, the medoid scenario in each leaf is assigned to be its representative inequality pattern. It is important to restate here that in each scenario the profitability of the districts is normalized between 0 and 1 where darker green color means higher relative profitability. Here, we illustrate how we can use either leaves-first or root-first analysis to interpret the results of the MRT. We will use root-first analysis to analyze the left branch of the tree and leaves-first analysis for the right branch of the tree.

The left branch of the tree as shown in Fig. 8 contains scenarios where the high dikes policy in An Giang is implemented. For illustration, we approach this side of the tree using root-first analysis. The subsequent decision node here is the degree of upstream dam development with a cutoff point of 2.5 (we have six levels of upstream dam development, with 0–2 being no to medium degree of upstream development and 3–5 being higher degrees of development). If upstream dam development is relatively small, the next deciding factors are the dikes policy in Dong Thap. Low dikes policy leads to inequality pattern in cluster 1, high dikes policy leads to cluster 2, while maintaining the current dikes distribution in Dong Thap leads to cluster 3.

It is interesting to compare cluster 2 and cluster 4, as, from the root-first perspective the only difference is the degree of upstream dam development. If high dikes are expanded in both provinces and many upstream dams are eventually built, districts alongside the river will become substantially better-off (cluster 4). However, a smaller degree of upstream dam construction will lead to a less striking difference in relative profitability (cluster 2). Cluster 6, although having a different narrative, has a similar inequality pattern as cluster 4. Even without expansion of high dikes in Dong Thap, a very large degree of upstream dam development in combination with high dikes policy in An Giang still make districts alongside the river better-off.

The right branch of the tree contains the remaining 12 leaves (Fig. 9). We approach the interpretation of this branch by following leaves-first analysis. We focus on three visually distinct inequality patterns. The first pattern is typified by the higher relative profitability of districts alongside the river, as observed in cluster 15 and 18. Both clusters actually have a similar narrative where no dikes policy is taken, and upstream dam development is relatively large. The second distinct pattern is exemplified by cluster 8 and 9. In both clusters, districts located to the north of the river have smaller relative profitability. Both clusters have a similar narrative of medium degree of upstream dam development, and with high dikes being expanded in Dong Thap.

The third distinct pattern is observed in cluster 10–14 and cluster 16. The main pattern here is that there are three districts located to the north of the river, three districts located to the south of the river, and one district on the northwest corner of the region who are better-off. This pattern can emerge from multiple future conditions. For example, a condition for cluster 10–14 to materialize is no extremely high upstream dam development. However, Cluster 16 shows that even if all planned upstream dams are built, a similar inequality pattern could emerge if all dikes in both An Giang and Dong Thap are reverted back into low dikes.

What can the VMD government learn from the concurrent approach? Through combining both root-first and leaves-first analyses, we can clearly see that similar narratives could lead to distinctive patterns of spatial distribution. At the same time, similar distribution patterns could

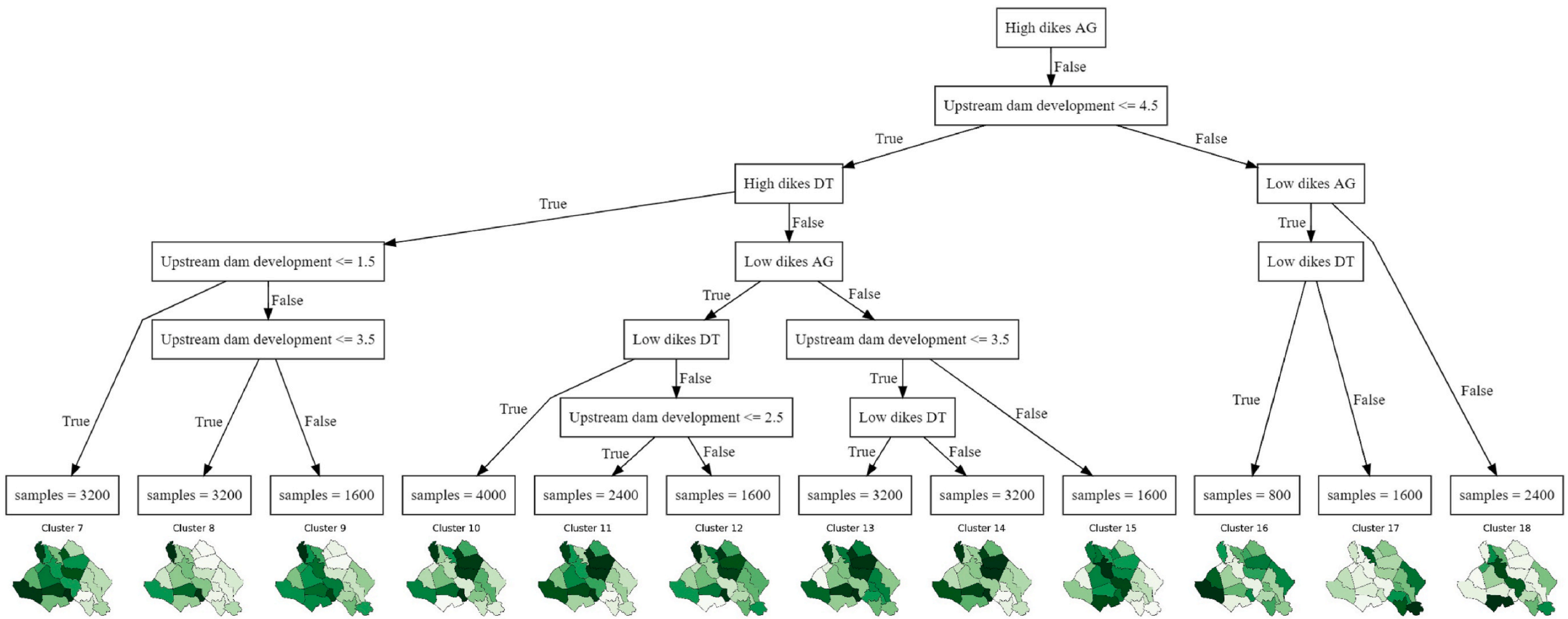

Fig. 9. Right branch of the multivariate regression tree and the corresponding representative inequality pattern.

Table 1
Comparison of output space separability.

	Within-cluster average distance	Between-cluster average distance
Sequential approach	2.997	6.329
Concurrent approach	2.628	6.007

emerge from distinctive narratives. The decision tree can easily help the government in understanding plausible inequality patterns and pathways that lead to those patterns, and thus preparing additional measures to compensate worse-off districts.

5. Discussion

5.1. Input space and output space separability

The induced input subspaces though scenario discovery have a perfect separability if each subspace is mutually exclusive with the others. While traditionally this has been quantified through the density and coverage indicators (Bryant and Lempert, 2010), the use of these indicators is not applicable for sequential multiclass scenario discovery as presented here. This is because in the sequential approach we do not set a strict boundary on the identified subspaces. However, from a visual inspection, we can see that some of the identified subspaces are overlapping with each other (see Fig. 7c). For example, the combination of current dikes and low to medium upstream dam development exists in the identified subspaces of cluster 2 and 7. In contrast, the concurrent approach produces completely separable input subspaces, as each end leaf has unique scenario narratives (i.e., combination of the scenario variables). Therefore, the concurrent approach leads to better input space separability compared to the sequential approach.

To quantify output space separability, we calculate the average Euclidean distance between the relative profitability of the 23 districts in all scenarios within each cluster (within-class dissimilarity) as well as between scenarios from different clusters (between-class dissimilarity). A better separability of output space thus entails low within-cluster distance and high between-cluster distance. From Table 1, we can see that neither approach is superior to the other. The concurrent approach has better within-cluster average distance compared to the sequential approach. This can be explained by the more granular separation of the output space, so that each cluster consists of more similar simulation results. In contrast, the sequential approach has better between-cluster average distance compared to the concurrent approach. This is explained by looking at the representative inequality patterns from the concurrent approach on Figs. 8 and 9, where there are many clusters that have similar inequality patterns. To this end, it is interesting to observe in more details the (dis)similarity of the resulting narratives from the two approaches.

5.2. Comparison of the resulting scenarios

In this section we compare the clusters of inequality patterns from the two approaches as well as the narratives behind each cluster. First, we see that the clusters of inequality patterns from the concurrent approach have a higher degree of variation as there are several clusters that have a comparable pattern. However, most of the patterns identified from the concurrent approach are also present in the sequential approach. For example, the inequality patterns of cluster 4, 6, 15, and 18 from the concurrent approach are comparable to the inequality pattern of cluster 1 from the sequential approach. Table 2 lists the other pairs of similar inequality patterns identified by the two approaches.

Two exceptions worth noting are cluster 15 and 16 from the concurrent approach. In general, cluster 16 has a similar inequality pattern to cluster 2 from the sequential approach. However, the most profitable districts in cluster 16 are the two districts in the westernmost part of the region. Furthermore, the easternmost district is also slightly better-off than many of the other districts. Cluster 15 has similar inequality pattern to cluster 1 from the sequential approach. The difference is that many districts to the north of the river are also relatively better-off in cluster 15.

Most of the narratives behind each inequality pattern identified by the two approaches are also comparable (see Table 2). For example, cluster 4 and 5 from the sequential approach have similar narratives to their counterparts from the concurrent approach. Since in the concurrent approach we do not limit our analysis to only four most important factors, this approach yields slightly richer and more detailed narratives for some of the clusters. In the concurrent approach, the low dikes policy in Dong Thap is identified as an important part of the narratives for some of the clusters (i.e., cluster 1, 10, and 16 from the concurrent approach).

The more aggregated results of the sequential approach conceal some diversity within the scenarios. For example, maintaining the current dike system in combination with a medium to high degree of upstream dam development is part of the narrative for cluster 6 from the sequential approach. However, the same narrative leads to different inequality patterns if we follow the decision tree from the concurrent approach (i.e., cluster 14 and 15). This is because combining the inequality patterns from cluster 14 and 15 of the concurrent approach will average out the profitability of districts that are better-off in each cluster, resulting in a more equal distribution as exemplified by the representative inequality pattern of cluster 6 from the sequential approach.

5.3. Reflection for practice

The comparisons above show that the concurrent approach outperforms the sequential approach. However, it comes with a caveat of having a larger number of final narratives. This raises the question of whether the benefit of better input space separability in the concurrent approach does outweigh the drawback of having more clusters and narratives. To answer this, we need to first revisit the main purpose of scenario discovery itself, which is to craft narratives about system outcomes under certain combinations of uncertainties/polices (Bryant and Lempert, 2010; Greeven et al., 2016; Lempert et al., 2006). In particular, attention needs to be given on the decision-making contexts, and on the use-case of the narratives generated from the multiclass scenario discovery exercise.

Past scenario discovery studies have a varying level of stakeholder involvement. Some studies indicate a relatively low degree of interactions with stakeholder (e.g., Hidayatno et al., 2020; Lamontagne et al., 2018; Moallemi et al., 2017). In these studies, the generated narratives are mainly aimed at defining plausible future pathways, which are to be used by other institutions in other contexts. Other studies indicate a more frequent and thorough interactions (e.g., Groves et al., 2019; Hamarat et al., 2013; Trindade et al., 2019). The aim of such studies is often more specific, such as for stress testing alternative policies and identifying vulnerabilities. Accordingly, narratives generated in such use-cases are used solely for the purpose of the project. The sequential approach, with a relatively lower number of narratives, is more suitable for the former type of use-cases (relatively little stakeholder engagement, narratives to be transferred for other contexts). The concurrent approach, with better separability performance but more narratives, is more suitable for the latter type of use-cases (more intense stakeholder engagement, more focused analysis).

In addition to the characteristics of the use-cases, there are two further important points to note. First, an important strength of scenario discovery is to facilitate deliberation, and this obviously requires thorough engagements with clients and stakeholders. Accordingly, narratives from scenario discovery should not be shared as-is with stakeholders. Rather, the analyst should always be at the interface

Table 2
Comparable inequality patterns and narratives from the sequential and the concurrent approach.

Cluster from the sequential approach	Comparable cluster from the concurrent approach	Narratives from the sequential approach	Narratives from the concurrent approach
1	4, 6, 18	- High to very high degree of upstream dam development and high dikes in An Giang and/or Dong Thap - Very high degree of upstream dam development and maintaining current dike system	- Medium to high degree of upstream dam development and high dikes in An Giang and Dong Thap - Very high degree of upstream dam development and high dikes in An Giang - Very high degree of upstream dam development and maintaining current dike system
2	10, 12	- Low dikes in An Giang - Low to medium degree of upstream dam development and maintaining current dike system	- Low dikes in An Giang and in Dong Thap - Medium degree of upstream dam development and low dikes in An Giang
3	1, 3, 5	- Low to high degree of dam development and high dikes in An Giang	- Low degree of upstream dam development, high dikes in An Giang and low dikes in Dong Thap - Low degree of upstream dam development and high dikes in Dong Thap - Medium to high degree of upstream dam development and high dikes in An Giang
4	17	- Very high degree of upstream dam development and low dikes in An Giang	- Very high degree of upstream dam development and low dikes in An Giang
5	8, 9	- Medium degree of upstream dam development and high dikes in Dong Thap	- Medium degree of upstream dam development and high dikes in Dong Thap
6	2	- Medium to high degree of upstream dam development and maintaining current dike system - Low degree of upstream dam development and high dikes either only in An Giang or both in An Giang and Dong Thap	- Low degree of upstream dam development and high dikes both in An Giang and Dong Thap
7	7, 11, 13, 14	- Low to medium degree of upstream dam development and maintaining current dike system or low dikes in An Giang - Low degree of upstream dam development and high dikes in Dong Thap	- Low degree of upstream dam development in combination with either high dikes in Dong Thap or low dikes in An Giang - Low to medium degree of upstream dam development in combination with either low dikes in Dong Thap or maintaining the current dike system

between the policy problem and the tool used to support the policy analysis (Cuppen et al., 2021). So, the issue with having a larger number of scenarios is that the analyst might have to do more work to distill the message from the analysis before conveying it to others, and this can be done through consultation with the stakeholders.

Second, as some clusters from the concurrent approach have similar inequality patterns, presenting them simultaneously might not be appropriate. Without the help of the results from the sequential approach, the regrouping of similar clusters from the concurrent approach can be performed through either root-first or leaves first analysis (Smith et al., 2019). In leaves-first analysis, the key step is to identify clusters with similar representative inequality patterns. This can be done qualitatively through visual inspection (as done in Table 2) or through consultation with stakeholders. To aid this process, the analyst can calculate the average distance between any pair of clusters and combine those with relatively low distance. The final choice of the number of narratives should not be the analyst's call, but instead, decided in a participatory and interactive setting with stakeholders. If what is of more interest is the narratives, instead of the resulting inequality patterns, the root-first analysis can be followed instead.

6. Conclusion

Adaptation policies and uncertainties, and the interactions between them, almost unavoidably yield unequal consequences to different people. The task of exploring future inequality patterns and understanding their drivers fits the nature of scenario discovery. In scenario discovery, one maps back the output space of a model (in this case, inequality patterns) with its input space (policy levers and exogenous uncertain factors). In this study, we contribute to the advancement of scenario discovery in two ways. First, we propose two novel criteria to evaluate the quality of multiclass scenario discovery results: output space separability and the number of resulting narratives. Second, we propose a novel concurrent approach for multiclass scenario discovery by using Multivariate Regression Trees (MRT).

Using agriculture adaptation planning for the Vietnam Mekong Delta as a case study, we demonstrate the application of both the established sequential and the novel concurrent approach for multiclass scenario discovery. We find that the concurrent approach performs considerably better in terms of input space separability. The MRT algorithm guarantees a perfect separation of the input space when clustering the simulation results. This, however, does result in a larger number of clusters of output space, and subsequently, narratives. While the sequential approach results in seven scenarios, the concurrent approach produces eighteen scenarios. Both approaches have a fairly comparable output space separability performance, with the sequential approach results in better between-cluster dissimilarity and the concurrent approach results in better within-class similarity. Despite the differences in performance, we show how most of the narratives and representative inequality patterns identified by the two approaches are similar, with some exceptions. The concurrent approach provides richer insights as it unravels two additional representative inequality patterns that are not captured by the sequential approach.

Based on the case study results, we argue for the use of the concurrent approach for future multiclass scenario discovery. The concurrent

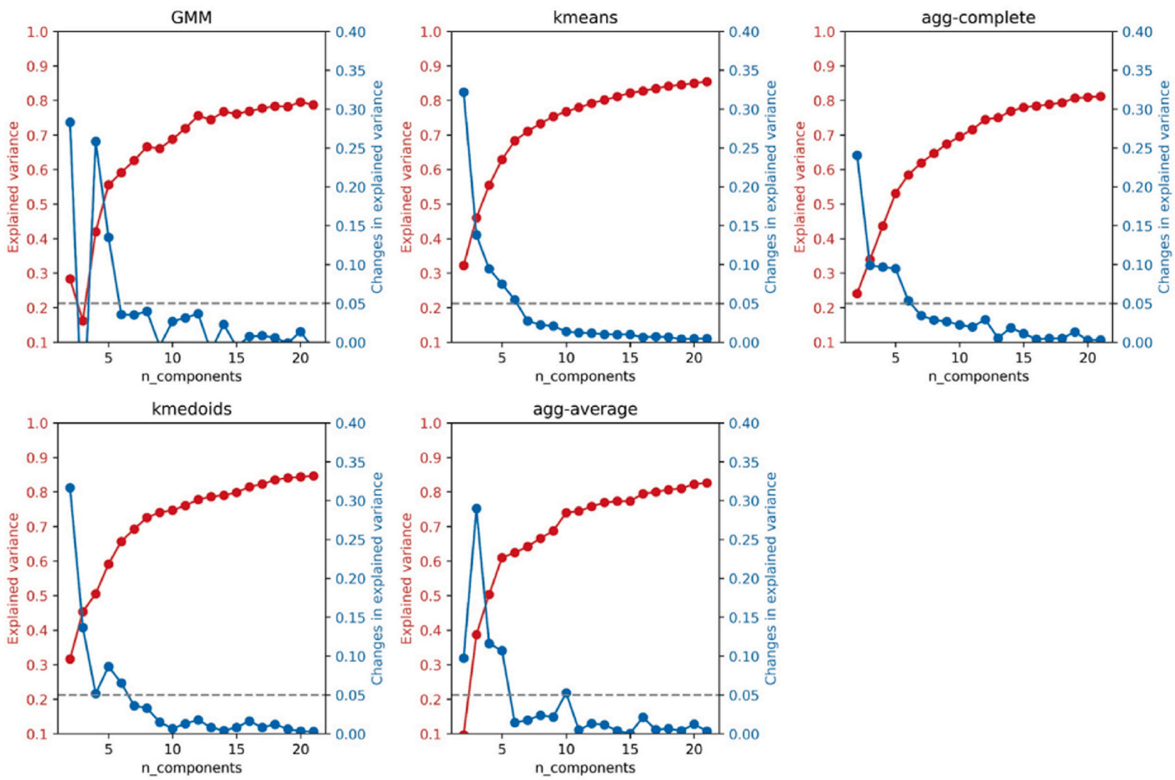

Fig. 10. Explained variance for varying numbers of clusters using five different clustering algorithms. The horizontal dashed line is the 5% threshold of delta/changes of explained variance used to determine the optimal number of clusters.

Table 3
Explained variance of each clustering algorithm for the selected number of clusters.

Algorithm	Selected number of clusters	Explained variance
Gaussian Mixture Model	7	0.626
k-Means	7	0.711
k-Medoids	7	0.693
Agglomerative clustering – complete linkage	7	0.619
Agglomerative clustering – average linkage	6	0.624

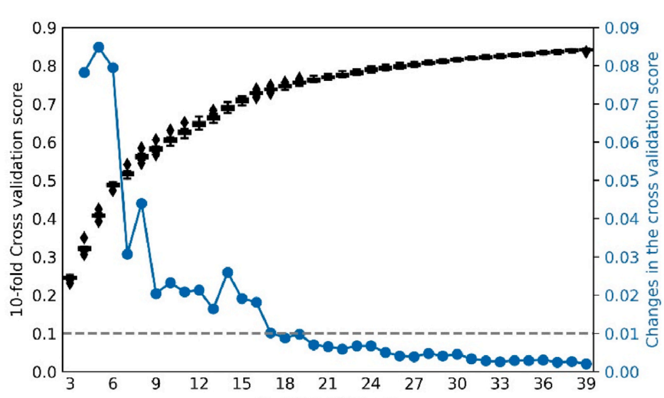

Fig. 11. 10-fold cross validation scores of the MRT with increasing tree sizes. The left-hand y-axis corresponds to the boxplot of the 10-fold cross validation score, while the right-hand y-axis corresponds to the points of changes in the cross validation scores. The horizontal dashed line shows the 0.01 threshold of the changes of cross validation scores.

approach guarantees perfect input space separability without sacrificing too much in terms of output space separability. Furthermore, the concurrent approach captures richer and more distinctive trade-off patterns between outcome variables (in our case, inequality patterns) compared to the sequential approach. One caveat is that the concurrent approach requires one to make extra effort to distill insights from these richer results.

In light of the presented results, we see several directions for future research. The first one is related to the selection of representative inequality patterns. In this study, we take a pragmatic approach by using the medoid scenario in each cluster. Other approaches include averaging the relative profitability of each actor across all scenarios in a cluster, or selecting the scenario which has the most distinctive inequality pattern relative to the other clusters (Carlsen et al., 2016). The second direction is assessing the limits and scalability of clustering when a higher number of stakeholders, which leads to a larger number of outcome variables, is considered. While alternative high-dimensional clustering techniques are available (Kriegel et al., 2009; Xu and Tian, 2015), their usefulness in the context of scenario discovery remains to be evaluated. The third direction is to assess the impacts of different spatial aggregations. As we aggregate farms profitability at a district level, within-district inequality is ignored. The statistical bias resulting from spatial aggregation, well-known as the modifiable areal unit problem (Fotheringham and Wong, 1991), can have profound implications for the emerging spatial pattern. Sensitivity or robustness analysis could be applied to understand the stability of the representative inequality patterns under different aggregation levels.

Inequalities can be viewed from various dimensions (across people in different locations (interregional), with different income, different socioeconomic background, or across actors) and variables (inequality of profitability, benefits from policies, exposure to and impacts of climate change) (Harrison et al., 2016; Jafino et al., 2021b; Rao et al., 2017). Irrespective of the dimension and variable of inequality, there is still a methodological need to explore plausible inequality patterns to support equitable adaptation planning. For the purpose of showing the merits of multiclass scenario discovery for this methodological need, we used one dimension of inequality (interregional inequality of profitability). Without loss of generality, the sequential and concurrent approaches could be applied to other conceptualizations of inequality, as we only need to slice the population differently based on our variables of interest. However, it is important to highlight the limitation of this approach. In planning for climate change, distributional consequences can be seen from intra-generational (between people, and assuming they live within the same generation) and intergenerational (between generations) perspectives (Jafino et al., 2021b). Multiclass scenario discovery is applicable for exploring intra-generational, but not intergenerational inequalities. The topic of discounting is more applicable for the latter, with recent works proposing alternative discounting methods that account for equity (Asheim, 2017; Dietz and Asheim, 2012).

Software and data availability

The code for conducting both sequential and concurrent approaches, as well as the data required to perform the analysis can be accessed at https://github.com/bramkaarga/inequality_pattern_exploration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by NWO Top Sector Water Call: Adaptation Pathways for socially inclusive development of urbanizing deltas (research number OND1362814).

Appendix A. Details of clustering results from the sequential approach

The first step in the sequential approach is determining the clustering algorithm and the number of clusters to proceed with. For each algorithm, we perform clustering with an increasing number of clusters from 2 to 21. We calculate the explained variance for each number of cluster (Fig. 10). By sweeping across different numbers of clusters we can observe the progression and the convergence of the explained variance. At the end of the iteration, i.e., with 21 clusters, the explained variance from all algorithms clusters converges to 0.8. As explained in the Methods section, we set a threshold of 0.05 for the changes in explained variance in order to select an optimal number of clusters from each algorithm.

Note that the selection of the threshold T for the changes in explained variance is a subjective choice. We need to balance the explained variance of the selected number of clusters at which the threshold T is being met, the potential gain in explained variance when using a higher number of clusters, and the potential loss in interpretability when a higher number clusters is used. Table 3 shows the number of clusters from each clustering algorithm when the threshold $T = 0.05$ is met and the corresponding explained variance. K-means algorithm yields the best performance. It performs slightly better than k-Medoids and clearly outperforms the other clustering algorithms. Its explained variance of 0.711 is also not too distant from the overall explained variance convergence of 0.8. Hence, in the remainder of this sequential approach we proceed with the 7 clusters of output space as identified by the k-Means algorithm.

Appendix B. Details of tree selection in the concurrent approach

In the concurrent approach we start directly with determining the size of the tree based on the evolution of the cross-validation scores. Fig. 11 shows the increase of the 10-fold cross validation scores with increasing number of leaves. Similar to the clustering results in the sequential approach, the cross-validation score seems to converge to 0.8. However, the cross-validation score has not stagnated yet even after being grown to having 40 leaves. As the score keeps increasing even after the tree has become quite complex, it is advised to set a threshold of increase in cross validation scores in order to select an appropriate tree size (Smith et al., 2019). We choose a threshold of 0.01 (dashed line on Fig. 11) and this threshold is reached when the number of leaves is 18.

References

Adler, M., Anthoff, D., Bosetti, V., Garner, G., Keller, K., Treich, N., 2017. Priority for the worse-off and the social cost of carbon. Nat. Clim. Change 7 (6), 443–449. https://doi.org/10.1038/nclimate3298.

Asheim, G.B., 2017. Sustainable growth. Soc. Choice Welfare 49 (3), 825–848. https://doi.org/10.1007/s00355-016-0977-9.

Bandaru, S., Ng, A.H.C., Deb, K., 2017. Data mining methods for knowledge discovery in multi-objective optimization: Part B - new developments and applications. Expert Syst. Appl. 70, 119–138. https://doi.org/10.1016/j.eswa.2016.10.016.

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

Bárcena, J.F., Camus, P., García, A., Álvarez, C., 2015. Selecting model scenarios of real hydrodynamic forcings on mesotidal and macrotidal estuaries influenced by river discharges using K-means clustering. Environ. Model. Software 68, 70–82. https://doi.org/10.1016/j.envsoft.2015.02.007.

Bartholomew, E., Kwakkel, J.H., 2020. On considering robustness in the search phase of robust decision making: a comparison of many-objective robust decision making, multi-scenario many-objective robust decision making, and many objective robust optimization. Environ. Model. Software 127, 104699. https://doi.org/10.1016/j.envsoft.2020.104699.

Berntsen, P.B., Trutnevyte, E., 2017. Ensuring diversity of national energy scenarios: bottom-up energy system model with Modeling to Generate Alternatives. Energy 126, 886–898. https://doi.org/10.1016/j.energy.2017.03.043.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC press, Florida.

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

Carlsen, H., Lempert, R., Wikman-Svahn, P., Schweizer, V., 2016. Choosing small sets of policy-relevant scenarios by combining vulnerability and diversity approaches. Environ. Model. Software 84, 155–164. https://doi.org/10.1016/j.envsoft.2016.06.011.

Chapman, A., Darby, S., 2016. Evaluating sustainable adaptation strategies for vulnerable mega-deltas using system dynamics modelling: rice agriculture in the Mekong Delta's an Giang Province, Vietnam. Sci. Total Environ. 559, 326–338. https://doi.org/10.1016/j.scitotenv.2016.02.162.

Chapman, A.D., Darby, S.E., Hồng, H.M., Tompkins, E.L., Van, T.P.D., 2016. Adaptation and development trade-offs: fluvial sediment deposition and the sustainability of rice-cropping in an Giang Province, Mekong Delta. Climatic Change 137 (3–4), 593–608. https://doi.org/10.1007/s10584-016-1684-3.

Ciullo, A., Kwakkel, J.H., De Bruijn, K.M., Doorn, N., Klijn, F., 2020. Efficient or fair? Operationalizing ethical principles in flood risk management: a case study on the Dutch-German rhine. Risk Anal. https://doi.org/10.1111/risa.13527.

Cuppen, E., Nikolic, I., Kwakkel, J., Quist, J., 2021. Participatory multi-modelling as the creation of a boundary object ecology: the case of future energy infrastructures in the Rotterdam Port Industrial Cluster. Sustainability Science 16 (3), 901–918. https://doi.org/10.1007/s11625-020-00873-z.

De'ath, G., 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 83 (4), 1105–1117. https://doi.org/10.1890/0012-9658(2002)083[1105:MRTANT]2.0.CO;2.

De'ath, G., 2007. Boosted trees for ecological modeling and prediction. Ecology 88 (1), 243–251. https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2.

Dietz, S., Asheim, G.B., 2012. Climate policy under sustainable discounted utilitarianism. J. Environ. Econ. Manag. 63 (3), 321–335. https://doi.org/10.1016/j.jeem.2012.01.003.
Dutta, D., Herath, S., Musiake, K., 2003. A mathematical model for flood loss estimation. J. Hydrol. 277 (1), 24–49. https://doi.org/10.1016/S0022-1694(03)00084-2.
Eker, S., van Daalen, E., 2015. A model-based analysis of biomethane production in The Netherlands and the effectiveness of the subsidization policy under uncertainty. Energy Pol. 82, 178–196. https://doi.org/10.1016/j.enpol.2015.03.019.
Fotheringham, A.S., Wong, D.W.S., 1991. The modifiable areal unit problem in multivariate statistical analysis. Environ. Plann. 23 (7), 1025–1044.
Franssen, M., 2005. Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. Res. Eng. Des. 16 (1–2), 42–56. https://doi.org/10.1007/s00163-004-0057-5.
Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139. https://doi.org/10.1006/jcss.1997.1504.
Friedman, J.H., Fisher, N.I., 1999. Bump hunting in high-dimensional data. Stat. Comput. 9 (2), 123–143. https://doi.org/10.1023/A:1008894516817.
GAEN-View, 2013. Viewing system for global agriculture and environment: using HighFrequency earth observation satellites. Retrieved from. http://gaenview.dc.affrc.go.jp.
Gardiner, S.M., 2010. Ethics and climate change: an introduction. WIREs Climate Change 1 (1), 54–66. https://doi.org/10.1002/wcc.16.
Garschagen, M., Diez, J.R., Nhan, D.K., Kraas, F., 2012. Socio-Economic development in the mekong delta: between the prospects for progress and the realms of reality. In: Renaud, F.G., Kuenzer, C. (Eds.), The Mekong Delta System: Interdisciplinary Analyses of a River Delta. Springer Netherlands, Dordrecht, pp. 83–132.
Gerst, M.D., Wang, P., Borsuk, M.E., 2013. Discovering plausible energy and economic futures under global change using multidimensional scenario discovery. Environ. Model. Software 44, 76–86. https://doi.org/10.1016/j.envsoft.2012.09.001.
Giudici, F., Castelletti, A., Giuliani, M., Maier, H.R., 2020. An active learning approach for identifying the smallest subset of informative scenarios for robust planning under deep uncertainty. Environ. Model. Software 127, 104681. https://doi.org/10.1016/j.envsoft.2020.104681.
Gourevitch, J.D., Singh, N.K., Minot, J., Raub, K.B., Rizzo, D.M., Wemple, B.C., Ricketts, T.H., 2020. Spatial targeting of floodplain restoration to equitably mitigate flood risk. Global Environ. Change 61, 102050. https://doi.org/10.1016/j.gloenvcha.2020.102050.
Grasso, M., 2007. A normative ethical framework in climate change. Climatic Change 81 (3–4), 223–246. https://doi.org/10.1007/s10584-006-9158-7.
Green, D., 2016. The spatial distribution of extreme climate events, another climate inequity for the world's most vulnerable people. Environ. Res. Lett. 11 (9), 091002.
Greeven, S., Kraan, O., Chappin, É.J., 2016. The emergence of climate change and mitigation action by society: an agent-based scenario discovery study. J. Artif. Soc. Soc. Simulat. 19 (3), 9.
Groves, D.G., Bonzanigo, L., Syme, J., Engle, N.L., Rodriguez Cabanillas, I., 2019. Preparing for Future Droughts in Lima, Peru: Enhancing Lima's Drought Management Plan to Meet Future Challenges. World Bank, Washington, DC. Retrieved from: https://openknowledge.worldbank.org/handle/10986/31695.
Groves, D.G., Lempert, R.J., 2007. A new analytic method for finding policy-relevant scenarios. Global Environ. Change 17 (1), 73–85. https://doi.org/10.1016/j.gloenvcha.2006.11.006.
GSO, 2019. *Statistical Yearbook of Vietnam 2018*. Vietnam: general statistics office of Viet Nam. Retrieved from. https://www.gso.gov.vn/default_en.aspx?tabid=515&idmid=5&ItemID=19299.
Guivarch, C., Rozenberg, J., Schweizer, V., 2016. The diversity of socio-economic pathways and CO2 emissions scenarios: insights from the investigation of a scenarios database. Environ. Model. Software 80, 336–353. https://doi.org/10.1016/j.envsoft.2016.03.006.
Halim, R.A., Kwakkel, J.H., Tavasszy, L.A., 2015. A Scenario Discovery Study of the Impact of Uncertainties in the Global Container Transport System on European Ports. Futures.
Hamarat, C., Kwakkel, J.H., Pruyt, E., 2013. Adaptive robust design under deep uncertainty. Technol. Forecast. Soc. Change 80 (3), 408–418. https://doi.org/10.1016/j.techfore.2012.10.004.
Hansson, S.O., 2007. Philosophical problems in cost–benefit analysis. Econ. Philos. 23 (2), 163–183. https://doi.org/10.1017/S0266267107001356.
Harrison, P.A., Dunford, R.W., Holman, I.P., Rounsevell, M.D.A., 2016. Climate change impact modelling needs to include cross-sectoral interactions. Nat. Clim. Change 6, 885. https://doi.org/10.1038/nclimate3039.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media, New York.
Herman, J., Reed, P., Zeff, H., Characklis, G., 2015. How should robustness Be defined for water systems planning under change? J. Water Resour. Plann. Manag. 141 (10), 04015012 https://doi.org/10.1061/(ASCE)WR.1943-5452.0000509.
Hidayatno, A., Jafino, B.A., Setiawan, A.D., Purwanto, W.W., 2020. When and why does transition fail? A model-based identification of adoption barriers and policy vulnerabilities for transition to natural gas vehicles. Energy Pol. 138, 111239. https://doi.org/10.1016/j.enpol.2020.111239.
Hoang, L.P., van Vliet, M.T.H., Kummu, M., Lauri, H., Koponen, J., Supit, I., Ludwig, F., 2019. The Mekong's future flows under multiple drivers: how climate change, hydropower developments and irrigation expansions drive hydrological changes. Sci. Total Environ. 649, 601–609. https://doi.org/10.1016/j.scitotenv.2018.08.160.
Jafino, B.A., Kwakkel, J., Klijn, F., Dung, N.V., van Delden, H., Haasnoot, M., Sutanudjaja, E., 2021a. Accounting for multisectoral dynamics in supporting equitable adaptation planning: a case study on the rice agriculture in the Vietnam Mekong Delta. Earth's Future 9 (5), e2020EF001939. https://doi.org/10.1029/2020EF001939.
Jafino, B.A., Kwakkel, J.H., Taebi, B., 2021b. Enabling assessment of distributive justice through models for climate change planning: a review of recent advances and a research agenda. Wiley Interdisciplinary Reviews: Climate Change, e721. https://doi.org/10.1002/wcc.721.
Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recogn. Lett. 31 (8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011.
Käkönen, M., 2008. Mekong delta at the crossroads: more control or adaptation? AMBIO A J. Hum. Environ. 37 (3), 205–212, 208.
Kasprzyk, J.R., Nataraj, S., Reed, P.M., Lempert, R.J., 2013. Many objective robust decision making for complex environmental systems undergoing change. Environ. Model. Software 42, 55–71. https://doi.org/10.1016/j.envsoft.2012.12.007.
Kasprzyk, J.R., Reed, P.M., Hadka, D.M., 2016. Battling arrow's paradox to discover robust water management alternatives. J. Water Resour. Plann. Manag. 142 (2), 04015053 https://doi.org/10.1061/(ASCE)WR.1943-5452.0000572.
Ketchen, D.J., Shook, C.L., 1996. The application of cluster analysis in strategic management research: an analysis and critique. Strat. Manag. J. 17 (6), 441–458. https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G.
Kind, J., Wouter Botzen, W.J., Aerts, J.C.J.H., 2017. Accounting for risk aversion, income distribution and social welfare in cost-benefit analysis for flood risk management. Wiley Interdisciplinary Reviews: Climate Change 8 (2). https://doi.org/10.1002/wcc.446.
Knox, J.W., Haro-Monteagudo, D., Hess, T.M., Morris, J., 2018. Identifying trade-offs and reconciling competing demands for water - integrating agriculture into a robust decision-making framework. Earth's Future. https://doi.org/10.1002/2017ef000741.
Kriegel, H.-P., Kröger, P., Zimek, A., 2009. Clustering high-dimensional data: a survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Trans. Knowl. Discov. Data 3 (1), 1–58.
Kwakkel, J.H., 2017. The Exploratory Modeling Workbench: an open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. Environ. Model. Software 96, 239–250. https://doi.org/10.1016/j.envsoft.2017.06.054.
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. https://doi.org/10.1016/j.techfore.2012.09.012.
Kwakkel, J.H., Jaxa-Rozen, M., 2016. Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. Environ. Model. Software 79, 311–321. https://doi.org/10.1016/j.envsoft.2015.11.020.
Lamontagne, J.R., Reed, P.M., Link, R., Calvin, K.V., Clarke, L.E., Edmonds, J.A., 2018. Large ensemble analytic framework for consequence-driven discovery of climate change scenarios. Earth's Future 6 (3), 488–504. https://doi.org/10.1002/2017EF000701.
Larsen, D.R., Speckman, P.L., 2004. Multivariate regression trees for analysis of abundance data. Biometrics 60 (2), 543–549. https://doi.org/10.1111/j.0006-341X.2004.00202.x.
Lauri, H., de Moel, H., Ward, P.J., Räsänen, T.A., Keskinen, M., Kummu, M., 2012. Future changes in Mekong River hydrology: impact of climate change and reservoir operation on discharge. Hydrol. Earth Syst. Sci. 16 (12), 4603–4619. https://doi.org/10.5194/hess-16-4603-2012.
Lempert, R.J., Groves, D.G., Popper, S.W., Bankes, S.C., 2006. A general, analytic method for generating robust strategies and narrative scenarios. Manag. Sci. 52 (4), 514–528.
Lempert, R.J., Popper, S.W., Bankes, S.C., 2003. Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis. Rand Corporation.
Manh, N.V., Dung, N.V., Hung, N.N., Kummu, M., Merz, B., Apel, H., 2015. Future sediment dynamics in the Mekong Delta floodplains: impacts of hydropower development, climate change and sea level rise. Global Planet. Change 127, 22–33. https://doi.org/10.1016/j.gloplacha.2015.01.001.
Manh, N.V., Dung, N.V., Hung, N.N., Merz, B., Apel, H., 2014. Large-scale suspended sediment transport and sediment deposition in the Mekong Delta. Hydrol. Earth Syst. Sci. 18 (8), 3033–3053. https://doi.org/10.5194/hess-18-3033-2014.
Manh, N.V., Merz, B., Apel, H., 2013. Sedimentation monitoring including uncertainty analysis in complex floodplains: a case study in the Mekong Delta. Hydrol. Earth Syst. Sci. 17 (8), 3039–3057. https://doi.org/10.5194/hess-17-3039-2013.
Mekong Delta Plan Consortium, 2013. Mekong Delta Plan: long-term vision and strategy for a safe, prosperous and sustainable delta. https://www.wur.nl/upload_mm/2/c/3/b5f2e669-cb48-4ed7-afb6-682f5216fe7d_mekong.pdf.
Minderhoud, P.S.J., Coumou, L., Erban, L.E., Middelkoop, H., Stouthamer, E., Addink, E.A., 2018. The relation between land use and subsidence in the Vietnamese Mekong delta. Sci. Total Environ. 634, 715–726. https://doi.org/10.1016/j.scitotenv.2018.03.372.
Moallemi, E.A., de Haan, F., Kwakkel, J., Aye, L., 2017. Narrative-informed exploratory analysis of energy transition pathways: a case study of India's electricity sector. Energy Pol. 110, 271–287. https://doi.org/10.1016/j.enpol.2017.08.019.
Moallemi, E.A., Elsawah, S., Ryan, M.J., 2018. An agent-monitored framework for the output-oriented design of experiments in exploratory modelling. Simulat. Model. Pract. Theor. 89, 48–63. https://doi.org/10.1016/j.simpat.2018.09.008.
Moallemi, E.A., Kwakkel, J., de Haan, F.J., Bryan, B.A., 2020. Exploratory modeling for analyzing coupled human-natural systems under uncertainty. Global Environ. Change 65, 102186. https://doi.org/10.1016/j.gloenvcha.2020.102186.
Ndong, G.O., Villerd, J., Cousin, I., Therond, O., 2020. Using a multivariate regression tree to analyze trade-offs between ecosystem services: application to the main

cropping area in France. Sci. Total Environ. 142815. https://doi.org/10.1016/j.scitotenv.2020.142815.
Ngan, L.T., Bregt, A.K., van Halsema, G.E., Hellegers, P.J.G.J., Nguyen, L.-D., 2018. Interplay between land-use dynamics and changes in hydrological regime in the Vietnamese Mekong Delta. Land Use Pol. 73, 269–280. https://doi.org/10.1016/j.landusepol.2018.01.030.
O'Neill, B.C., Kriegler, E., Ebi, K.L., Kemp-Benedict, E., Riahi, K., Rothman, D.S., Solecki, W., 2017. The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century. Global Environ. Change 42, 169–180. https://doi.org/10.1016/j.gloenvcha.2015.01.004.
Okereke, C., 2010. Climate justice and the international regime. WIREs Climate Change 1 (3), 462–474. https://doi.org/10.1002/wcc.52.
Paavola, J., Adger, W.N., 2006. Fair adaptation to climate change. Ecol. Econ. 56 (4), 594–609. https://doi.org/10.1016/j.ecolecon.2005.03.015.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Dubourg, V., 2011. Scikit-learn: machine learning in Python. the Journal of machine Learning research 12, 2825–2830.
Rao, N.D., 2013. Distributional impacts of climate change mitigation in Indian electricity: the influence of governance. Energy Pol. 61, 1344–1356. https://doi.org/10.1016/j.enpol.2013.05.103.
Rao, N.D., van Ruijven, B.J., Riahi, K., Bosetti, V., 2017. Improving poverty and inequality modelling in climate research. Nat. Clim. Change 7 (12), 857.
Rohmer, J., Idier, D., Paris, F., Pedreros, R., Louisor, J., 2018. Casting light on forcing and breaching scenarios that lead to marine inundation: combining numerical simulations with a random-forest classification approach. Environ. Model. Software 104, 64–80. https://doi.org/10.1016/j.envsoft.2018.03.003.
Rozenberg, J., Guivarch, C., Lempert, R., Hallegatte, S., 2014. Building SSPs for climate policy analysis: a scenario elicitation methodology to map the space of possible future challenges to mitigation and adaptation. Climatic Change 122 (3), 509–522. https://doi.org/10.1007/s10584-013-0904-3.
Sakamoto, T., Cao Van, P., Kotera, A., Nguyen Duy, K., Yokozawa, M., 2009. Detection of yearly change in farming systems in the Vietnamese mekong delta from MODIS time-series imagery. Jpn. Agric. Res. Q.: JARQ 43 (3), 173–185. https://doi.org/10.6090/jarq.43.173.
Sattari, S.Z., van Ittersum, M.K., Bouwman, A.F., Smit, A.L., Janssen, B.H., 2014. Crop yield response to soil fertility and N, P, K inputs in different environments: testing and improving the QUEFTS model. Field Crop. Res. 157, 35–46. https://doi.org/10.1016/j.fcr.2013.12.005.
Sayers, P., Penning-Rowsell, E.C., Horritt, M., 2018. Flood vulnerability, risk, and social disadvantage: current and future patterns in the UK. Reg. Environ. Change 18 (2), 339–352. https://doi.org/10.1007/s10113-017-1252-z.
Schapire, R., Freund, Y., 2012. Boosting: Foundations and Algorithms. MIT Press, Cambridge, Massachusetts.
Schweizer, V., 2018. A few scenarios still do not fit all. Nat. Clim. Change 8 (5), 361–362. https://doi.org/10.1038/s41558-018-0148-3.
Shortridge, J.E., Zaitchik, B.F., 2018. Characterizing climate change risks by linking robust decision frameworks and uncertain probabilistic projections. Climatic Change 151 (3), 525–539. https://doi.org/10.1007/s10584-018-2324-x.
Smith, R., Kasprzyk, J., Rajagopalan, B., 2019. Using multivariate regression trees and multiobjective tradeoff sets to reveal fundamental insights about water resources systems. Environ. Model. Software 120, 104498. https://doi.org/10.1016/j.envsoft.2019.104498.
Son, N.-T., Chen, C.-F., Chen, C.-R., Duc, H.-N., Chang, L.-Y., 2013. A phenology-based classification of time-series MODIS data for rice crop monitoring in Mekong Delta, Vietnam. Rem. Sens. 6 (1), 135–156. https://doi.org/10.3390/rs6010135.
Steinmann, P., Auping, W.L., Kwakkel, J.H., 2020. Behavior-based scenario discovery using time series clustering. Technol. Forecast. Soc. Change 156, 120052. https://doi.org/10.1016/j.techfore.2020.120052.
Sutanudjaja, E.H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J.H.C., Drost, N., Bierkens, M.F.P., 2018. PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. Geosci. Model Dev. (GMD) 11 (6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018.
Szekely, G.J., Rizzo, M.L., 2005. Hierarchical clustering via joint between-within distances: extending ward's minimum variance method. J. Classif. 22 (2), 151–183. https://doi.org/10.1007/s00357-005-0012-9.
Taconet, N., Méjean, A., Guivarch, C., 2020. Influence of climate change impacts and mitigation costs on inequality between countries. Climatic Change 160 (1), 15–34. https://doi.org/10.1007/s10584-019-02637-w.
Thomas, K., Hardy, R.D., Lazrus, H., Mendez, M., Orlove, B., Rivera-Collazo, I., Winthrop, R., 2019. Explaining differential vulnerability to climate change: a social science review. WIREs Climate Change 10 (2), e565. https://doi.org/10.1002/wcc.565.
Toan, T.Q., 2014. 9 - climate change and sea level rise in the mekong delta: flood, tidal inundation, salinity intrusion, and irrigation adaptation methods. In: Thao, N.D., Takagi, H., Esteban, M. (Eds.), Coastal Disasters and Climate Change in Vietnam. Elsevier, Oxford, pp. 199–218.
Tran, D.D., van Halsema, G., Hellegers, P.J.G.J., Ludwig, F., Wyatt, A., 2018. Questioning triple rice intensification on the Vietnamese mekong delta floodplains: an environmental and economic analysis of current land-use trends and alternatives. J. Environ. Manag. 217, 429–441. https://doi.org/10.1016/j.jenvman.2018.03.116.
Triet, N.V.K., Dung, N.V., Hoang, L.P., Duy, N.L., Tran, D.D., Anh, T.T., Apel, H., 2020. Future projections of flood dynamics in the Vietnamese Mekong Delta. Sci. Total Environ. 742, 140596. https://doi.org/10.1016/j.scitotenv.2020.140596.
Triet, N.V.K., Dung, N.V., Merz, B., Apel, H., 2018. Towards risk-based flood management in highly productive paddy rice cultivation – concept development and application to the Mekong Delta. Nat. Hazards Earth Syst. Sci. 18 (11), 2859–2876. https://doi.org/10.5194/nhess-18-2859-2018.
Trindade, B.C., Reed, P.M., Characklis, G.W., 2019. Deeply uncertain pathways: integrated multi-city regional water supply infrastructure investment and portfolio management. Adv. Water Resour. 134, 103442. https://doi.org/10.1016/j.advwatres.2019.103442.
Weaver, C.P., Lempert, R.J., Brown, C., Hall, J.A., Revell, D., Sarewitz, D., 2013. Improving the contribution of climate model information to decision making: the value and demands of robust decision frameworks. Wiley Interdisciplinary Reviews: Climate Change 4 (1), 39–60. https://doi.org/10.1002/wcc.202.
Winsemius, H.C., Van Beek, L.P.H., Jongman, B., Ward, P.J., Bouwman, A., 2013. A framework for global river flood risk assessments. Hydrol. Earth Syst. Sci. 17 (5), 1871–1892. https://doi.org/10.5194/hess-17-1871-2013.
Witt, C., Dobermann, A., Abdulrachman, S., Gines, H.C., Guanghuo, W., Nagarajan, R., Olk, D.C., 1999. Internal nutrient efficiencies of irrigated lowland rice in tropical and subtropical Asia. Field Crop. Res. 63 (2), 113–138. https://doi.org/10.1016/S0378-4290(99)00031-3.
Xu, D., Tian, Y., 2015. A comprehensive survey of clustering algorithms. Annals of Data Science 2 (2), 165–193. https://doi.org/10.1007/s40745-015-0040-1.
Zhu, J., Zou, H., Rosset, S., Hastie, T., 2009. Multi-class adaboost. Stat. Interface 2 (3), 349–360.