Visual Reconciliation of Alternative Similarity Spaces in Climate Modeling

Category: Research

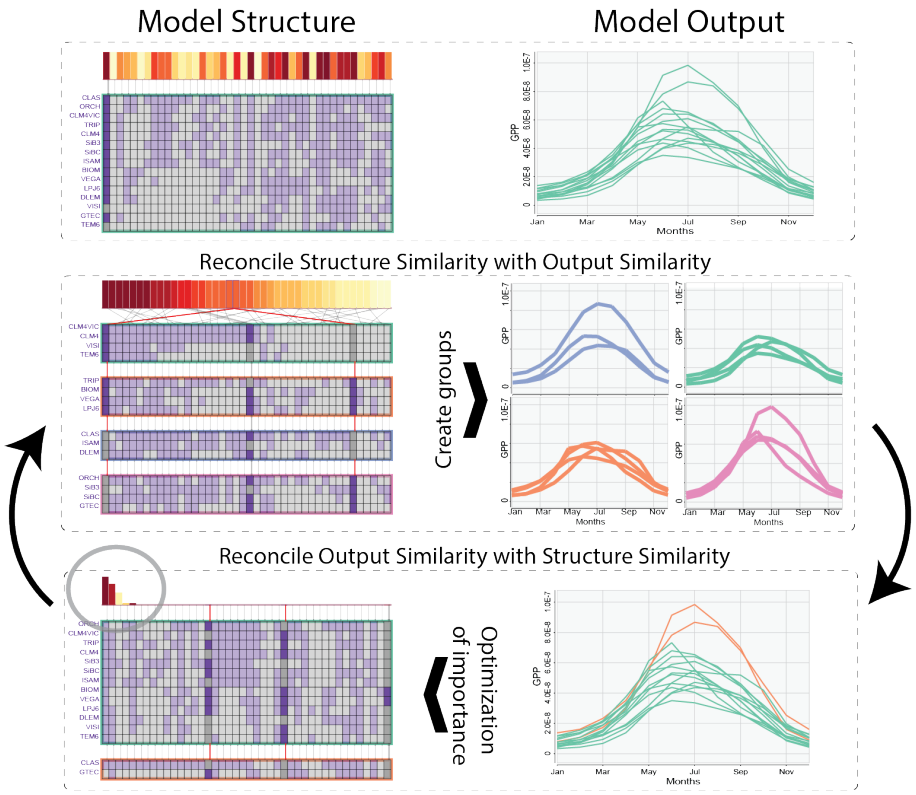


Fig. 1: Iterative visual reconciliation of groupings based on climate model structure and model output. Visual inspection of similarity coupled with an underlying computation model facilitates iterative refinement of the groups and flexible exploration of the importance of the different parameters.

Abstract— Visual data analysis often requires grouping of data objects based on their similarity. In many application domains researchers use algorithms and techniques like clustering and multidimensional scaling to extract groupings from data. While extracting these groups using a single similarity criteria is relatively straightforward, comparing alternative criteria poses additional challenges. In this paper we define visual reconciliation as the problem of reconciling multiple alternative similarity spaces through visualization and interaction. We derive this problem from our work on model comparison in climate science where climate modelers are faced with the challenge of making sense of alternative ways to describe their models: one through the output they generate, another through the large set of properties that describe them. Ideally, they want to understand whether groups of models with similar spatio-temporal behaviors share similar sets of criteria or, conversely, whether similar criteria lead to similar behaviors. We propose a visual analytics solution based on linked views, that addresses this problem by allowing the user to dynamically create, modify and observe the interaction among groupings, thereby making the potential explanations apparent. We present cases studies that demonstrate the usefulness of our technique in the area of climate science.

1 Introduction

Grouping of data objects based on similarity criteria is a common analysis task. In different application domains, computational methods such as clustering, dimensionality reduction, are used for extracting groupings from data. However, in the real world, with the growing variety of collected and available data, group characterization is no longer restricted to a single set of criteria; it usually involves alternative sets. Exploring the inter-relationship between groups defined by such alternative similarity criteria is a challenging problem. For example, in health care, an emerging area of research is to reconcile patient groups based on their demographics and based on their disease history, for targeted drug development [34]. In climate science, an open problem is to analyze how similar outputs from model simulations can be linked with similarity in the model structures, characterized by diverse sets of criteria. Analyzing features of model structures and their impact on model output, can throw light into important global climate change indicators [19].

To meet these challenges, the data mining area has developed redescription algorithms for quantifying and exploring relationships among multiple data descriptors [23]. These techniques have focused on mining algorithms for binary data, where objects are characterized by the presence or absence of certain features. Group extraction based on such computational methods are generally heavily influenced by

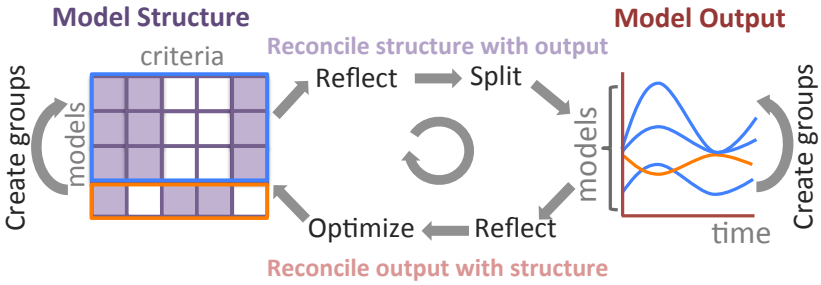


Fig. 2: **Conceptual model of visual reconciliation** between binary model structure data and time-varying model output data. Iterative creation of groups and derivations of relationship between output similarity and importance of the different model structure criteria. Blue and orange indicate different groups of models.

parameter selections. Also, it usually takes multiple iterations to find a perfect solution; and in most cases, only approximate solutions can be found. Domain experts need to be involved in this iterative process, utilizing their expertise for controlling the parameters. This necessitates a visual analytics approach towards group extraction and reconciliation of such groups created based on alternative similarity criteria. Currently, there is a lack of visual analytics techniques that can handle the complexity of the reconciliation process, especially involving domain experts.

To fill this gap, we introduce a novel visual analytics paradigm: *visual reconciliation*, which is an iterative, human-in-the-loop process for reconciling alternative similarity spaces. The reconciliation technique involves a synergy of computational methods, adaptive visual representations, and a flexible interaction model, for communicating the relationships among the similarity spaces. While the technique is domain-independent, and generally applicable to different similarity spaces, we use climate science as a specific use case for illustrating the benefits of our technique.

Our concept of visual reconciliation is grounded in our experience of collaborating with climate scientists as part of the Multi-Scale Synthesis and Terrestrial Model Inter-comparison Project (MsTMIP). An open problem in climate science research is how to analyze the effect that similarity and differences in climate model structures have on the temporal variance in model outputs. Recent research has shown model structures can have significant impact on variability of outputs [14], and that, some of these findings need to be further investigated in details for exploring different hypotheses.

To achieve these goals, we propose an analysis paradigm for reconciling alternative similarity spaces, that leverages the high bandwidth of human perception system and exploits the pattern detection and optimization capabilities of computing models [2, 16] The key contributions of this work stems from a visual reconciliation technique (Figure 2) that i) helps climate scientists understand the dependency between alternative similarity spaces for climate models, ii) facilitates iterative refinement of groups with the help of a feedback loop, and iii) allows flexible multi-way interaction and exploration of the parameter space for drilling down into patterns of interest.

2 Motivation

Why do we need to define a new visual analytics technique? Reconciling alternative similarity spaces is challenging on several counts: i) data descriptors can comprise of different attribute types. From a human cognition point-of-view, reconciling the similarity of climate models across two different visual representations is challenging. There needs to be explicit encoding of similarity [10] that helps in efficient visual comparison and preserve the mental model about similarity. Moreover adaptation of similarity needs to be reflected by dynamic linking between views and by preventing change blindness. ii) for aligning two different similarity spaces, say computed by two clustering algorithms, we will in most cases get an approximate result. The result will need to be iterated upon with subsequent parameter tuning to achieve higher accuracy. This necessitates iteration, and therefore a human-in-the-loop approach. iii) domain experts want to *trust* the methodology working at the backend and the flexibility to tune parameters and understand their interaction. Fully automated methods do not allow that. Thereby, a user-driven approach is necessary where parameters in similarity computation can be influenced by user selections and filters. In this section we provide context to the visual reconciliation technique by discussing the background with respect to climate models. As mentioned before, the technique is not restricted to climate models, but for simplifying our discussion, in this paper we specifically discuss the applicability of the technique in the climate modeling context.

2.1 Problem Characterization

Climate models, specifically, Terrestrial Biosphere Models (TBM) represent time and space variable ecosystem processes, like, simulations of photosynthesis and respiration, using different algorithms. Our visual reconciliation technique has been developed in the context of structure and output of these TBMs.

Model Structure: A model simulation algorithm can have different implementations of a process. These implementations are different from each other due to the presence or absence of different criteria, that control the specific process. For example, if a model simulates photosynthesis, a group of criteria like `simulating carbon pools`, influence of `soil moisture`, and `stomatal conductance` can be either present or absent. Thus, a model structure is a function of these criteria. If there are c criteria, there can be 2^c combinations of this function. There are 4 different classes of criteria, with each class comprising of criteria which is of the order of 20 to 30 in number.

Model Output: Model simulation outputs are ecosystem variables that help climate scientists predict the rates of carbon dioxide increases and changes in the atmosphere. For example, Gross Primary Productivity (`GPP`) is arguably the most important ecosystem variable, indicating the total amount of energy that is fixed from sunlight, before respiration and decomposition. Climate scientists need to understand patterns of `GPP` in order to predict rates of carbon dioxide increases and changes in atmospheric temperature.

Relationship between model structure and output: One of the open research questions in the TBM domain is how similarity or differences in model output can be correlated with that in model structures. The heterogeneity of model structure and model output data make it complex to derive one-to-one relationships among them. Currently, in absence of an effective analysis technique, scientists manually browse through the theoretically exponential number of model structure combinations, and analyze their output. This process is inefficient and also ineffective owing to the large parameter space which can easily cause important patterns to be missed.

To address this problem we focus on using visual analytics methods for addressing the following high-level analysis questions: i) given all other factors remain constant, analyze how different combination of parameters within model structure cause similarity or difference in model output, and ii) by examining time-varying model outputs at different regions, understand which combination of parameters cause the same clusters or groups in model output.

2.2 Visual reconciliation goals

For addressing the aforementioned challenges, we have devised the iterative visual reconciliation technique comprising of automated computation of similarity functions and coordinated multiple views. As illustrated in Figure 2, the visual reconciliation technique enables climate scientists to i) start analyzing model structure and use that as feedback for reconciling similarity or differences in model output, and ii) start analyzing model output and use that as a feedback for comparing similarity or differences in model structure. The reconciliation framework focuses on three key goals which are as follows:

Similarity encoding and linking: For providing guidance on choosing the starting points of analysis, the visual representations of both structure and output encode similarity functions. Subsequently, scientists can use those initial seed points for reconciling structure characteristics with output data, or conversely, for reconciling output data with structure characteristics.

Flexible exploration of parameters: The visual feedback and interaction model adapts to the analysts' workflow. Scientists can choose different combinations of parameters, customize clusters on model structure and model output side and accordingly the visual representations change, different indicators of similarity are highlighted.

Iterative refinement of groups: By incorporating user feedback in conjunction with a computation model, the reconciliation technique allows users to explore different group parameters in both data spaces and iteratively refine the groupings. The key goal here is to understand, which criteria in model structures are most important in determining how the outputs are similar or different over time.

3 Related Work

We discuss the related work in the context of the following threads of research: i) automated clustering methods proposed in the data mining community for handling different data descriptors, ii) integration of user feedback for handling distance functions in high-dimensional data, and iii) visual analytics solutions for climate modeling.

3.1 Clustering Methods

Different clustering methods have been proposed in the data mining community have been proposed for dealing with alternative similarity spaces. Pfitzner et al. proposed a theoretical framework for evaluating the quality of clusterings through pairwise estimation of similarity ([24]). The area of multi-view clustering [3] analyzes cases when data can be split into two independent subsets. In that case either subset is conditionally independent of each other and can be used for learning. Similarly, authors have proposed approaches towards combining multiple clustering results into one clustered output, suing similarity graphs [20] . Although we are also dealing with multiple similarity functions, the goal is to reconcile one with respect to the other.

In this sense, the most relevant research in data mining community is that which looks into learning the relationship between different data descriptor sets. The reconciliation idea is similar, in principle, to redescription mining which looks at binary feature spaces and uses automated algorithms for reconciling those spaces. [25, 23]. While redescriptions mostly deal with binary data, we handle both binary data and time-varying data in our technique.

Our work is also inspired by the consensus clustering concept, which attempts to find the consensus among multiple clustering algorithms [21] in the context of gene expression data. Consensus clustering has also been applied in other applications in biology and chemistry [8, 6]. In our case, while we are interested in the consensus between similarity of model structure and model output, we also aim at quantifying and communicating the contribution of the different parameters towards that consensus or the lack thereof. We adopt a human-in-the-loop approach, which is especially crucial when domain scientists are involved. The automated methods do not provide adequate transparency to the clustering parameters, and also in most cases, iteration is necessary to accurately quantify the reconciliation results. Our visual reconciliation technique allows domain experts to supervise the iterative process of tuning parameters and visualizing the dependency between the similarity spaces.

3.2 User Feedback for Adaptive Distance Functions

Recently, there has been a lot of interest in the visual analytics community for investigating how computation and tuning of distance functions can be steered by user interaction and feedback. Recently Gleicher proposed a system called Explainers that attempts to alleviate the problem of multidimensional projection, where the axes have no semantics, by providing named axes based on experts' input [9]. Eli et al. present a system that allows an expert to interact directly with a visual representation of the data to define an appropriate distance function, without having to modify different parameters ([4]). In our case. the parameter space is of high interest to the user; therefore we create a visual representation of the parameters, that is the weights of the criteria on the model structure side, and allow direct user interaction with them. Our user feedback mechanism based weighted optimization method is inspired by the work on manipulating distance functions by Hu et al.([12]). However, the interactivity and conceptual implementation is different, since we are working with two different data spaces, without using multidimensional projections. The modification of distance functions have also been used for spatial clustering, where user selected groups are given as input to the algorithm [22]. Our reconciliation method is similar, in principle to this approach, where the system suggests grouping in one data space, based on the same in other space, by a combination of user selection and computation.

3.3 Visual Analytics for Climate Modeling

Similarity analysis of model simulations is an emerging problem in climate science. This is especially relevant for understanding global climate change patterns. While there has been some work for developing visualization solutions for climate data [17], most of these focus on addressing the problem at the level of a single model and understanding its spatio-temporal characteristics. For example, Steed et al. introduced EDEN [29], a tool based on visualizing correlations in an interactive parallel coordinates plot, focused on multivariate analysis. Recently, UVCDAT [32] has been developed which is a provenance-enabled framework for climate data analysis. However, like most other tools, UV-CDAT does not support multi-model analysis. In our case, we are not only comparing multiple models, but also comparing two different data spaces: model structure and model output. Climate scientists have found that different combinations of model structure criteria can potentially throw light into different simulation output behavior [14]. However, to the best of our knowledge, no visual analytics solution currently exists in climate science to address this problem. For developing a solution, formulating an analysis paradigm precedes tool development because of the complexities involved in handling multiple descriptor spaces. Although there has been some work on hypothesis generation [15] and task characterization [27] for climate science, they are not sufficient for reconciling the alternative similarity spaces. In this sense we advance the state-of-the-art in problem characterization in the climate science domain by introducing the visual reconciliation paradigm.

4 Coordinated Multiple Views

An important component of the visual reconciliation technique is the interaction between multiple views [26]. In this case we have binary model structure data and time-varying model output data. As we had shown in Figure 2, the goal is to let domain scientists create and visualize groups on both sides, and understand the importance of the different criteria in creating those groups. In this section we provide an overview of the different views and describe the basic interactions between those.

Matrix View: To display the model structure data, which is a two-dimensional matrix of 0's and 1's, we use a color-coded matrix Figure 3, which serves as a presence/absence representation of the different criteria for the model structure. This is inspired from Bertin's reorderable matrix [1] and the subsequent interactive versions of the matrix [28].

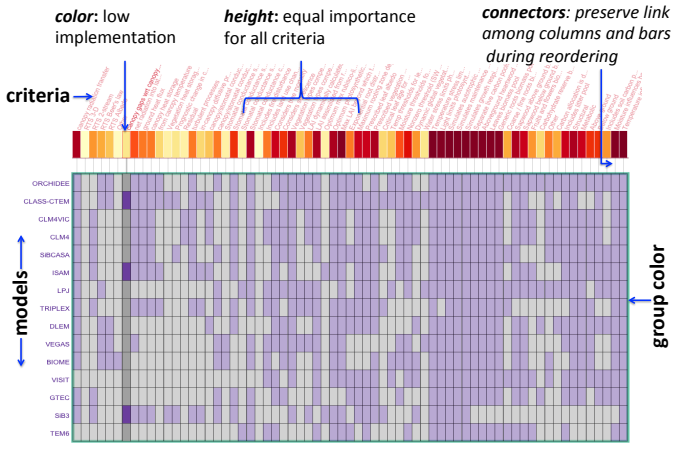


Fig. 3: **Matrix view for model structure data**: Rows represent models and columns represent criteria. The variation of average implementation of a criterion for all models is shown by a color gradient from light yellow to red, with red signifying higher implementation. In the default view, all criteria have equal importance or weights, indicated by the heights of the bars. Connectors help visually link the columns and bars when they are reordered independently.

Purple color is used for denoting presence and gray for absence. Visual salience of a heat map depends on the order of the rows and columns and numerous techniques have been developed till data fore reordering [5, 33] and seriation [18]. In this case, the main motivation is to let the scientists visually separate the criteria which have high average the non-implementation (indicated by 0) and those with high average implementation. So, for providing visual cues on potential groups within the data, we reorder the rows and columns, based on a function that puts the criteria that are present to the upper left of the matrix and pushes those are absent, to the bottom right.

The colored bars on top of the matrix serve a dual purpose. The heights of the bars indicate the importance or weight of each criteria for creating groups in model structure. The colors of the bars, with a light yellow to red gradient indicate the average implementation of a criterion. For example, as indicated in Figure 3, the yellow bar indicates that only three models have implemented that criterion. This gives a quick overview of which criteria are most implemented, and which ones, the least. The grey lines connecting the bars and the columns are essential for connecting the link when columns are reordered. The criteria bars and the columns in the matrix can be reordered independently.

Groups can be created by selecting the different criteria. For a single criteria, there can be two groups of models: those which do not implement the criteria and have a value 0, and those which implement criteria, and have a value 1. With multiple selections, there can be 2^c combinations, with c being a criteria. In most cases practically, only a subset of those combinations exist. These are highlighted in the matrix view.

Time Series View: We display the model output data, which comprises of a time series for each model, we use a line chart comprising of multiple time series (Figure 4,a). But effective visual comparison of similarity among multiple groups is difficult using this view because of two reasons. First, due to similar trajectory of the series, there is a a lot of overlap, leading to clutter. Second, we are unable to show the degree of clustering using this approach. To resolve these design problems, we use small multiples. Small multiples [31] have been used extensively in visualization, one problem with them is when there are a large number of them, it becomes difficult to group them visually without any additional cues. To prevent this, we create a small multiple for each group. When there are time series for different region, a small multiple can also be created for each region to compare groupings across different regions.

Interaction: An overview of the steps in the interactive workflows between the matrix view and the time series view are shown in Figure 2. These actions and operations are described below:

Create Groups: While reconciling model structure with model output, scientists can first observe similarity among the models based on their criteria, and accordingly create groups. This is part of the reconciliation workflow described in Section 5.1. In the matrix view, groups can be created on interaction. In the time-series view, groups are either suggested by the system or selected by the user through direct manipulation. This is part of the reconciliation workflow described in Section 5.2.

Reflect: Creation of groups triggers reflection of the groups in both views. On the matrix side, this is through grouping of the rows. On the time series side, this is done by color coding the lines.

Split: In the time series view, groups can be reflected by clustering the models into small multiples.

Optimize: While reconciling model output with structure, to handle the variable importance of the criteria, an *optimization* step is necessary. This workflow starts with the scientist selecting groups in the output, which get *reflected* in the matrix view. Next they can choose to *optimize* the importance or the weights, which leads to subsequent iteration. This reconciliation workflow is described in detail in Section 5.2.

5 Reconciliation Workflows

In this section we describe how we instantiate the conceptual model of visual reconciliation described in Figure 2 by incorporating the coordinated multiple views, user interaction and an under lying computational model. The following workflows provide a step-by-step analysis of how the views and interactions can be leveraged by climate scientists for getting insight into structure similarity and output similarity.

5.1 Reconcile Structure Similarity with Output Similarity

In Figure 4 we show the different steps in the workflow when the starting point of analysis is the model structure. This workflow relies on visual inspection of structure similarity by using matrix manipulation, and observing the corresponding patterns in output by creation of small multiples. The steps are described as follows:

Create groups: For reconciling model structure with output, it is necessary to first provide visual cues about which models are more similar with respect to the different criteria. For this the default layout of the matrix is sorted from left to right, by high to low average implementation of the different criteria. This is indicated in Figure 4b by the transition of the importance bars from red to yellow. This gives the scientists an idea of which criteria create more evenly sized groups with 0's and 1's. The criteria which are colored dark red and light yellow will create groups which are skewed: either too many models implement the criteria or they do not. Selecting criteria from the criteria which are deep yellow and orange, gives more balanced clusters, with around 50 per cent implementation. The highlighted column indicates the criterion with the highest percentage of implementation.

The selected columns are indicated in Figure 4c. These two criteria creates four groups. For showing groups of models within the matrix, we introduce vertical gaps between groups, and then draw colored borders around each group. Reordering by columns is also allowed for each group independently as shown in Figure 4c . In that case, the weighted ordering of the bars is kept fixed. For visually indicating the change in ordering we link the criteria by lines. Lines that are parallel indicate that those criteria have not moved due to reordering and share the same position for different groups. Since too many crossing lines can cause clutter, we render the lines with varying opacity. For indicating movement of criteria, we render those lines with higher opacity. To highlight where a certain criteria is within a group, on selection we highlight the line by coloring it red as shown in the figure.

If columns in each group is reordered independently, that shows the average implementation patterns for each group clearly. But it becomes difficult to compare the implementations of a set of criteria across the different groups. To enable this comparison, user can select a specific group which will be reordered column-wise, and the columns in other groups will be sorted by that order. This is shown

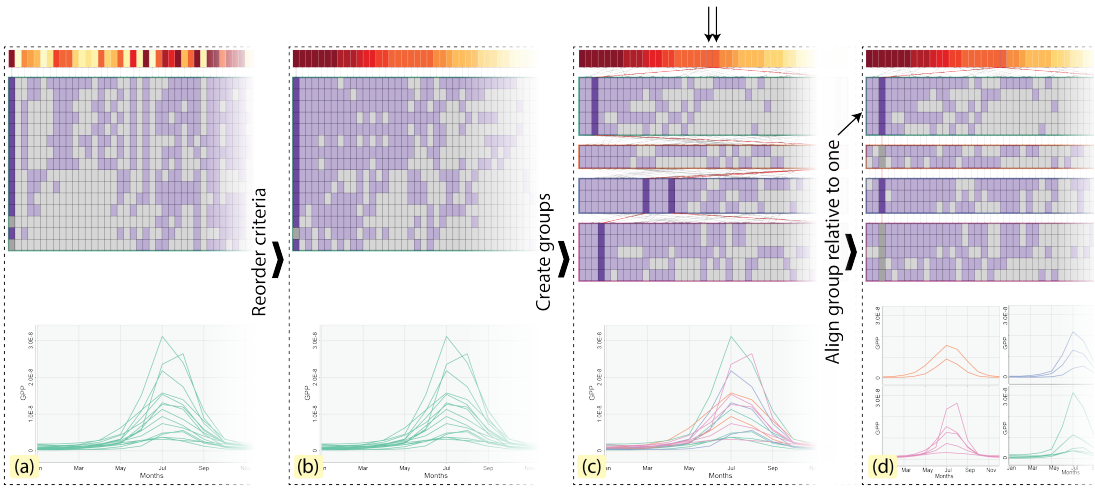


Fig. 4: **Workflow for reconciling model structure with model output**: This linear workflow relies on matrix manipulation techniques and visual inspection of grouping patterns in the matrix view and the small multiple view.

in Figure 4d, where the first group from the top is reordered based on the columns, and other groups are aligned relative to that group. As observed, this enables more efficient comparison relative all the implemented and non-implemented criteria in the first group. For example, we can easily find that the rightmost criteria is not implemented by the first group of models, but is implemented by all other groups.

Reflect: The creation of groups in the structure is reflected in the output by the color of the groups. Users can see the names of the models on interaction.

Split: Small multiples can be created for each group (Figure 4d). The range of variability of models in each small multiple group reflects how similar or different they are. This is comparison is difficult to achieve in a time series overloaded with too many lines. This also enables a direct reconciliation of the quality of grouping in model structure with that of the output. For example, as shown in the figure, only the orange group has low variability across models, denoting that the groups based on the criteria in model structure do not create groups where models produce similar output behavior.

5.2 Reconcile Output Similarity with Structure Similarity

To reconcile output with structure and complete the loop, we need to account for the fact that different criteria can have different weights or importance in the creation of groups. One of the goals of the reconciliation models is to enable scientists explore different combinations of these criteria that can create groups that are similar to the corresponding model output. However, naive visual inspection is inefficient to analyze all possible combinations without any guidance from the system. For this, we developed a weighted optimization algorithm that complements the human interaction. We describe the algorithm, provide an outline of its validation, and the corresponding workflow, as follows.

5.2.1 Weighted Optimization

Using the model structure data and the model output data, we can create two distance matrices. The eventual goal is to learn a similarity function from the output distance matrix and modify the weights of the criteria in the structure distance function for adapting to the output similarity matrix. We describe the problem formulation below.

Let $\hat{M}$ be a matrix representing the model output with size $n \times p$ and $\tilde{M}$ represents the model structure with size $n \times q$. Similarity in model output is computed by the function $\hat{d} : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$. This function can be any specialized distance function such as Euclidean, Cosine, etc. For the model structure we use weighted euclidean distance $\tilde{d}^w : \mathbb{R}^q \times \mathbb{R}^q \rightarrow \mathbb{R} = \sum_{k=1}^{q} \sqrt{w^k (y_i^k - y_j^k)^2}$, where w^k is a weight assigned to each dimension on $\tilde{M}$.

Using $\hat{d}$ we encode the similarity information of the model output in a distance matrix $\hat{D}$. Our goal would be to find the weights' vector $\mathbf{w} = \{w^1, ..., w^q\}$ which could create a distance matrix for the model structure $\tilde{D}$ containing approximately the same similarity information as the model output. This problem can be formulated as the minimization of the square error of the two distance functions:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \sum_{i=1}^{n} \sum_{j=1}^{n} \|\tilde{d}^w(x_i, x_j)^2 - \hat{d}(y_i, y_j)^2\|^2 \\ \text{subject to} \quad & w_k \geq 0,\ k = 1, \ldots, q. \end{aligned} \tag{1}$$

where $\|.\|$ is the L_2 norm.

Using this vector $\mathbf{w}$ we can define which criteria are important in the model structure to recreate the same similarity information from the model output. Note that in the previous formulation we have not taken into account the the user's feedback. The weights computation step is similar to the one used in weighted metric multidimensional scaling [11] technique.

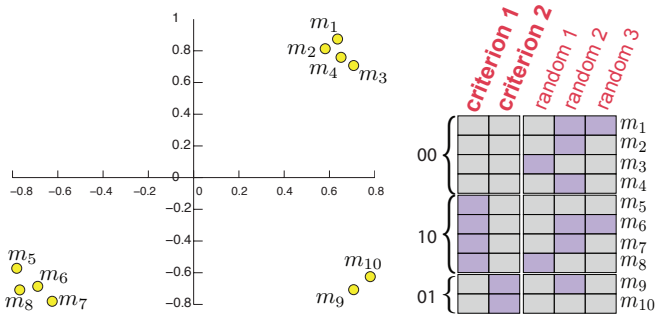
If we want to incorporate user's feedback into our formulation we can multiply the square errors in Eq. 1 by a coefficient $r_{i,j}$. This number represents the importance of each pair of elements in the minimization problem. In our approach we allow the user to define groups on the model output, then $r_{i,j}$ will be almost zero or zero for all the elements i, j in a group. Now, we need to minimize:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \sum_{i=1}^{n} \sum_{j=1}^{n} r_{i,j} \|\tilde{d}^w(x_i, x_j)^2 - \hat{d}(y_i, y_j)^2\|^2 \\ \text{subject to} \quad & w_k \geq 0,\ k = 1, \ldots, q. \end{aligned} \tag{2}$$

Both equations above can be converted into quadratic problems and solved using any quadratic programming solvers, such as JOptimizer [30] for Java or *quadprog* in MATLAB.

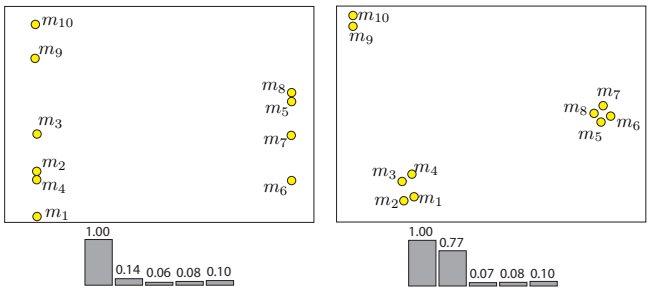
Our approach of incorporating user feedback for computation of the weights is similar to the cognitive feedback model, namely V2PI-MDS [13]. Mathematically the approaches are similar but conceptually they are different on two counts. First, in their case the projected data space is another representation of the high-dimensional data space and they attempt to reconcile the two. In our case however, the underlying data spaces are entirely different. We handle this problem by using interactive visualization as a means to preserve the mental model of the scientists about the characteristics of the data. We could also have used multidimensional projections. But as found in previous work, domain scientists tend not to trust the information loss caused by the dimensionality reduction and prefer transparent visualizations, where the raw data is represented.[7].

Second, the user interaction mechanism for providing feedback to the computation model is also different than the V2PI-MDS model. We allow users to define groups within the data, as opposed to direct



(a) Model Output (b) Model Structure

Fig. 5: **Synthetic data for validating weighted optimization**. Using the model output data in (a) and model structure data in (b), we validate the accuracy of the optimization algorithm.



(a) Automatic optimization (b) Optimization based on user feedback

Fig. 6: **Validation of user feedback based optimization.** As we can observe in (b), optimization based on user's feedback gives highest weights to the two criteria which are splitting the models into three groups.

manipulation and movement of data points in a projection; which is not applicable in our case. Our focus is on the relationship between the weights of the dimensions and the similarity they induce. As a result, we let users explore different groupings by using the sorted weights and modifying the views accordingly. This results in a rich interactive analysis for reconciling the two similarity spaces.

5.2.2 Validation

To validate our optimization, we use two synthetic datasets, one for model output and the other one for model structure. The purpose of this validation to demonstrate the accuracy of the algorithm in the best case scenario, i.e., when a perfect grouping based on some criteria exists in the data. In most real-world cases, however the optimization will only create an approximation of the input groups.

Our model output is a two-dimensional dataset and we use scatter plot to visualize it (Figure 6). We can notice that we have three well defined groups $\{m_1,m_1,m_3,m_4\}$, $\{m_5,m_6,m_7,m_8\}$ and $\{m_9,m_{10}\}$. Figure 5(b) shows our synthetic model structure data which contains boolean values. Each row represents a different model (m_i) and each column a different criteria. The first two criteria were chosen specifically to split the dataset into the same three groups as the model output. For instance when $criteria_1 = 0$ and $criteria_2 = 0$ we can create the group $\{m_1,m_1,m_3,m_4\}$.The next three columns are random values (zero or one).

First, we solve the Eq. 1 using our synthetic dataset and Euclidean distance for the model output; and we get $\mathbf{w} = \{1.00,0.14,0.06,0.08,0.10\}$. For visualizations purpose we use the classical multidimensional scaling algorithm to project the model structure data using the Weighted Euclidean distance. We normalized the weights between zero and one for visualization purpose, but the weighted Euclidean distance uses the unnormalized weights. Figure 6(a) shows the two dimensional data. Our vector $\mathbf{w}$ was able to capture some similarity information from the model output. For example, $\{m_1,m_1,m_3,m_4\}$ is a well defined group. Even though $\{m_5,m_6,m_7,m_8\}$ and $\{m_9,m_{10}\}$ are not mixed, they are not well defined groups.

Next, we incorporate user feedback and set the coefficient $r_{i,j}$ to zero for all pair combinations in the groups $\{m_1,m_1,m_3,m_4\}$, $\{m_5,m_6,m_7,m_8\}$ and $\{m_9,m_{10}\}$. Solving Eq. 2 we get the vector $\mathbf{w} = \{1.00,0.77,0.07,0.08,0.10\}$. Figure 6(b) shows the two-dimensional projection of the model structure using the weighted Euclidean distance and $\mathbf{w}$. We notice that now the three groups are well defined. Our algorithm gave the highest weights to the first two criteria (criteria1= 1.0 and criteria2= 0.7) which we knew a priori have the best combination to split the model structure in the same groups as the model output.

These two experiments show that our formulation accurately gives the highest weights to the most relevant criteria for splitting models into groups, and this will be used to guide the user during the exploration process. In Section 6 we will show how this approach works with real data; where in most cases, an approximation of the output group is produced by the algorithm.

5.2.3 Workflow

In Figure 7 we show how the complete loop starting from output to structure, and back, is executed by user interaction and the optimization algorithm described above. This workflow relies on human inspection of structure similarity through manipulation of the matrix view and observation of the corresponding output in the small multiples of time series. The steps are described as follows:

Create groups in output: For suggesting groups of similar outputs, the system uses clustering of time series by Euclidean distance or correlation (Figure 7a). While other metrics are available for clustering time series, for this case scientists were only interested in these two. Accordingly, the clusters are updated in the output view.

Reflect in structure: These clusters are reflected in the model structure side by reordering the matrix based on the groups (Figure 7b). All the criteria are given equal weights by default, as indicated by the uniform height of the bars. The two views are linked by the color of the groups. Users can also select groups by direct manipulation in the output view.

Optimize weights: Next on observing the system-defined clusters, one can choose to optimize the weights for the criteria on the structure side. As shown in Figure 7c, the columns are reordered from left to right based on weights. These weights serve as hints to the user for creating groups on the structure side. The groups are not immediately created to prevent change blindness. The system needs the user to intervene to select the criteria, based on which the groups can be created.

The underlying optimization algorithm as described earlier creates an approximate grouping based on the input. In many cases, as shown in the figure, the highest weight may not give a perfect grouping. By perfect grouping we mean, the optimization algorithm is able to create the exact same groups as the input from the output side. In most cases, the weights for an exact solution might not even exist. By using the optimization, all we get is a group of structure clusters which are as closely aligned with the output as possible.

Create groups in structure: Based on the suggested weights, a user can select the two highest weights and create groups, as shown in Figure 7d. There are four possible combinations of these two criteria (with 0's and 1's) and all of them are shown in their own group. In many cases all possible combinations might not exist.

Reflect/Split in output:The creation of the groups are also reflected on the output side by indicating the group membership of each model by color-coding or by creation of small multiples (Figure 7e), the output groups created are not perfect, as they do not exactly match with the output groups in the previous step. From this however, the scientists can judge the effect of the two criteria on model output. For example, if for the selected criteria, the presence or absence does not have an impact on the output, that will be reflected in the time series,

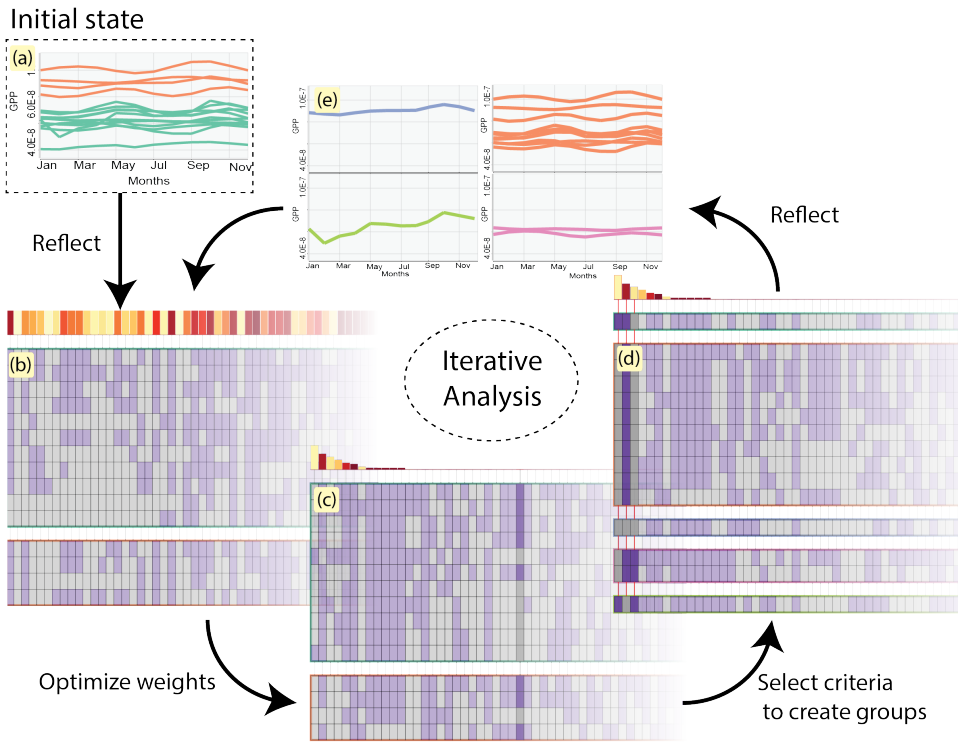


Fig. 7: **Workflow for reconciling output with structure through feedback**: This iterative workflow relies on weighted optimization, based on Equations 1 and 2, and human initiated parameter tuning and selection for reconciling model output with model structure.

by their spread or lack of any significant correlation. For inspecting if combining other criteria can give a more perfect grouping on the structure side, that matches with the output, scientists need to continue the iteration and repeat the previous steps.

6 Case Study

We collaborated with 2 climate scientists from the Oak Ridge National Lab and one other climate scientist from the United States Forest Service, as part of the Multi-Scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP). Each of them have at least ten years of experience in climate modeling and model inter-comparison. MsTMIP is a formal multi-scale synthesis, with prescribed environmental and meteorological drivers shared among model teams, and simulations standardized to facilitate comparison with other model results and observations through an integrated evaluation framework [14]. One key goal of MsTMIP is to understand the sources and sinks of the greenhouse gas carbon dioxide, the evolution of those fluxes with time, and their interaction with climate change. To accomplish these goals, inter-annual and seasonal variability of models need to be examined using multiple time-series. Early results from MsTMIP have shown that variation in model outputs could be traced to the same in model structure. Using visual reconciliation, climate scientists wanted to further understand whether similarity or differences in model structure play a role in the inter-annual variability of Gross Primary Productivity (`GPP`) for different regions. Inclusion of particular combinations of simulated processes may exaggerate `GPP` or its timing more than any component in isolation. Inclusion of a patently incorrect model structure could dramatically sour model output by itself

We describe two cases where our collaborators could find relationships between model structure and model output using our visual reconciliation technique. The model structure data is segmented into different classes like energy, carbon, vegetation, etc. In this case the scientists wanted to understand the relationship between criteria belonging to energy and vegetation, and their `GPP` variability in `Polar` and `North American Temperate` regions. Each of the model structure datasets consist of about 15 models and about 20 to 30 criteria.

6.1 Reconciling seasonal cycle similarity with structural similarity

The seasonal cycle of a climate model is given by the trajectory of the time series and the peaks and crests for the different months in a year. Exploring the impact of seasonal cycles for different models with respect to `GPP` is an important goal in climate science, since the amount and timing of energy fixation provides a baseline for almost all other ecosystem functions, and models must accurately capture this behavior for all regions and conditions before other, more subtle ecosystem processes can be accurately modeled. The motivation for this scenario was to find if there is any dependency between regional seasonal cycles of models and included model structures with respect to this overarching energy criterion.

The scientists started their analysis in the `Polar` region by selecting the `BIOME` and `VEGAS` models which appeared to be similar with respect to both their `GPP` values and the timing of their seasonal cycles, as shown in Figure 8a. Their intent was to observe which energy parameter causes `BIOME` and `VEGAS` to behave similarly in one group, and the rest in another. They optimized the matrix view to find the most important criterion, which was found to be `Stomatal conductance`. After this step they chose to select this criteria to split the models into two groups, shown in Figure 8b and reflected in Figure 8c. The underlying optimization algorithm thus gave a perfect grouping, with the models that implement Stomatal conductance in the orange group, while the rest are in another group. The climate scientists were already able to infer that `Stomatal conductance` has strong impact on the seasonal cycles of `BIOME` and `VEGAS`.

Next the scientists selected the `SiB3` and `SiBCASA` models in the North American Temperate (`NAT`) region, which appear to be similar with respect to their seasonal cycle and `GPP` output (Figure 8d). This grouping is already intuitive and inspires confidence, because of its consistency with the known genealogical relationship of these two models as siblings. With the same goal as the previous case, they optimized the matrix view, and found that `Prognostic change` was the most important structural criterion to approximately create the two groups. This structural criterion provided a near-perfect seg-

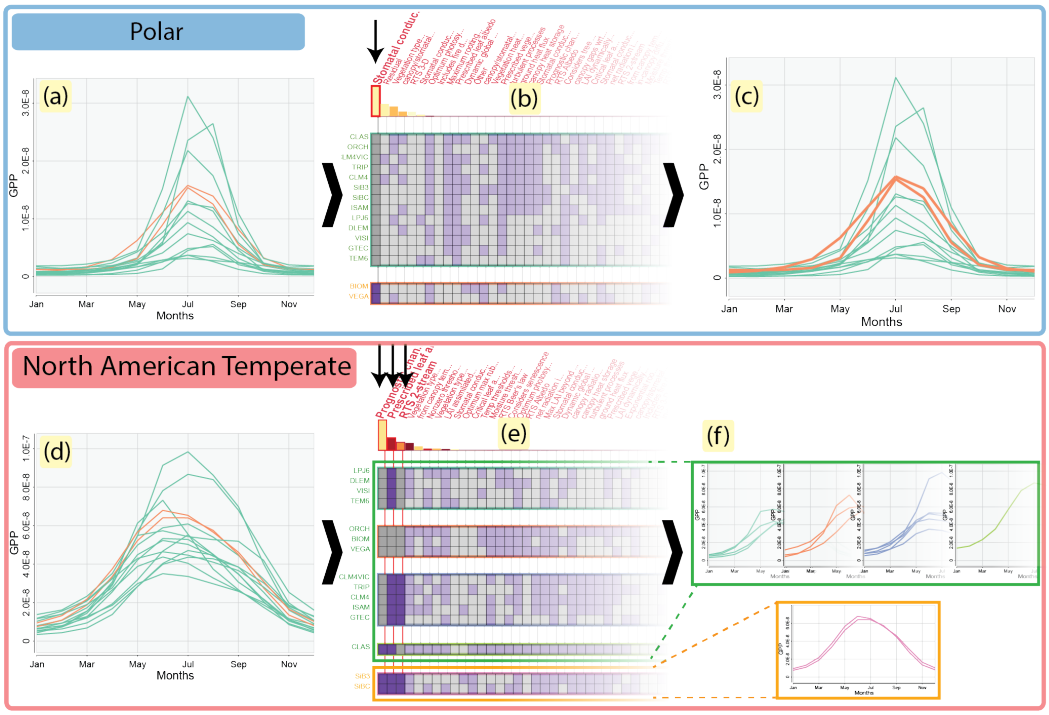


Fig. 8: **Reconciling seasonal cycle with model structure** using the workflow described in Section 5.1. (a) Initial user selection in Polar region output. (b) Weighted optimization, (c) Corresponding output; (d) Initial user selection in North American Temperate region, (e) Creating groups based on the first three criteria after optimization. (f) Small multiple groups of models.

mentation, except for the `CLASS-CTEM` model, which also implements this parameter, as shown in Figure 8e. In an attempt to get the exact segmentation, they selected the next two most important criteria, which are `prescribed leaf index` and `RTS2-stream`. `SiB3` and `SiBCASA` implement both of these criteria and are in one output group, while the other green output group is split into three sub-groups based on their implementation of these three criteria. The implementation of these three criteria thus has a significant effect on the grouping of these two models with respect to their `GPP`. The scientists could continue in this way to find more inferences from the implementation or non-implementation of these three structural criteria, by further observing their output in small multiples, as shown in Figure 8f. This shows that the blue group, none of which implement `Prognostic change`, but all of which implemented the other two, show a greater spread of `GPP` output values than any other group. In this way, the scientists could reconcile the impact of different energy criteria on the seasonal cycle and regional variability of `GPP`.

6.2 Iterative exploration of structure-output dependency

In this case, the scientists started by looking at the model structure data for discovering structure criteria that could explain model groups having high and low `GPP` values across both `Polar` and `NAT` regions. A simple sequential search for criteria is inefficient for reconciliation. To start their analysis, as shown in Figure 9a, the matrix view is first sorted from left to right by the columns having high numbers of implementations. The sorting enabled the scientists to group using a criterion that would cause balanced clusters, i.e., divide the models into equal groups. In this view, these criteria would lie in the center, having orange or deep yellow color. In course of this exploration, they found that the `canopy/stomatal conductance whole canopy` structural criterion splits the group into nearly equal halves. These clusters are represented in the output by green, i.e., not implementing that criterion, and orange, i.e., implementing that criterion. Further, looking at the output, as shown in Figure 9b, scientists found that the orange group has higher `GPP` values and the green group has lower values. In other words, the models that have implemented `stomatal conductance` have higher `GPP` values than the ones that have not implemented this criterion. This grouping is consistent for the North American Temperate region, with the exception of the `CLASS-CTEM` model, as shown in Figure 9c.

Next, the scientists wanted to verify whether by performing optimization, they can get the same criterion to be the most important for the behavior of `GPP` within the `Polar` region, which represents a different, extreme combination of ecological conditions. They selected the green group, as shown in Figure 9d, and then chose to optimize the matrix view. They found the same criterion (`canopy/stomatal conductance whole canopy`) to have the highest weight, reinforcing the reconciling power of this same group of model structures for explaining differences in `GPP` across two extreme eco-regions. Thus, the same criterion that they discovered interactively could be verified algorithmically. Note that, as shown in Figure 9d, only one of the models is classified in a different group than the user-selected group.

For the `NAT` region, the scientists wanted to drill-down to determine what was causing `Class-CTEM` to behave differently, as was found during the initial exploration. They defined two groups, with one of them only having `Class-CTEM` as shown in Figure 9,g. Once they chose to optimize the matrix, they found that no single criterion could produce the same output groups. However, by combining the two most important criteria, that is `vegetation heat` and `canopy-stomatal sunlit shaded` (Figure 9,h), `Class-CTEM` was put in a separate group by itself. It was the only model that implemented both of these criteria. Additionally, the scientists also saw that the models in the green group, which did not implement any of these structures, had a larger range of `GPP` variability than the other model groups (Figure 9,i). They concluded that,by allowing both `more- and less-productive sunlit` and `shaded canopy leaves`, respectively, models which implement these differential processes seem to stabilize the production of `GPP`, even across extremely different eco-regions, possibly accurately reflecting the actual effect of these processes in nature.

7 Conclusion and Future Work

We have presented a novel visual reconciliation technique, using which climate scientists can understand the dependency relationships between model structure similarity and model output similarity.

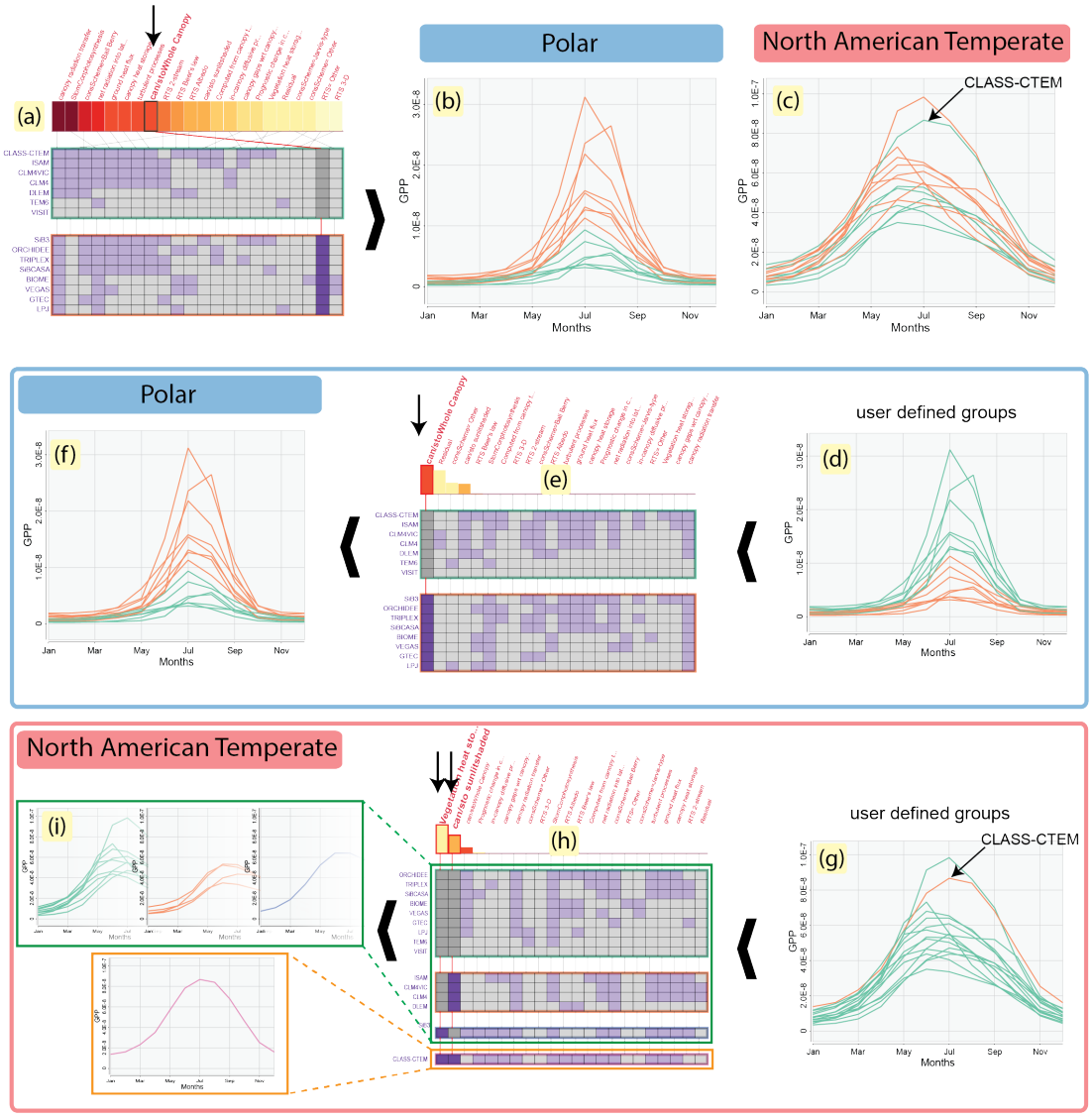


Fig. 9: **Iterative exploration of structure-output dependency** using a combination of the two workflows for reconciliation. (a) Initial user creation of groups, (b,c) Corresponding groups in regions, (d,e,f) Workflow for verifying user-defined groups, (g,h,i) Workflow for finding the criteria that can potentially cause CLASS-CTEM to be an outlier, and then looking at range of variability in small multiple outputs.

Impact: By exploiting visual linking and user-steered optimization, we are able to communicate to the scientists, the effects of different groups of criteria on the variability of model output. Using this technique, scientists could form and explore hypotheses about reconciling the two different similarity spaces, which was not possible before. One of the climate scientists observed that: "One of the most valuable functions of the technique is to effectively remove from consideration the complications created from model structures, that have little to no effect on outputs, and to effortlessly show and rank the differential effects on output created by seemingly related or unrelated model structures."

Challenges: There are several open issues to consider for improving the technique. Currently we are handling only two descriptor sets. More diverse descriptor data will cause visual complexity and it poses a significant challenge for effective visualization design and interaction. Although we are using about 15 models and not more than 30 criteria, we do not foresee a scalability problem, as matrix visualizations do not require much screen real estate. Some of the interactions like showing groups, however, have to be adapted accordingly, for example, by using focus-and-context techniques for zooming on one group and rendering other ones as context with lower resolution.

Generalization: As observed before, the visual reconciliation technique is not restricted to the climate science domain. As a next step, we will apply this technique in the healthcare domain, where the goal is to reconcile patient similarity with drug similarity for personalized medicine development [34]. Another potential application is in the product design domain. For example in the automotive market, car models can be qualified by multitude of features. It will be of interest to automotive companies to reconcile similarity of car models based on their descriptors, with the similarity based on transaction data. In short, we posit that visual reconciliation can potentially serve as an important analytics paradigm for making sense of the ever-growing variety of available data and their diverse similarity criteria.

REFERENCES

[1] J. Bertin. Semiology of graphics: diagrams, networks, maps. 1983.

[2] E. Bertini and D. Lalanne. Surveying the complementary role of automatic data analysis and visualization in knowledge discovery. In *Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration*, pages 12–20. ACM, 2009.

[3] S. Bickel and T. Scheffer. Multi-view clustering. In *ICDM*, volume 4, pages 19–26, 2004.

[4] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-function: Learning distance functions interactively. In *IEEE COnference on Visual Analytics*

Science and Technology, pages 83–92, 2012.

[5] C.-H. Chen, H.-G. Hwu, W.-J. Jang, C.-H. Kao, Y.-J. Tien, S. Tzeng, and H.-M. Wu. Matrix visualization and information mining. In *Proceedings in Computational Statistics*, pages 85–100. Springer, 2004.

[6] C.-W. Chu, J. D. Holliday, and P. Willett. Combining multiple classifications of chemical structures using consensus clustering. *Bioorganic & medicinal chemistry*, 20(18):5366–5371, 2012.

[7] J. Chuang, D. Ramage, C. Manning, and J. Heer. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 443–452. ACM, 2012.

[8] V. Filkov and S. Skiena. Heterogeneous data integration with the consensus clustering formalism. In *Data Integration in the Life Sciences*, pages 110–123. Springer, 2004.

[9] M. Gleicher. Explainers: Expert explorations with crafted projections. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2042–2051, 2013.

[10] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[11] M. Greenacre. Weighted metric multidimensional scaling. In H.-H. Bock, W. Gaul, M. Vichi, P. Arabie, D. Baier, F. Critchley, R. Decker, E. Diday, M. Greenacre, C. Lauro, J. Meulman, P. Monari, S. Nishisato, N. Ohsumi, O. Opitz, G. Ritter, M. Schader, C. Weihs, M. Vichi, P. Monari, S. Mignani, and A. Montanari, editors, *New Developments in Classification and Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 141–149. Springer Berlin Heidelberg, 2005.

[12] X. Hu, L. Bradel, D. Maiti, L. House, and C. North. Semantics of directly manipulating spatializations. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2052–2059, 2013.

[13] X. Hu, L. Bradel, D. Maiti, L. House, C. North, and S. Leman. Semantics of directly manipulating spatializations. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2052–2059, 2013.

[14] D. N. Huntzinger, C. Schwalm, A. M. Michalak, K. Schaefer, et al. The north american carbon program multi-scale synthesis and terrestrial model intercomparison project - part 1: Overview and experimental design. *Geoscientific Model Development Discussions*, 6(3):3977–4008, 2013.

[15] J. Kehrer, F. Ladstadter, P. Muigg, H. Doleisch, A. Steiner, and H. Hauser. Hypothesis generation in climate research with interactive visual data exploration. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1579–1586, 2008.

[16] D. A. Keim, F. Mansmann, and J. Thomas. Visual analytics: how much visualization and how much analytics? *ACM SIGKDD Explorations Newsletter*, 11(2):5–8, 2010.

[17] F. Ladstädter, A. K. Steiner, B. C. Lackner, B. Pirscher, G. Kirchengast, J. Kehrer, H. Hauser, P. Muigg, and H. Doleisch. Exploration of climate data using interactive visualization*. *Journal of Atmospheric and Oceanic Technology*, 27(4):667–679, 2010.

[18] I. Liiv. Seriation and matrix reordering methods: An historical overview. *Statistical analysis and data mining*, 3(2):70–91, 2010.

[19] D. Masson and R. Knutti. Climate model genealogy. *Geophysical Research Letters*, 38(8), 2011.

[20] S. Mimaroglu and E. Erdil. Combining multiple clusterings using similarity graph. *Pattern Recognition*, 44(3):694–703, 2011.

[21] S. Monti, P. Tamayo, J. Mesirov, and T. Golub. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine learning*, 52(1-2):91–118, 2003.

[22] E. Packer, P. Bak, M. Nikkila, V. Polishchuk, and H. J. Ship. Visual analytics for spatial clustering: Using a heuristic approach for guided exploration. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2179–2188, 2013.

[23] L. Parida and N. Ramakrishnan. Redescription mining: Structure theory and algorithms. In *AAAI*, volume 5, pages 837–844, 2005.

[24] D. Pfitzner, R. Leibbrandt, and D. Powers. Characterization and evaluation of similarity measures for pairs of clusterings. *Knowledge and Information Systems*, 19(3):361–394, 2009.

[25] N. Ramakrishnan, D. Kumar, B. Mishra, M. Potts, and R. F. Helm. Turning cartwheels: an alternating algorithm for mining redescriptions. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 266–275. ACM, 2004.

[26] J. C. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization*, pages 61–71. IEEE, 2007.

[27] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2366–2375, 2013.

[28] H. Siirtola. Interaction with the reorderable matrix. In *In Proc., Conf. on Information Visualization*, pages 272–277. IEEE, 1999.

[29] C. A. Steed, G. Shipman, P. Thornton, D. Ricciuto, D. Erickson, and M. Branstetter. Practical application of parallel coordinates for climate model analysis. *Procedia Computer Science*, 9(0):877 – 886, 2012. Proceedings of the International Conference on Computational Science, {ICCS} 2012.

[30] A. Tivellato. JOptimizer. `http://www.joptimizer.com/`.

[31] E. R. Tufte and P. Graves-Morris. *The visual display of quantitative information*, volume 31. Graphics press, 1983.

[32] D. N. Williams, T. Bremer, C. Doutriaux, J. Patchett, S. Williams, G. Shipman, R. Miller, D. R. Pugmire, B. Smith, C. Steed, E. W. Bethel, H. Childs, H. Krishnan, P. Prabhat, M. Wehner, C. T. Silva, E. Santos, D. Koop, T. Ellqvist, J. Poco, B. Geveci, A. Chaudhary, A. Bauer, A. Pletzer, D. Kindig, G. L. Potter, and T. P. Maxwell. Ultrascale visualization of climate data. *Computer*, 46(9):68–76, 2013.

[33] H.-M. Wu, Y.-J. Tien, and C.-h. Chen. Gap: A graphical environment for matrix visualization and cluster analysis. *Computational Statistics & Data Analysis*, 54(3):767–778, 2010.

[34] P. Zhang, F. Wang, J. Hu, and R. Sorrentino. Towards personalized medicine: Leveraging patient similarity and drug similarity analytics. *target*, 1(1):1.