Visual Analytics of Ensemble Data Using Coupled Subspaces

Fan Hong, Hanning Shao, Xiyao Mei, Alex Pang, and Xiaoru Yuan

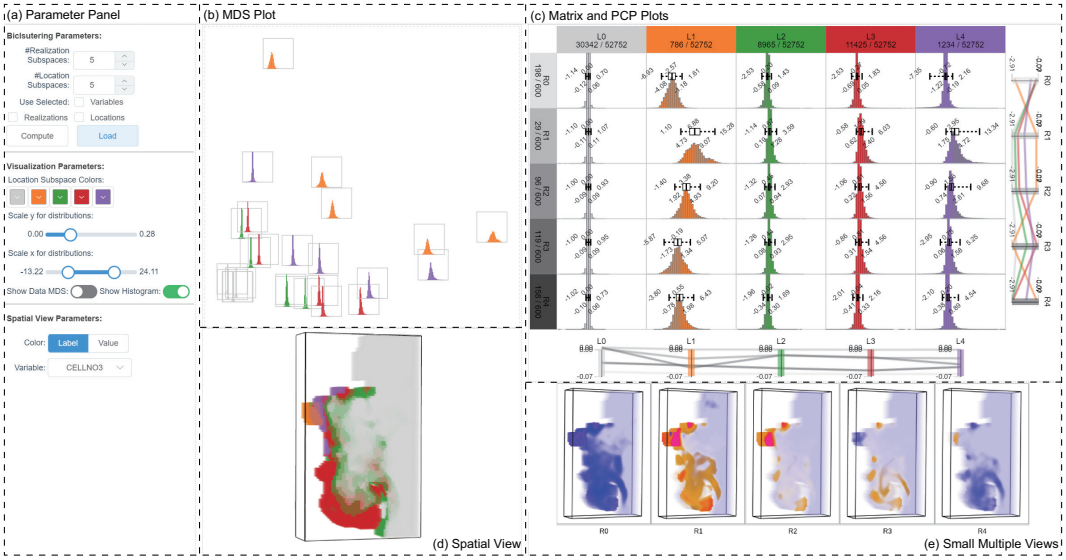

Fig. 1. System interface: (a) Parameter Panel; (b) Bicluster MDS Plot; (c) Bicluster Matrix View; (d) Spatial View; (e) Small Multiple View.

Abstract— In this work, we introduce a novel visual analysis method to explore correlations between locations and realizations of ensemble data with large realization set. In our approach, the ensemble data is transformed into a matrix with two dimensions: location and realization. The matrix is subdivided into cells which indicate behaviors of corresponding location-realization combinations. Biclustering is employed to simultaneously partition the location domain and the realization set, the two dimensions of the matrix, into several subspaces, where every intersection subspace shares similar coordinated behaviors. With the matrix subspaces and their coordinated behaviors, a visual analytics workflow is designed to support visualization and analysis of location-realization correlations and their variation across different location and realization. Case studies show that our system is able to reveal location-realization correlations, such as the opposite behaviors of realization subspaces in different regions, which are normally difficult to be discovered by previous methods.

Index Terms—Ensemble data, uncertainty, distribution, biclustering, subspace

1 Introduction

Ensemble simulations are important in many scientific research areas, such as in climatology, ocean circulation, and aerodynamics, etc. In many cases, scientists need to explore the impact brought by different simulation parameters. The sensitivity of simulation parameters can be studied by running a model several times with different initial values or boundary conditions. This can also reduce the uncertainty of a single simulation, and the results can be used to further improve the model. In this paper, each independent run of the simulation is referred to as a *realization* that represents an output of the simulation under a set of parameters. Together, all the *realizations* can be referred to as the ensemble dataset. Due to the complexity of ensemble simulations, it is challenging to extract and compare features from ensemble datasets.

Location and realization are two indispensable and closely related dimensions in ensemble data. By investigating the relations between these two dimensions, implicit features and data insights could be discovered. Existing ensemble visualization and analysis methods can be classified into location-based and feature-based [28]. Both of these methods only focused on one dimension regardless of the other correlated dimension such as realization. Hence, using these methods, it is impossible to extract and study the correlations between the location and realization as the information in other dimensions was lost due to the data aggregation and reduction. In location-based methods, a spatial field of distributions can be derived from realizations for visualization and analysis [3, 37, 46, 6]. However, intensive aggregation to distributions ignores information in realizations, which could lead to misleading visualization. For example, assuming two locations always have the same values but opposite signs (+/-) in every realization, their behaviors will be treated as the same in the above approaches, due to their identical distributions. In feature-based methods, spatial features are extracted and compared from individual realizations [39, 51, 34]. Due to potential visual clutter in the spatial domain, either sparse features, such as isocontours of a single isovalue, or fieldlines of a single

- *Fan Hong, Hanning Shao, Xiyao Mei, and Xiaoru Yuan are with Key Laboratory of Machine Perception (Ministry of Education), School of Intelligence Science and Technology, Peking University. E-mail: {fan.hong, hanning.shao, xiyao.mei, xiaoru.yuan}@pku.edu.cn.*
- *Xiaoru Yuan is also with National Engineering Laboratory for Big Data Analysis and Application, Peking University.*
- *Alex Pang is with the Computer Science Department, University of California, Santa Cruz, Santa Cruz, CA 95064. E-mail: pang@cse.ucsc.edu*
- *Xiaoru Yuan is the corresponding author.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

seed point, are used for visual comparison, or certain distance metrics are defined to measure the overall differences of features. Either way reduces the information regarding the spatial correlations and variations over the whole domain. It is possible that feature-based methods show that two realizations are very different in terms of defined features, yet the realizations can indeed have similar behaviors in local regions. As the two examples above demonstrated, current location-based and feature-based methods are difficult or unable to discover the valuable information regarding the location-realization correlations. The main reason is the severe aggregation on either the location dimension or the realization dimension, which makes it difficult for a system to unfold the dimension being aggregated and relate it to the other dimension once again.

In this work, to enable the extraction and analysis of the location-realization correlations, we propose a subspace-based approach that simultaneously investigates both the location and realization dimensions of ensemble data. Our approach is inspired by subspace analysis in high-dimensional data visualization. The concepts of subspaces are extended to ensemble data analysis. The *realization subspace* can be defined as a subset of realizations, and the *location subspace* is defined as a subset of locations – which can be connected or disconnected regions. Given a realization subspace, if some locations share similar behaviors in the realizations, we refer to them as *location clusters* in the realization subspace. The similar behaviors can also be treated as shared by *realization clusters* in a location subspace. Here, we use the general term *behavior* to indicate any information possessed by location-realization combinations, such as a single scale value from one variable or value vectors formed by multiple variables. The behaviors shared by one subspace cluster embody the correlation between locations and realizations, which is denoted as *coordinated behaviors*. Furthermore, with extracted subspaces, we can study how the coordinated behaviors change in the clusters of the same location/realization subspace, which corresponds to the variations of correlations.

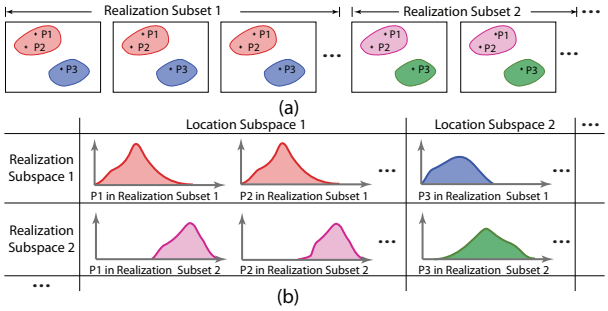

Fig. 2. Illustration of ensemble subspaces and coordinated behaviors in forms of (a) small multiples and (b) bicluster matrix.

In Fig. 2, we provide an illustration of ensemble subspaces and related concepts. Fig. 2(a), locations P1, P2, and other locations in the red regions exhibit similar behaviors in realization subspace 1, indicating correlations between the red region and realization subspace 1. The coordinated behaviors can be interpreted in two types of subspace clusters: location clusters in realization subspace 1 or realization clusters in location subspace 1, as shown in Fig. 2(b). Probability density function curves are used to describe their shared behaviors. Another important aspect is the variations in these coordinated behaviors across different subspaces. When transitioning to realization subspace 2, the behaviors of the location cluster become different, as the distribution changes from red to purple. Similarly, the behaviors of realization clusters change when transitioning to location subspace 2, as the distribution changes from red to blue. Therefore, through subspace analysis, we can not only discover location-realization correlations in ensemble data but also analyze data variations along the dimensions of either location or realization.

In our work, we employ biclustering techniques for the extraction of subspaces and coordinated relationships. Biclustering is a widely used subspace analysis method. By transforming ensemble data into a matrix, the biclustering algorithm can simultaneously partition the rows and columns into several subsets, revealing coordinated behaviors as sub-matrices, also known as biclusters. The row and column subsets correspond to realization and location subspaces, respectively, and the biclusters represent subspace clusters. Based on the biclustering results, we further develop a prototype system that supports a visual analytics workflow for interactive and progressive exploration of ensemble datasets. Our system supports three levels of visualization and analysis tasks: the data level, bicluster level, and subspace level. Specifically, it enables the visualization of subspaces and coordinated behaviors in biclusters, as well as the comparison of behaviors across biclusters and subspaces for studying the location-realization correlations and their variations. Progressive exploration based on previous biclustering results is also supported, allowing users to incorporate their knowledge into the visual analytics process. We demonstrate the usage of our system on datasets from various applications, including computational fluid dynamics, ocean circulation, and climate simulations. The results show that our system is capable of analyzing ensemble datasets with hundreds of realizations and enables the identification of richer details regarding the location-realization correlations that previous approaches were unable to discover.

2 Related Work

Our work falls within the domain of ensemble data visualization, which can be categorized into two main approaches: feature-based and location-based. In the subsequent paragraphs, we provide a comprehensive literature review in this field. The concept we introduce, ensemble subspaces, is inspired by subspace analysis in high-dimensional data visualization. We briefly review the relevant literature on biclustering techniques and subspace analysis.

2.1 Basic Biclustering Algorithms

Biclustering is a series of algorithms to extract subspaces from high-dimensional data by simultaneously clustering the rows and columns of a matrix. The problem and first algorithm was originally introduced by Hartigan [14] in 1972 and has been widely used in bioinformatics to study the coordinated relationships between genes and expressions. In the following, a concise introduction to the biclustering problem and the algorithm is given.

Given a data matrix $A = \{a_{ij}\}$ with n rows and m columns.A *bicluster* is a subset of rows I which exhibits coordinated behaviors across a subset of columns J, and vice versa. In other words, each bicluster represents a submatrix in the original data matrix. We denote a bicluster as $A_{ij} = (I_i, J_j)$, which involves a set of rows I_i and a set of columns J_j. In this way, the problem of biclustering can be defined as: given a data matrix A, we expect to identify a set of submatrix $\{A_{ij}\}$ which represents a set of biclusters $\{(I_i, J_j)\}$ such that cell values in each bicluster satisfy some definitions of similarities. If the matrix A represents a high-dimensional dataset with n data items and m dimensions, then I_i and J_j are data and dimension subspace, respectively, and A_{ij}s are subspace clusters.

The above definition is a general description of biclustering problem. As the application scenario varies, the problem could be specialized according to the bicluster structure it produces. In this work, we apply the non-overlapping biclusters with checkerboard structure, that is more commonly used in applications. In the checkerboard structure, the entire set of rows I and columns J are partitioned into K and L subsets respectively, denoted as $\{I_1, \cdots, I_K\}$ and $\{J_1, \cdots, J_L\}$ accordingly. For those row subsets, they satisfy the following conditions:

$$I_i \cap I_j = \emptyset, 1 \leq i, j \leq K, I_1 \cup I_2 \cup ... \cup I_K = I.$$

Likewise, those columns subsets also satisfy similar conditions. Thus $K \times L$ non-overlapping biclusters are revealed, and we can denote them using the submatrix mark A_{ij} or bicluster mark (I_i, J_j) as mentioned. If we reorder rows and columns so that elements in any row subset and column subset stay together, the checkerboard structure is revealed in the matrix (Fig. 3 top-right).

The original work of biclustering [14] provided a simplest but also highly extendable algorithm to discover checkerboard structures in a data matrix. The algorithm inputs include the data matrix $A = \{a_{ij}\}$ and the expected numbers of row and column clusters K, L. Then

labels are given for each row and column so that rows or columns with the same label will form clusters. And the aim of the algorithm is to simultaneously optimize the clusters of rows and columns. The outputs can be represented as two mappings s, t that maps row indexes and column indexes to labels:

$$s : \{1 \cdots n\} \to \{1 \cdots K\}, t : \{1 \cdots m\} \to \{1 \cdots L\}.$$

And we will use $s^{-1}(\cdot)$ and $t^{-1}(\cdot)$ to represent the subsets of rows and columns with certain labels. Following steps are executed until convergence is reached:

1. Initialize clustering labels for rows s and columns t.
2. Calculate the number of rows or columns assigned to each label, denoted as r_i for row label i and c_j for column label j.
3. Evaluate the biclustering quality, here we calculate the mean value of each bicluster. For bicluster (I_k, J_l):

$$m_{kl} = \frac{1}{r_k \times c_l} \sum_{i \in s^{-1}(k)} \sum_{j \in t^{-1}(l)} a_{ij} \tag{1}$$

4. Update clustering labels for rows and columns:

$$s(i) = \underset{k \in \{1, \cdots, K\}}{\arg\min} \sum_{j=1}^{m} (a_{ij} - m_{k,t(j)})^2, 1 \le i \le n. \tag{2}$$

$$t(j) = \underset{l \in \{1, \cdots, L\}}{\arg\min} \sum_{i=1}^{n} (a_{ij} - m_{s(i),j})^2, 1 \le j \le m. \tag{3}$$

5. Jump to step 2 if labels $s(\cdot)$ and $t(\cdot)$ are changed, otherwise stop.

When the algorithm stops, $s(\cdot)$ and $t(\cdot)$ are the final partition schemes for rows and columns respectively, where $s^{-1}(\cdot)$ and $t^{-1}(\cdot)$ are corresponding subsets, respectively.

2.2 Ensemble Visualization

Ensemble simulations are commonly utilized to investigate model uncertainty and parameter sensitivities. The visualization methods for ensemble data have considerable overlap with those used in uncertainty visualization and comparative visualization [21, 24, 31]. These methods can be classified into two categories: feature-based and location-based [28].

In feature-based methods, spatial features are first extracted from each individual realization and then compared and visualized. For scalar fields, there has been extensive research on uncertainty visualization of spatial features such as isocontours and isosurfaces. Several methods address the extraction of isocontours from scalar fields with uncertainty [2, 1, 36, 35, 33]. Alternatively, isocontours can be directly extracted from each realization and their uncertainty visualized. For instance, Ensemble-vis [39] introduced the spaghetti plot technique to visualize isocontour ensembles from each realization. Building upon the spaghetti plot, Noodles [41] employed glyphs and confidence ribbons to highlight their uncertainty. More advanced techniques have been developed to describe uncertainty, such as contour boxplots [51] and structural variability [34]. When isocontours vary over time, Ferstl et al. [9] employed hierarchical clustering to study the divergence of isocontour ensembles along the time dimension. Zhang et al. [56] utilized kernel density modeling to provide ensemble modeling with different levels of abstraction. In the case of ensemble vector fields, topological structures are commonly extracted as spatial features for comparison [29, 30, 32]. Apart from visualizing uncertainty in spatial features derived from simulation realizations, some methods directly define and calculate overall similarities between realizations and employ clustering techniques for analysis [20, 13].

Location-based methods visualize or compare distributions obtained from every location in the spatial domain. Traditional visual encodings involve statistical values such as mean, standard deviation, moment, and interquartile range [25, 38]. Shu et al. [48] visualized regions with high variation using a storyline metaphor to depict the evolution of uncertainty. However, statistical values can only capture a few major characteristics of value distributions. Particularly, when the distributions are not assumed to be Gaussian, statistical values have limitations in describing the distributions. Subsequent methods often utilize histograms or density probability functions for visualization [3, 37, 46]. Demir et al. [6] applied clustering methods to histograms to help identify regions with similar distributions.

When dealing with vector fields, certain works directly visualize the uncertainty of directional speeds using glyphs [52, 23, 15]. Jarema et al. [20] further employed Gaussian mixture models to fit directional distributions and visualized the Gaussian components using glyphs. To handle non-Gaussian bivariate distributions, Hollister et al. [16] introduced interpolation methods to address integration requirements of ensemble vector fields. Several approaches focus on visualizing uncertainty through streamlines or pathlines [47]. As the number of simulation runs increases, curve boxplots [27] and streamline variability plots [7] were proposed to summarize the distribution of streamlines originating from a single location. A similar concept is applied to isocontours, where users can examine the joint occurrence of isocontours in local regions with interactive techniques [8]. However, these approaches face challenges in visualizing the variation of streamlines originating from multiple grid points due to visual clutter. Therefore, methods with a higher level of summarization have been proposed. Hummel et al. [18] and Liu et al. [22] defined and calculated transport variances at each grid point to represent overall uncertainty. Hollister et al. [17] used DBSCAN clustering to analyze divergence from multimodal distributions within streamlines. Guo et al. [12] measured variation in the attribute space rather than the geometric space.

However, both feature-based and location-based methods have their limitations. Feature-based methods focus on highlighting differences and uncertainty between realizations, but they provide limited locational information, relying on aggregated features such as isocontours, isosurfaces, or topology. On the other hand, location-based methods can visualize and compare locational information through distributions, but they often lose spatial correlation information within each realization. In our approach, we propose a method for the coupled analysis of locations and realizations, which can be considered a hybrid between the two aforementioned approaches.

2.3 Subspace Analysis and Biclustering Techniques

In the field of high-dimensional data visualization, subspace analysis is a category of approaches that focuses on identifying meaningful features in high-dimensional spaces. Since many data clusters are not easily distinguishable in the full-dimensional space, researchers have started exploring clusters in lower-dimensional subspaces, which are spaces spanned by a subset of dimensions. Dimension projection matrix/tree [55] enables interactive and progressive exploration to identify significant subspaces and subspace clusters. Tatu et al. [45] defined appropriate subspace similarity functions for automatic searching, grouping, and filtering of subspaces, generating numerous interesting subspaces for users to gain different perspectives on the data. To facilitate exploration, a framework was developed to decompose high-dimensional data into a continuum of generalized 3D subspaces, allowing data analysts to interactively visualize and understand the data using visual tools while utilizing familiar trackball interfaces and smooth transitions to adjacent subspaces for enhanced comprehension.

In addition to interactive subspace exploration, numerous automatic subspace clustering algorithms have been proposed to detect data clusters in various subspaces. Biclustering, a type of subspace clustering method extensively used in bioinformatics to study the coordinated relationships between genes and expressions [26], has also been introduced into the visualization community to aid in data visualization. Watanabe et al. [49] employed biclustering for correlated subspace mining in multivariate data, as multivariate data can naturally be represented as a data matrix. Subsequently, biclustering techniques have been adopted in various applications, such as exploring cross-view data relationships [44] and mining events co-occurrence [54]. He et al. [50] utilized biclustering to investigate relationships among variables in multivariate data. In our work, we utilize biclustering techniques for subspace analysis in ensemble data to extract coordinated behaviors between locations and realizations.

The visualization of biclustering results poses a non-trivial challenge. Sun et al. [43] proposed a five-level design framework to summarize bicluster visualization, which includes visualizations at the entity level, group level, and bicluster level. Matrix-based visualizations [42, 10, 57], parallel coordinated-based visualizations [11, 53], and node-link diagrams [40] have been widely used in many works. Some of our visualization approaches are inspired by these works, and we also provide data-specific visualizations to enhance the understanding of ensemble data.

3 Requirements and Tasks Analysis

For ensemble data, locations and realizations naturally form two indispensable and closely related dimensions. Basically, there are analysis requirements from three aspects when dealing with ensemble data:

R1: Location aspect: Find the similarities and differences between locations. Given a set of realization, analyzers want to identify the location clusters that share common features.

R2: Realization aspect: Find the behavior patterns between realizations. Analyzers want to identify the realization clusters that have similar influence on selected location domain.

R3: Data aspect: Check the original data at selected locations under assigned realizations.

Previous works mostly focus on analyzing single dimension, either the location or the realization. For example, the comparison between realizations over the whole domain or specifically assigned locations (**R2**), or the exploration of the distributions of cluster locations over the entire realizations (**R1**). However, the distributions of locations over all realizations are not isotropic, neither the behaviors of realizations on the location domain.

So, there is another one that involves both dimensions:

R4: Coordination of location and realization: Analysis the coordinated behaviors of locations and realizations simultaneously. Exploration and comparison across both dimensions are required, where analyzers can extract clusters sharing similarities from locations and realizations at the same time.

In this work, we apply biclustering-based approach to take both dimensions into consideration simultaneously (**R4**). Table 1 compares our approach with some other works with regard to those requirements. As we can see, realization-based methods [39, 9, 20, 13, 56] can barely meet the requirements in location aspect (**R1**), while location-based methods [6, 48, 5] can hardly meet the requirements in realization aspect (**R2**), not to mention the coordinated analysis requirement (**R4**). Exceptionally, a few works [20, 48] meet part of **R4** that they allow users to extract clusters firstly on one dimension and then explore the distribution on the other one. This two-steps approach satisfies part of the coordinated analysis requirement while still has inevitable limitations compared to simultaneous methods like biclustering. We will discuss about this later, and they are marked with a star in the table, indicating that they partially support coordinated analysis requirements.

Table 1. Comparison of design requirements supported by different ensemble analysis approaches.

	Location	Realization	Coordinated	Data
Our approach	+	+	+	+
Ensemble-Vis [39]	-	+	-	-
Ferstl et al. [9]	-	+	-	+
Hummel et al. [20]	-	+	*	+
Hao et al. [13]	-	+	-	+
Demir et al. [6]	+	-	-	-
EnsembleGraph [48]	+	+	*	-
EnConVis [56]	-	+	*	-
Souza et al. [5]	+	+	-	-

According to the four types of analysis requirements, we further conclude five analysis tasks that the visual analytic system should support to explore the ensemble data.

T1: Visualize bicluster structures, showing the subspace partitions.

T2: Visualization of location-realization coordinated behaviors in individual biclusters, showing the data distribution and basic statistics.

T3: Comparison of behaviors of biclusters in single subspace, to study how coordinated behaviors vary in one subspace.

T4: Comparison of biclusters behaviors in multiple subspaces, to study the positive/negative correlations between subspaces.

T5: Comparison of behaviors of arbitrary biclusters, to identify the overall distribution of coordinated behaviors.

T1 shows the clustering results, combining with **T2**, the analysis requirement of data (**R3**) is satisfied since users can check the data behind each single clusters. **T3** meets the requirements from single data dimension (**R2** & **R3**) that users can check the behavior patterns within selected dimension of the ensemble datasets. **T3** limits the comparison within single subspace, whichever dimension the subspace is extracted from, so it does not involve any cross dimensions analysis. While T4 and T5 support the coordinated analysis of two dimensions (**R4**), since subspaces from different dimensions are compared together. In table 2, we summarize the correspondence between them.

Table 2. The correspondence between analysis requirements, tasks, and the view designs.

	Requirements				View Designs				
Tasks	R1	R2	R3	R4	MDS	PCP	Matrix	Spatial	Small Multiples
T1			+					*	*
T2			+		*		*	*	*
T3	+	+			*	*	*	*	*
T4	+	+		+	*	*	*		
T5	+	+		+	*		*		

4 Ensemble Subspace Extraction

In this work, we introduce a visual analytics system that utilizes a biclustering algorithm to facilitate ensemble analysis based on the concept of subspaces. Our approach focuses on extracting subspaces from both realizations and locations simultaneously. The pipeline of our biclustering-based approach is shown in Fig. 3.

In this section, we introduce the adaptations made to the original biclustering algorithm mentioned previsouly to enable support for matrics of value vectors for multivariate ensemble data. In addition, we also extend the Calinski-Harabasz criterion from clustering to biclustering to assist the choosing of the bicluster number.

4.1 Biclustering for Ensemble Data

To apply biclustering method to ensemble subspace analysis, the important step is to transform the ensemble data to a data matrix (Fig. 3 top-middle). The rows and columns in the biclustering algorithm correspond to the realizations and locations of ensemble data naturally. The extracted biclusters correspond to certain coordinated behaviors shared by realization clusters over location subspaces, and vice versa. Since rows and columns are interchangeable, we just let the rows represent realizations and the columns for locations.

The domain of the ensemble data can be represented as $E : \{L_1, \cdots, L_M\} \times \Omega$, where $\Omega \subseteq R_D$ denotes the D-dimensional spatial domain, and M denotes the number of realizations. The range of E depends on the type of ensemble datasets, such as $\mathbb{R}$ for a single scalar field, $\mathbb{R}^{|V|}$ for multivariate scalar fields with V variables, etc. Given a spatial location $\mathbf{p}_j \in \Omega$, the value produced by realization i can be denoted as $E(i, \mathbf{p}_j)$. To employ biclustering algorithm for ensemble subspace analysis, we need to define certain transformations to form a data matrix, that is $a_{ij} = \mathcal{F}(E(i, \mathbf{p}_j))$. Experts can define the most suitable $\mathcal{F}$ according to their data and analytic requirements. For ensemble data with a single scalar field, we directly use the scalar values in the data matrix, that is $a_{ij} = \mathcal{F}(E(i, \mathbf{p}_j)) = E(i, \mathbf{p}_j)$. It is also possible to choose other transformations such as normalized values when needed.

And for ensemble datasets with multivariate scalar fields, which are also common in scientific simulations such as climate simulation and ocean circulation simulation, we apply two alternative manners to handle multivariate ensemble data. We denote $E(i, \mathbf{p}_j) =$

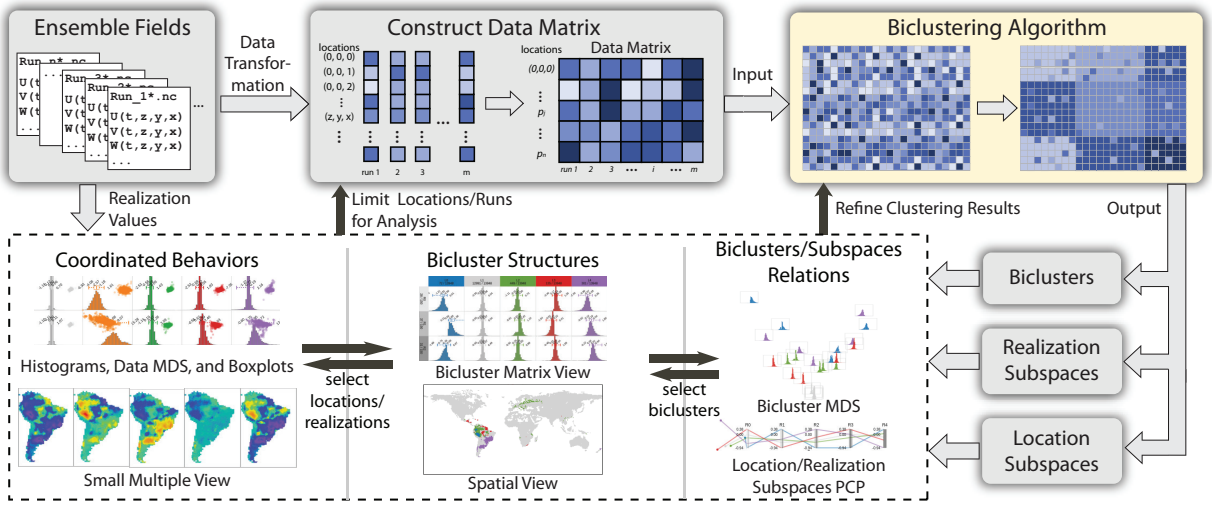

Fig. 3. The pipeline of our approach, including data transformation, biclustering algorithm, and interactive visualization.

$\vec{v}_{ij} = (v_1, \cdots, v_V)_{ij}$ as the value vector at location $\mathbf{p}_j$ in i-th realization.The two mainly used ways to handle multivariate data is extending univariate biclustering algorithms, and reducing value vectors to single values. For the first approach, we extend the Hartigan's algorithm introduced previously, where we modify the steps of mean value calculation and label update. For the former step, summation operations in Equation 1 can be directly applied onto vectors.For the label update step, differences between each cell and mean vectors need to be computed. The costs of label assignments are adjusted from the mean squared differences to L2-norm of vectors, i.e. $\sum_{j=1}^{m} ||a_{i,j} - m_{i,t(j)}||^2$ and $\sum_{i=1}^{n} ||a_{i,j} - m_{s(i),j}||^2$ respectively. Direct L2-norm could introduce bias for different variables, due to unmatched value ranges. So the value vectors are normalized when needed.

Another approach is to reduce value vectors to single values, but keep their distances as much as possible. Following the L2-norm differences measure, we can choose a transformation $\mathcal{F}$, so the total error of similarities caused is minimized, that is

$$\min_{\mathcal{F}}(\sum_{s,t}(|\mathcal{F}(\vec{v}_s) - \mathcal{F}(\vec{v}_t)| - \|\vec{v}_s - \vec{v}_t\|)^2).$$

This objective function can be solved by Multidimensional Scaling (MDS), a commonly used technique to represent high-dimensional data in a low-dimensional space. In our situation, we embed the multivariate vectors into 1-dimensional space [19] and use it as the matrix cell values. The benefits brought by the compact representation are considered to overcome the accuracy loss, especially when the multivariate data has other dimensions, which are locations, realizations, and variables in our case.

This two approaches for multivariate scalar field datasets, extending the Hartigan's algorithm or reducing the vectors to single values, have their own advantages under different conditions. The former one keeps the original vectors and thus may improve the biclustering results compared with the latter one, since recuding the vectors to single values actually change the data. However, it also brings great computing complexity to the biclustering algorithm for doing vector arithmetic. While the negative effect of vector dimension reduction can be eased through carefully designed reduction methods. It is noteworthy that both approaches are solutions of finding a function $\mathcal{F}$ to transform the ensemble data into biclustering matrix. In our system, we provide these two approaches, while there are still alternative chooses. Experts can select the most suitable one dealing with their data and special analysis tasks.

4.2 Decision of the Number of Biclusters

The most important parameters for biclustering algorithm are the numbers of biclusters $K \times L$, i.e. the number of realization and location subspace.However, it is not trivial to set appropriate values to obtain ideal results. Although users can keep trying different parameters until good results are obtained, such trial-and-error way is quite inefficient. For this reason, we extend Calinski-Harabasz criterion [4] from 1D to 2D biclustering to measure the clustering quality of given parameters.

The intuition of Calinski-Harabasz criterion is that the overall within-cluster variance should be as small as possible, while the overall between-cluster variances should be as large as possible. The overall within-cluster variance SS_W is defined as $SS_W = \sum_{i=1}^{k} \sum_{x \in c_i} ||x - m_i||^2$, where k is the number of clusters, c_i is the ith cluster, x is a data point, and m_i is the center of cluster c_i. The overall between-cluster variance SS_B is defined as $SS_B = \sum_{i=1}^{k} n_i ||m_i - m||^2$, where m_i is the center of cluster c_i, n_i is the size of cluster c_i, and m is the overall mean of the sampled data. Then Calinski-Harabasz criterion is defined as $S = \frac{SS_B}{SS_W} \times \frac{N-k}{k-1}$, where N is the total number of data points. Larger scores indicate better clustering quality. In the biclustering problem, since the algorithm cluster cells, rows, and columns at the same time, we evaluate the quality of all three clustering results. Specifically, we extend Calinski-Harabasz criterion as a weighted combination of three individual scores, i.e. $S = S_{\text{cell}} + m \times S_{\text{row}} + n \times S_{\text{column}}$. The scores for rows and columns are scaled with factors m and n respectively, to lift them to the same scale of the cell scores.

Fig. 4(a) shows the criterion matrix with different parameter settings ranging from 1×1 to 10×10. The scores in the matrix show a significant increase when the number of location subsets or realization subsets increases from 1 to 2. This observation supports our motivation for subspace analysis, indicating the presence of correlations and variations between the two dimensions of location and realization. Additionally, it can be observed that when the number of realization subspaces K reaches certain values (e.g., 2, 3, 5, etc.), the scores become noticeably larger than others, suggesting these values as good choices. In our work, we choose $K = 5$ as the default value.

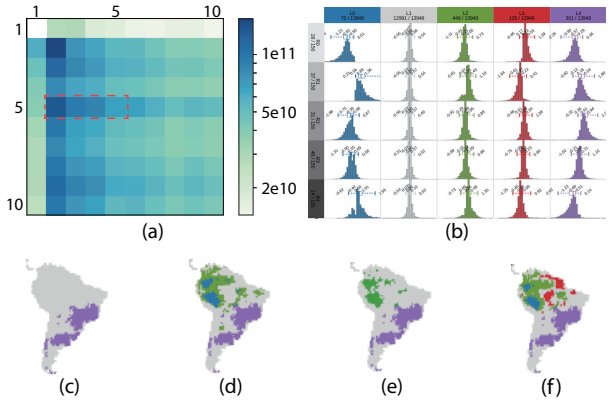

Fig. 4. (a) Calinski-Harabasz criterion matrix. (b) Distributions of extracted biclusters. (c-f) Spatial distributions of location subspaces keep stable as the number of biclusters increases.

We further explore the clustering results over location subspaces in terms of the bicluster numbers. Fig. 4(c-f) shows the spatial partition results, i.e. location subspaces, with bicluster numbers from 5×2 to 5×5. As the location subset number increases, we can observe that previously extracted subspaces are kept well, new subspace only being separated from the largest gray one. For example, the purple region in Fig. 4(c) is also extracted as a cluster in (d-f) when biclustering parameter L increases from 2 to 5. From the bicluster matrix view (b), we can also find that the value distributions of biclusters in the new subspaces are indeed different from previous ones. From this aspect, the stability of the biclusterings on location is adequately well. As a result, users can choose the setting of parameter L according to the data and analysis tasks freely. In this work, we adopt $L = 5$ as default.

In all, to tackle the problem of parameter selection, we study the settings of bicluster numbers and their stability. As a result, our system renewcommands a $K \times L = 5 \times 5$ biclustering setting as default, while users are allowed to change those parameters when needed.

5 Prototype System and Interface

To support users to comprehensively explore coordinated behaviors in ensemble subspace and their variations, we further implement a visual analytic system driven by the adapted biclustering algorithm described above. In the system, several linked views are presented to visualize the results from different facets as shown in Fig. 1.

Our system supports visual analytics workflow as shown in Fig. 5, so that users are able to analyze the ensemble data progressively and comprehensively.

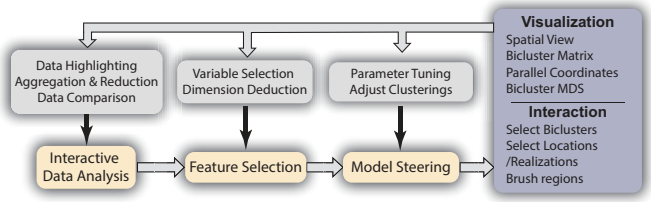

Fig. 5. Visual analytics workflow supported by our system.

Users are provided with interactive parameter setting for the biclustering algorithm, and the clustering results are displayed in multiple linked views, facilitating interactive analysis. Users can select and highlight specific data of interest, ranging from individual realizations and locations to biclusters and subspaces. The system enables comparison tasks by allowing users to perform linked operations across different views. They can gain insights into the relationships between biclusters and subspaces, uncovering coordinated behaviors within the ensemble data. Furthermore, our system supports flexible and progressive analysis. Users have the option to modify the biclustering parameters to re-run the clustering process. Alternatively, they can also focus on selected subspaces to refine the provided results, that is to apply biclustering algorhtm exactly within selected subspaces . Through interactions, users can further divided or re-clustering selected subspaces to have a deeper exploration of the data. With the proposed analysis workflow, our system empowers users to interactively and progressively explore ensemble data from various levels and perspectives, catering to their specific analysis tasks.

5.1 Visualization System

A five-level design framework is presented by Sun et al. [43] to summarize biclustering visualization, where the entity level, group level, and bicluster level are almost used in most existing works. When it comes to our application scenarios, i.e. ensemble data analysis, we present more data-specific visualization techniques: volume rendering techniques (2D and 3D), distribution-related visualization, and biclusters MDS plot for comparison.

In our interface, the bicluster and subspace labels are used globally. We use R_i and L_j to represent **R**ealization and **L**ocation subspaces respectively. Then, each bicluster is denoted as R_iL_j, that is the crossing of the location subspace and realization subspace. In addition, we adopt consistent color schemes for easy recognition of subspaces in different views. For location subspaces, we choose colors with different hues, while for realization subspaces, to avoid confusion, we use gray color with different intensities. Users can adjust the colors to meet their needs better freely.

Bicluster Matrix and PCP Plots

The matrix gives (Fig. 1(c)) an initial perception of the structure of biclustering results. In the bicluster matrix, each row and column corresponds to one realization and location subspace, respectively. The cells represent biclusters. In the top and left header areas of the matrix, we use colored blocks to indicate subspaces, as well as their labels and the number of elements contained. Users are allowed to select one subspace to conduct progressive computation described previously.

Histograms are derived from the values of realization-location combinations within each bicluster. These histograms are displayed in the cells of the matrix. The axes of the histograms are unified to facilitate easy comparison. The histograms provide a clear presentation of the coordinated behaviors within the biclusters, including measures such as mean, variance, skewness, and Gaussianness. In addition, boxplots are superimposed to show basic statistics. For multivariate ensemble data, users can decide to show either histogram of single variables and/or an MDS plot deriving from multiple variables.

Brushing operation is also supported in the histogram of one bicluster. The value distributions of selected location-realization combinations in other biclusters of the same subspaces are highlighted. Based on distribution shapes and statistics, as well as the data behaviors across different subspaces, users can decide if it is necessary to split one subspace further for deeper investigation.

Two parallel coordinates plots (PCP) are used to visualize the behaviors of biclusters across different subspaces just beside the matrix. In the PCP for locations (right of the bicluster matrix), each axis represents one realization subspace aligned with the rows of the bicluster matrix, while each line represents one location subspace, whose values are the averages of their crossing biclusters. Users can easily identify similarities and variations of coordinated behaviors within biclusters in one subspace.

Bicluster MDS Plot

A bicluster MDS plot (Fig. 1(b)) is provided to show the similarities between arbitrary biclusters in addition to the bicluster matrix. The distance between any two biclusters is defined as the Jensen-Shannon divergence of their distributions. Biclusters with similar distributions are projected to near positions in the plot because of the property of MDS algorithm. Similar to the cells in the bicluster matrix, 1D histograms and/or 2D data MDS plots are shown in each bicluster cell.

The bicluster MDS plot can be exploited in two ways. Firstly, it enables the identification of clusters of distributions with distinctive properties such as left-/right-skewed or normal distributions. These distribution clusters indicate that the corresponding location-realization combinations share similar coordinated behaviors. Users can select these distribution clusters to examine their associated biclusters and subspaces, facilitating the exploration of relationships between realizations and different subspaces. This approach enhances the understanding of correlations within the ensemble data. Secondly, users can select subspaces in the bicluster matrix view to observe their similarity and difference in the MDS plot. In the cases, we have found some subspaces show very similar distributions in their containing biclusters. While for some others, the distributions of their biclusters could be very different, indicating significant variations of correlations in single subspaces. The information regarding the variations of correlations is especially difficult and inconvenient to be found with existing approaches.

Spatial View and Small Multiple View

To obtain the physical meaning of extracted subspaces and biclusters in ensemble data analysis, it is necessary to relate biclustering results with their spatial distribution. We provide the spatial view (Fig. 1(d)) and small multiple view (Fig. 1(e)) for this purpose.

In the spatial view, the location subspace partition is rendered in user defined color scheme. Users can easily observe the spatial distribution of location subspaces, and relate them to physical environment. Besides basic navigation interactions, such as zooming and panning, we support brushing function as well as country-wise selection in some specific datasets. The brushing operation is linked with the

histograms in the bicluster matrix, so that users can analyze the behaviors of brushed locations in detail.

In the small multiple view, for each realization subspace, we render the averaged fields over the corresponding realizations. Users can easily compare behaviors of different regions within one realization subspace. At the same time, all small multiple views use a unified diverging colormap: the orange-red color indicates the realization subspace produces larger values than the average, while the blue-purple color is for smaller values. Comparison of regional behaviors across different realization subspaces is enabled.

In all, we provide general biclustering visualization as well as ensemble data-specific visualization to help users understand the coordinated behaviors from different levels and facets. Comprehensive exploration on biclusters and subspaces are enabled through interactions and multiple linked views. In addition, as users' knowledge about the ensemble data grows, users are allowed to progressively explore the data with biclustering algorithm to obtain a deeper understanding.

6 Case Studies

In this section, we describe case studies in two ensemble datasets: carbon emission and ocean circulation. We demonstrate how to make discoveries through introduced analysis workflow with our system. All these datasets contain a large number of realizations, and thus the coordinated behaviors in realization subspaces are more statistically significant. On the other hand, such a large number of realizations also makes it more critical to study their correlations with locations.

6.1 Carbon Emission Dataset

Carbon emission data comes from the global CO_2 flux simulation, which has 150 realizations and a spatial resolution of 360×180. We select one time slice at Jan. 2, 2010 from the original datasets and focus on the optimized carbon emission value from the terrestrial vegetation source. Significantly, only land region has valid values.

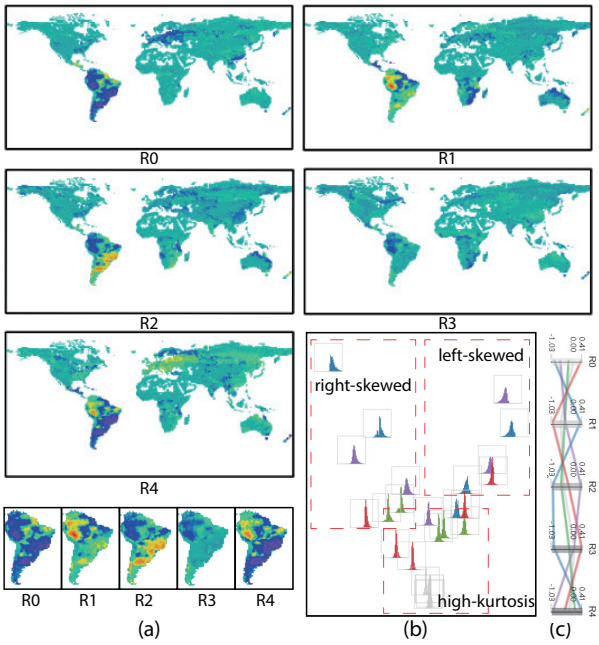

Fig. 6. (a) Small multiples view rendering the average fields of every realization subspace. (b) Bicluster MDS plot showing the similarity among biclusters. (c) Parallel coordinates plot supporting direct comparison of location subspaces across different realization subspaces.

The biclustering results are shown in Fig. 4 and Fig. 6. In the spatial view, Fig. 4(f), subspace L1 covers most of the land mass, and shows high kurtosis concentrated near 0. Subspaces L0 and L4, the blue and purple regions, cover most of South America. Subspace L2, the green region, mainly covers Europe and the coastal region of southeast Africa. Subspace L3, the red region, is scattered around North and South American. In Fig. 6(a), the land mass is colored according to the averaged field value for each realization subspace. Looking at the region covered by the L0, L3 and L4 subspaces, i.e. South America, we can observe that there is quite a significant amount of variance across the different realization subspaces. Scientists explain that the high forest cover in South America significantly affects the field value, i.e. carbon emission from terrestrial vegetation source. Except L1, all other location subspaces correspond to regions with forest cover larger than about 25%. Furthermore, we can make the hypothesis that regions with higher forest cover also have larger variances across realizations. The variances across different realization subspaces are also supported by checking the bicluster matrix (Fig. 4(b)).

Next, we study the relationships between biclusters and subspaces. In the bicluster MDS plot (Fig. 6(b)), the overall similarity relations among biclusters are shown. From the projection space, all biclusters are roughly divided into three categories: left- and right-skewed, and high-kurtosis distributions. Especially, for the blue location subspace L0, the distributions of biclusters in it are either strongly left-skewed or right-skewed, indicating strong variations of coordinated behaviors across different realization subspaces. Similar behaviors of variations can also be observed for the purple location subspace L4. From the parallel coordinates view of the location subspace (Fig. 6(c)), we can further observe that L0 and L4 show similar distributions in R0 and R1, but exhibit distributions with opposite skewness in R2, R3, and R4. These positive and negative correlations among distributions, which have not been reported in previous works, provide opportunities for studying the simulation conditions.

6.2 Ocean Circulation Dataset

The ocean circulation data comes from a simulation model which covers a region of Massachusetts Bay on the east coast of USA. The simulation is run over a 3-dimensional spatial domain with a spatial resolution of $90 \times 53 \times 16$.Part of the domain has no data, which represents the land. In our figures, these are rendered in totally transparency. This ensemble contains 600 realizations. The data contains 9 scalar variables that researchers are interested in, such as $temp$ (temperature), $salt$ (salinity), $NO3$ (nitrate concentration), $CELLNO3$ (cellular nitrogen from nitrate), etc. We employ the two methods aforementioned to process this multivariate ensemble data: direct computation using value vectors, and using 1D MDS reduced values.

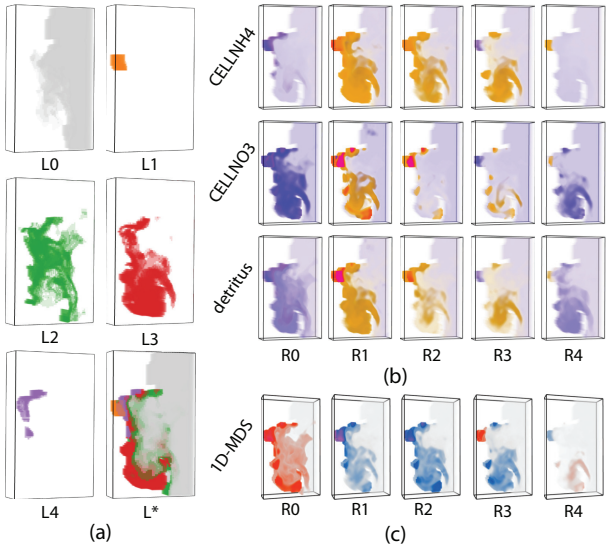

Fig. 7. Visualization of biclustering results for ocean circulation dataset (together with Fig. 1): (a) Spatial view showing the coverage of individual location subspaces and the overall partition; (b) Small multiple views for average fields of variables CELLNH4, CELLNO3, and detritus of each realization subspace, respectively. (c) Small multiple views for average fields of 1D MDS values.

In Fig. 1(b,c), we present the overall biclustering structures in the matrix and the similarity relationships of all biclusters in the MDS plot. The value distributions of the variable CELLNO3 are visualized in the cells. At the same time, the spatial coverage of location

subspaces is shown in Fig. 7(a). By relating above visualizations, we find that as we move deeper and westward, the variations of biclusters in corresponding location subspaces become larger. The increasing variances can be observed in all realization subspaces, even R1 and R2 show different skewness from others. Domain scientists provide an explanation regarding the location subspaces and their coordinated behaviors. The western and deep seashore regions, represented by L1 and L4, are convergence zones for warm and cold currents, which lead to more disturbances and mixing. Therefore, the variations of coordinated behaviors in these regions are higher than those in other regions.

In Fig. 7(b) and (c), the results of two approaches handling the multivariate data are shown. In (b), we select three highly correlated variables, CELLNH4, CELLNO3, and detritus.We can easily find that, although realization subspaces show very different coordinated behaviors in each region, they show consistent behaviors in these three variables. In (c), we directly use the 1D MDS values for analysis. It is interesting that the shapes of visible features are consistent in the average field of each realization subspace. Note that the exact colors in (b) and (c) are not comparable, but we can observe their relative values. For example, in R3, L1 shows significant different values compared to L3, which can be observed in both (b) and (c) easily. The above observation indicates that both approaches are able to identify meaningful correlations among variables as well as the location-realization correlations. Domain scientists explain that the correlations among variables are strongly related to the nitrogen cycle in the ocean. Bacteria decompose and convert detritus of fish and other organisms into ammonia (CELLNH4) and nitrate (CELLNO3). Such conversion makes the concentration of the three variables change accordingly.

The two case studies above have demonstrated the usage and effectiveness of our system. It can reveal the coordinated behaviors in ensemble subspaces, especially in exploring the variations of these behaviors across the dimensions of locations and realizations, which are difficult for previous approaches.

In the ocean circulation case, we have demonstrated that our approach can effectively extract correlations among variables, as well as the correlation between location and realization. Our approach enables us to obtain meaningful results and explanations by analyzing these correlations simultaneously.

7 Discussion

In this section, we will primarily discuss the differences between our approach and existing works, as well as the scalability of our approach.

Comparison with Other Methods

The novelty of our approach comes from the coupled analysis of realizations and locations with subspaces, while existing methods typically focus on one aspect and heavily aggregate the other. In Table 3, we compare the analysis tasks supported by these approaches. The advantage of our approach is the automatic extraction of meaningful clusters in subspaces and the subsequent analysis of their coordinated behaviors and how they vary across different subspaces. Location- and realization-based approaches are unable to analyze their correlation as they lack information about the other dimension. Jarema et al. [20] proposed a method that supports both approaches subsequently in their analysis method. However, since their approach is still based on the derived distributions, realization correspondence is not in the consideration at the very beginning. Therefore, they are still unable to acquire knowledge regarding the actual location-realization correlations.

Table 3. Comparison of analysis tasks supported by different ensemble analysis approaches. * indicates that manual selection is needed for corresponding tasks, where the subsets reflect users' intention, rather than objective clusters or subspaces.

	Correlation and Variation of Locations Considering			Similarity and Uncertainty of Realizations Considering		
	A Single Run	A Run Subset	All Runs	A Single Location	A Location Subset	The Whole Domain
Location-based [25, 6, 52, 23, 15, 20, 18, 12, ...]	-	+*	+	+	-	-
Realization-based [39, 41, 51, 29, 32, ...]	+	-	-	-	+*	+
Our Approach	+	+	+	+	+	+

In addition to existing ensemble visualization and analysis approaches, one alternative approach to achieve coupled analysis is to perform clustering twice, either by clustering the data based on realizations first and then clustering locations within each realization cluster, or vice versa. However, our experiments have demonstrated that these alternatives are unable to accomplish our objective of coupled analysis. In Fig. 8, ensemble data is firstly clustered into 5 realization clusters, then locations are clustered within each realization cluster. The location clustering labels of Europe area in each realization cluster are shown in small multiples, where each sub-figure of five indicates one realization subspace, and the location subspaces are represented using different colors . Although coordinated behaviors can somehow be revealed, it is almost impossible to study how they change across different regions, since the location clusters are not aligned in different realization subspaces . For example, the large yellow part in the first sub-figure can not be found in other sub-figures. Therefore, without additional constrains in the two-level clustering, the results are prone to inconsistency. In contrast, our biclustering method imposes simultaneous restrictions on coordinated behaviors based on both realizations and locations. As a result, our approach allows for the discovery of abundant information regarding the correlation between locations and realizations, particularly the variations in coordinated behaviors across different subspaces.

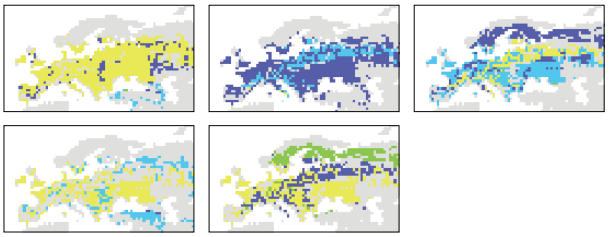

Fig. 8. Visualization of the alternative approach: First cluster data by realizations, and then by locations. Clustering labels of Europe area in each realization subset are shown.

Scalability of Our Approach

We discuss the scalability of our approach in two aspects: computation and visualization.

The computation in our approach mainly comes from the iterative biclustering algorithm. In each iteration, the mean value for each bicluster are calculated, and cluster labels are updated. For mean value calculation, the data matrix is traversed once and the time complexity is $O(nm)$. To update rows labels, each possible label $k \in \{1, \cdots, K\}$ is tested by computing corresponding costs for each row. So the time complexity is $O(Knm)$. The column labels update is similar, and the time complexity is $O(Lnm)$. Hence, the overall time complexity is $O((K + L)nm)$. In terms of computation, the total time cost is also influenced by the convergence speed, which generally needs less than 20 of iterations. In summary, our approach provides good scalability concerning the data size.

Previous approaches often disregard the number of realizations, since it is typically much smaller compared to the number of locations. However, in our scenario, we are dealing with ensemble data consisting of hundreds of realizations. The unique analysis tasks addressed in this work present a significant challenge to visualization scalability, as the number of biclusters can be on the order of the square of the typical number of clusters. In our visual analytics system, this will mainly influence the visual space of histograms in the bicluster matrix and the small multiple view. In our visual analytics system, this will mainly influence the visual space of histograms in the bicluster matrix and the small multiple view. In the system, we show 5 × 5 cells in matrix and five realization subspaces in small multiple view, and they do not suffer from overcrowding on a typical 1080p monitor. In addition, other techniques including lenses and LoD can be applied to further alleviate the pressure on visual space. Therefore, we think our visualization can fulfill most analysis requirements with visual analytics workflow.

8 Conclusion and Future Works

In this work, we present a biclustering-based approach for the coupled analysis of realizations and locations in ensemble data. With our system, we can study the coordinated behaviors between locations and realizations, and analyze how they vary across different subspaces. Our case studies have successfully demonstrated several findings that existing tools have difficulty or are unable to discover.

In the future, we would like to improve our approach in two aspects. Firstly, our approach can be further extended to other types of ensemble data, such as time-varying scalar fields, flow fields, in order to explore their coordinated features. This expansion requires addressing the challenge of defining appropriate transformations for such ensemble fields, which is a complex yet worthwhile endeavor. Secondly, we intend to establish connections between subspace exploration and the parameter space of simulation models to investigate their relationships.

Acknowledgments

This work is supported by NSFC No. 62272012.

References

[1] T. Athawale and A. Entezari. Uncertainty quantification in linear interpolation for isosurface extraction. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2723–2732, 2013.

[2] T. Athawale, E. Sakhaee, and A. Entezari. Isosurface visualization of data with nonparametric models for uncertainty. *IEEE Trans. Vis. Comput. Graph.*, 22(1):777–786, 2016.

[3] U. Bordoloi, D. L. Kao, and H. Shen. Visualization techniques for spatial probability density function data. *Data Science Journal*, 3:153–162, 2004.

[4] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[5] C. V. F. de Souza, P. d. C. L. Barcellos, L. Crissaff, M. Cataldi, F. Miranda, and M. Lage. Visualizing simulation ensembles of extreme weather events. *Computers & Graphics*, 104:162–172, 2022.

[6] I. Demir, C. Dick, and R. Westermann. Multi-charts for comparative 3d ensemble visualization. *IEEE Trans. Vis. Comput. Graph.*, 20(12):2694–2703, 2014.

[7] F. Ferstl, K. Bürger, and R. Westermann. Streamline variability plots for characterizing the uncertainty in vector field ensembles. *IEEE Trans. Vis. Comput. Graph.*, 22(1):767–776, 2016.

[8] F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann. Visual analysis of spatial variability and global correlations in ensembles of iso-contours. *Comput. Graph. Forum*, 35(3):221–230, 2016.

[9] F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann. Time-hierarchical clustering and visualization of weather forecast ensembles. *IEEE Trans. Vis. Comput. Graph.*, 23(1):831–840, 2017.

[10] P. Fiaux, M. Sun, L. Bradel, C. North, N. Ramakrishnan, and A. Endert. Bixplorer: Visual analytics with biclusters. *IEEE Computer*, 46(8):90–94, 2013.

[11] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. T. Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. *IEEE Trans. Vis. Comput. Graph.*, 19(10):1646–1663, 2013.

[12] H. Guo, X. Yuan, J. Huang, and X. Zhu. Coupled ensemble flow line advection and analysis. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2733–2742, 2013.

[13] L. Hao, C. G. Healey, and S. A. Bass. Effective visualization of temporal ensembles. *IEEE Trans. Vis. Comput. Graph.*, 22(1):787–796, 2016.

[14] J. A. Hartigan. Direct clustering of a data matrix. *Journal of the american statistical association*, 67(337):123–129, 1972.

[15] M. Hlawatsch, P. C. Leube, W. Nowak, and D. Weiskopf. Flow radar glyphs - static visualization of unsteady flow with uncertainty. *IEEE Trans. Vis. Comput. Graph.*, 17(12):1949–1958, 2011.

[16] B. E. Hollister and A. Pang. Bivariate quantile interpolation for ensemble derived probability density estimates. *International Journal for Uncertainty Quantification*, 5(2):123–137, 2015.

[17] B. E. Hollister and A. Pang. Visual analysis of transport similarity in 2D CFD ensembles. *Electronic Imaging*, 2016(1):1–11, 2016.

[18] M. Hummel, H. Obermaier, C. Garth, and K. I. Joy. Comparative visual analysis of Lagrangian transport in CFD ensembles. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2743–2752, 2013.

[19] D. Jäckle, F. Fischer, T. Schreck, and D. A. Keim. Temporal MDS plots for analysis of multivariate data. *IEEE Trans. Vis. Comput. Graph.*, 22(1):141–150, 2016.

[20] M. Jarema, I. Demir, J. Kehrer, and R. Westermann. Comparative visual analysis of vector field ensembles. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 81–88. IEEE, 2015.

[21] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Trans. Vis. Comput. Graph.*, 19(3):495–513, 2013.

[22] R. Liu, H. Guo, J. Zhang, and X. Yuan. Comparative visualization of vector field ensembles based on longest common subsequence. In *2016 IEEE Pacific Visualization Symposium (PacificVis)*, pages 96–103. IEEE, 2016.

[23] S. K. Lodha, A. Pang, R. E. Sheehan, and C. M. Wittenbrink. UFLOW: visualizing uncertainty in fluid flow. In *Proc. of IEEE Visualization*, pages 249–254, 1996.

[24] A. L. Love, A. Pang, and D. L. Kao. Visualizing spatial multivalue data. *IEEE Computer Graphics and Applications*, 25(3):69–79, 2005.

[25] A. Luo, D. Kao, and A. Pang. Visualizing spatial distribution data sets. In *VisSym*, volume 3, pages 29–38, 2003.

[26] S. C. Madeira and A. L. Oliveira. Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 1(1):24–45, 2004.

[27] M. Mirzargar, R. T. Whitaker, and R. M. Kirby. Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE Trans. Vis. Comput. Graph.*, 20(12):2654–2663, 2014.

[28] H. Obermaier and K. I. Joy. Future challenges for ensemble visualization. *IEEE Computer Graphics and Applications*, 34(3):8–11, 2014.

[29] M. Otto, T. Germer, H. Hege, and H. Theisel. Uncertain 2D vector field topology. *Comput. Graph. Forum*, 29(2):347–356, 2010.

[30] M. Otto, T. Germer, and H. Theisel. Uncertain topology of 3d vector fields. In *2011 IEEE Pacific Visualization Symposium*, pages 67–74. IEEE, 2011.

[31] A. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, 1997.

[32] C. Petz, K. Pöthkow, and H. Hege. Probabilistic local features in uncertain vector fields with spatial correlation. *Comput. Graph. Forum*, 31(3):1045–1054, 2012.

[33] T. Pfaffelmoser, M. Reitinger, and R. Westermann. Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. *Comput. Graph. Forum*, 30(3):951–960, 2011.

[34] T. Pfaffelmoser and R. Westermann. Correlation visualization for structural uncertainty analysis. *International Journal for Uncertainty Quantification*, 3(2), 2013.

[35] K. Pöthkow and H. Hege. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Trans. Vis. Comput. Graph.*, 17(10):1393–1406, 2011.

[36] K. Pöthkow, B. Weber, and H. Hege. Probabilistic marching cubes. *Comput. Graph. Forum*, 30(3):931–940, 2011.

[37] K. Potter, M. Kirby, D. Xiu, and C. R. Johnson. Interactive visualization of probability and cumulative density functions. *International journal for uncertainty quantification*, 2(4):397–412, 2012.

[38] K. Potter, J. Kniss, R. F. Riesenfeld, and C. R. Johnson. Visualizing summary statistics and uncertainty. *Comput. Graph. Forum*, 29(3):823–832, 2010.

[39] K. Potter, A. Wilson, P.-T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. R. Johnson. Ensemble-vis: A framework for the statistical visualization of ensemble data. In *2009 IEEE international conference on data mining workshops*, pages 233–240. IEEE, 2009.

[40] R. Santamaría, R. Therón, and L. Quintales. Bicoverlapper: A tool for bicluster visualization. *Bioinformatics*, 24(9):1212–1213, 2008.

[41] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. J. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Trans. Vis. Comput. Graph.*, 16(6):1421–1430, 2010.

[42] M. Sun, L. Bradel, C. L. North, and N. Ramakrishnan. The role of interactive biclusters in sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1559–1562, 2014.

[43] M. Sun, C. North, and N. Ramakrishnan. A five-level design framework for bicluster visualizations. *IEEE Trans. Vis. Comput. Graph.*, 20(12):1713–1722, 2014.

[44] M. Sun, A. R. Shaikh, H. Alhoori, and J. Zhao. Sightbi: Exploring cross-view data relationships with biclusters. *IEEE Trans. Vis. Comput. Graph.*, 28(1):54–64, 2022.

[45] A. Tatu, F. Maaß, I. Färber, E. Bertini, T. Schreck, T. Seidl, and D. Keim. Subspace search and visualization to make sense of alternative clusterings in high-dimensional data. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 63–72, Oct 2012.

[46] D. Thompson, J. A. Levine, J. C. Bennett, P.-T. Bremer, A. Gyulassy, V. Pascucci, and P. P. Pébay. Analysis of large-scale scalar data using hixels. In *2011 IEEE Symposium on Large Data Analysis and Visualization*, pages 23–30. IEEE, 2011.

[47] V. Verma and A. Pang. Comparative flow visualization. *IEEE Trans. Vis. Comput. Graph.*, 10(6):609–624, 2004.

[48] K. Watanabe, H.-Y. Wu, Y. Niibe, S. Takahashi, and I. Fujishiro. Biclustering multivariate data for correlated subspace mining. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pages 287–294. IEEE, 2015.

[49] K. Watanabe, H.-Y. Wu, Y. Niibe, S. Takahashi, and I. Fujishiro. Biclustering multivariate data for correlated subspace mining. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pages 287–294. IEEE, 2015.

[50] K. Watanabe, H.-Y. Wu, Y. Niibe, S. Takahashi, and I. Fujishiro. Biclustering multivariate data for correlated subspace mining. In *Proc. of IEEE Pacific Visualization Symposium*, pages 287–294, 2015.

[51] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2713–2722, 2013.

[52] C. M. Wittenbrink, A. Pang, and S. K. Lodha. Glyphs for visualizing uncertainty in vector fields. *IEEE Trans. Vis. Comput. Graph.*, 2(3):266–279, 1996.

[53] H.-Y. Wu, Y. Niibe, K. Watanabe, S. Takahashi, M. Uemura, and I. Fujishiro. Making many-to-many parallel coordinate plots scalable by asymmetric biclustering. In *Proc. of IEEE Pacific Visualization Symposium*, pages 305–309, 2017.

[54] W. Wu, J. Xu, H. Zeng, Y. Zheng, H. Qu, B. Ni, M. Yuan, and L. M. Ni. Telcovis: Visual exploration of co-occurrence in urban human mobility based on telco data. *IEEE Trans. Vis. Comput. Graph.*, 22(1):935–944, 2016.

[55] X. Yuan, D. Ren, Z. Wang, and C. Guo. Dimension projection matrix/tree: Interactive subspace visual exploration and analysis of high dimensional data. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2625–2633, 2013.

[56] M. Zhang, Q. Li, L. Chen, X. Yuan, and J. Yong. Enconvis: A unified framework for ensemble contour visualization. *IEEE Trans. Vis. Comput. Graph.*, 29(4):2067–2079, 2023.

[57] J. Zhao, M. Sun, F. Chen, and P. Chiu. Bidots: Visual exploration of weighted biclusters. *IEEE Trans. Vis. Comput. Graph.*, 24(1):195–204, 2018.