

A Probability Density-based Visual Analytics Approach to Forecast Bias Calibration

Renpei Huang, Quan Li, Li Chen, and Xiaoru Yuan

Abstract—Biases inevitably occur in numerical weather prediction (NWP) due to an idealized numerical assumption for modeling chaotic atmospheric systems. Therefore, the rapid and accurate identification and calibration of biases is crucial for NWP in weather forecasting. Conventional approaches, such as various analog post-processing forecast methods, have been designed to aid in bias calibration. However, these approaches fail to consider the spatiotemporal correlations of forecast bias, which can considerably affect calibration efficacy. In this work, we propose a novel bias pattern extraction approach based on forecasting-observation probability density by merging historical forecasting and observation datasets. Given a spatiotemporal scope, our approach extracts and fuses bias patterns and automatically divides regions with similar bias patterns. Termed *BicaVis*, our spatiotemporal bias pattern visual analytics system is proposed to assist experts in drafting calibration curves on the basis of these bias patterns. To verify the effectiveness of our approach, we conduct two case studies with real-world reanalysis datasets. The feedback collected from domain experts confirms the efficacy of our approach.

Index Terms—Weather forecast, pattern extraction, calibration, visual analytics

1 INTRODUCTION

As an important method in numerical weather prediction (NWP) for weather forecasting, ensemble forecasting has been widely applied in many disciplines due to the rapid improvement in modern computing power. Specifically, ensemble simulation forecasting is performed multiple times by using various models with different initial values, parameters, and hypotheses to include real-world distribution spaces as much as possible. Thus, the future probability density distribution of atmospheric state can be predicted quantitatively. However, the prediction accuracy of future weather is considerably affected by chaotic weather systems in which a small input error and idealized assumptions of numerical simulations can result in distinct outputs over time. The phenomenon, which is conventionally identified as a bias [1], [2], occurs when consistent differences arise between actual outcomes and previously generated forecasts of those quantities, i.e., forecasts may have a general tendency to be too high or too low. Biases may also follow certain spatiotemporal distributions. The characteristics of biases in a certain spatiotemporal interval are called bias patterns.

Bias calibration is a major issue in weather forecasting and has many important social implications. The quality of bias calibration directly affects the accuracy of the weather forecast, and the quality of meteorological services. Domain experts have proposed several automated methods, including statistics- [3], probability distribution- [4], and similarity-based methods [5]. Statistics-based methods, such as the bias-corrected relative frequency (BRF) [6] and fre-

quency matching [7], suggest a calibration value by modeling bias on the basis of the statistical characteristics of historical data. However, this kind of method cannot accurately model complex weather systems because they are only based on macroscopic characteristics. Thus, the calibration results cannot integrate specific weather parameters. Similarity-based methods, such as analog postprocessing forecast methods [8], generate calibration values through a few historical forecasts of a specific grid point in a similar atmospheric state, thereby serving a good recommendation for the calibration. However, similarity detection based on individual points ignores the potential spatiotemporal correlation among the grid points.

Although automatic bias calibration has demonstrated promising performance in forecast calibration [9], it often fails because of the complicated correlation of the atmospheric states across different timestamps and spatial locations. To address this problem, domain experts are often required to participate in the forecast process when finalizing bias calibration. Specifically, experts often need to analyze the reliability of the forecast results and increase the accuracy of weather forecasting through manual calibration. Thus, this approach considerably relies on the expertise of domain experts. However, the calibration results of forecasts can vary from person to person and are often accompanied by uncertainties, such as the different qualification levels and working status of the domain experts. Therefore, long-term and high-quality calibration that is solely based on experts' working experience is relatively difficult to maintain. The use of visualization and visual analytics to support the exploration of bias patterns and facilitate automated methods with human knowledge has been introduced for bias calibration [10], [11], [12]. However, two major challenges have emerged. The first is the representation of bias patterns. Bias patterns are a statistical abstraction of historical data as a mathematical model at a certain spatiotemporal in-

• Renpei Huang and Li Chen are with School of Software, BNRIST, Tsinghua University, Beijing, P. R. China.
• Quan Li is with AI Group, WeBank, Shenzhen, Guangdong, P. R. China.
• Xiaoru Yuan is with Key Laboratory of Machine Perception (Ministry of Education), and National Engineering Laboratory for Big Data Analysis and Application, Peking University.
• Li Chen is the corresponding author.

terval, which requires an effective representation to describe the abstract bias pattern concept. The second challenge is the scope detection of bias patterns. Bias patterns can only be established when the spatiotemporal interval has been determined. However, no ground-truth or theoretical basis is available to assist domain experts in finding an appropriate interval.

In this study, we first propose a probability density-based approach to model the bias pattern. Specifically, we convert time-separated reanalysis data (forecast simulation results with the corresponding observation data) on each spatial grid point into a forecast-observation probability density function (PDF). Subsequently, we aggregate the spatial grid points hierarchically on the basis of the similarity between the bias patterns of the grids. Thus, the bias pattern can be identified on the basis of the spatiotemporal similarity to simulate the process of obtaining the calibration experience from the reanalysis data. On the basis of the proposed pattern extraction approach, a novel visual analytics system called *BicaVis* is proposed with coordinated views and intuitive interactions to assist domain experts in exploring the bias pattern and calibrating weather forecast. Particularly, this system provides a transition detection of bias patterns across different time windows encoded by a scatter plot and an interactive semiautomatic calibration visualization based on bias patterns that cross time windows. *BicaVis* combines analysis, visualization, and interactions to assist forecasters in discovering bias patterns in reanalysis data. The primary contributions of our work are summarized as follows:

- We integrate a probability density-based method with visual interventions to convey the bias between simulation and real data and depict the spatiotemporal trends and data distribution of bias patterns for visual detection and comparative analysis.
- We construct a visual analytics system for bias patterns; the system has multiple interactive views to explore and compare the bias patterns of the historical data in a top-down manner and enables a “human-in-loop” process to draft the calibration curve on the basis of the identified bias patterns.
- We apply our system to weather forecast calibration through a collaboration with meteorologists. Case studies on real-world data verify the efficacy of our approach.

2 RELATED WORK

Studies that overlap with this work can be classified into three categories: multidimensional meteorological data visualization, uncertainty visualization for meteorological data, and calibration in meteorological forecast.

2.1 Multidimensional Meteorological Data Visualization

Visualization has played an important role in the field of meteorology for several decades [13], [14]. In addition, many tools, such as Ferret [15], GMT [16], Vis-5D [17], and Metview [18], have been developed to provide functions for meteorological data integration, reconstruction, and visualization. Rautenhaus et al. [19] presented Met.3D to facilitate the real-time visualization of meteorological ensemble

members by using advanced computer graphic techniques. Helbig et al. [20] proposed an interactive application termed MEVA for visualizing heterogeneous data with multiple 3D devices. Wang et al. [21] designed a feature-based comparison framework with coordinated views to support the comparison of the rain band generated by the Gaussian mixture model from the observation and forecast datasets. Liao et al. [10] developed a visual analytics system based on a novel voting framework to assist forecasters during the calibration of grid points. Dasgupta et al. [22] developed a multiway visual comparison approach based on various visual cues for detecting detailed differences while maximizing the accuracy of comparisons in climate models. Doraiswamy et al. [23] presented a visual exploration framework for cloud motion; the framework hierarchically summarizes cloud systems without prior knowledge, thereby assisting in identifying and detecting atmospheric data. On the basis of mutual information and Pearson’s chi-square test, Bernard et al. [24] proposed a visual analysis method for high-dimensional scalar meteorological datasets to show and compare the relationship of a subset of the datasets. These studies have helped domain experts understand large-scale meteorological data and have assisted them in their specific domain tasks. However, research that focuses on the calibration of weather forecasts and extraction, as well as the exploration of bias patterns, is limited. In this work, we focus on the extraction of bias patterns and explore the spatiotemporal relationship among these patterns.

2.2 Uncertainty Visualization for Meteorological Data

Uncertainty visualization is an important but challenging topic in data visualization [25]. In meteorological research, uncertainty is usually estimated on the basis of the means and variance of the meteorological data. Pöthkow et al. [26], [27] presented a *Probabilistic Marching Cubes* method based on the level-crossing probability field to visualize the spatial uncertainty of isocontours. Mirzargar et al. [28] proposed a nonparametric method called curve box plot to summarize 2D and 3D curves in the ensemble data, thereby extending a method from the descriptive statistics and data depth to the curves. Whitaker et al. [29] introduced contour box plots, which are a generalization of box plots, to support the visualization and exploration of ensembles of contours or level sets of functions. Ferstl et al. [30] described variability plots to visualize the statistical properties of pathlines or isocontours in ensemble data; they also applied the plots to flow fields, multidimensional scalar fields [31], and time-varying data [32] to explore different data trends. Pfaffelmoser et al. [33], [34] presented a set of circular glyphs and a color diffusion scheme to visualize the variability of gradients in uncertain 2D scalar fields. Potter et al. [35] proposed a framework called *EnsembleVis* to demonstrate the uncertainties of numerical weather ensemble data with multiple linked views. Sanyal et al. [36] designed *Noodles*, which is an interactive view based on ribbon and circular glyphs, to visualize the uncertainty of numerical weather ensemble data. Wang et al. [37] presented nested parallel coordinate plots, which integrate intraset parameter visualization and interset parameter comparison in a single view with diverse user interactions, to visualize and analyze the

high-dimensional convective parameter space in multiresolution climate ensemble data.

Despite these existing approaches, methods that explore the consistency between numerical weather ensembles and observation data remain limited; such limitation is a major concern of domain experts in weather forecasting. In this work, we initially generate bias patterns from the reanalysis data. We then sum them up and visualize the uncertainty of the bias between forecast results and the observation data through the historical dataset.

2.3 Calibration in Meteorological Forecast

Calibration refers to the identification and correction of potential bias in the NWP products; this process is conducted by forecasters in accordance with their work experience and the automated bias calibration method. Many approaches have been proposed to improve the calibration accuracy and quality prior to publishing weather forecast results. For example, Glahn et al. [3] applied model output statistics to calibrate forecast results through linear regression. Wilks et al. [38] compared the logistic regression method [39], the Gaussian set mixing method [40], and the nonhomogeneous Gaussian regression method [41] and improved these three methods by using the cumulative PDF of 25-year reanalysis data. The BRF method [6] was presented to calibrate the forecast values by aligning it with the observed cumulative distribution function (CDF) and its inverse function. Raftery et al. [42] combined the forecast distributions of different sources on the basis of the Bayesian model and then performed calibration by using weighting algorithms. Gneiting et al. [4] proposed an evaluation method for predictive performance on the basis of the sharpness model, which maximizes the calibrated forecast distribution. Hamill et al. [8] invented a method called *Analog*, which is based on the retrieval of similar weather conditions in historical data, to assist in the calibration of future weather forecasts. The above methods have improved calibration results. However, an enhanced calibration result can be obtained further by an experienced forecaster even without the use of any automated calibration methods. This finding implies that the combination of forecasters' experience and automated calibration methods can be a promising approach to improve the effect of calibration further. In this work, we initially use a divide-and-conquer algorithm to generate the calibration curve and optimize the segmented domain of the line with minimal loss of the original information. Then, the domain experts can interactively adjust the calibration curve in accordance with their expertise.

3 BACKGROUND AND OBSERVATIONAL STUDY

To understand how weather forecast bias is analyzed in practice, we worked with two experts, including one weather forecaster (E.1) and one forecast analyst (E.2). E.1 browsed isoline maps drawn by other numerical forecast products of the meteorological bureau prior to reporting the weather forecast daily. Then, he made a subjective judgment regarding the accuracy of these numerical forecast products by combining his own domain knowledge and the current weather conditions of the destination location. On the basis

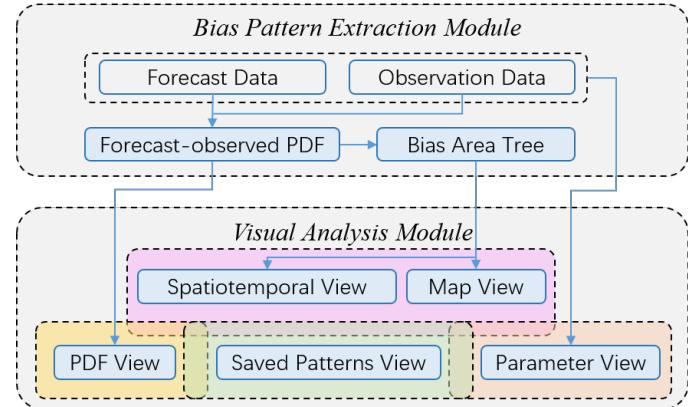


Fig. 1. Overview of *BicaVis*. The bias pattern extraction module abstracts the forecast-observation PDF from the reanalysis data and clusters the PDF into a bias area tree for further visualization. The visual analysis module visualizes the result of each phase of the extraction model and supports an interactive analysis for experts to analyze the data.

of his judgment, he adjusted the weighted mean value of each forecast product and determined the final released weather forecast result. He also referred to some automated algorithms for the analysis of the accuracy of each numerical forecast product to assist him in calibrating the result. E.2 mainly analyzed the error between the prediction and actual results. Specifically, he analyzed the weather characteristics when errors occurred and the prediction calibration techniques behind such errors. General analysis methods included leveraging the *threat score* (TS) and the *equitable TS* (ETS) [43] to evaluate the recent performance or special weather conditions of different forecast calibration products. He then analyzed the guidance conclusions with various meteorological conditions, forecast product characteristics, pros and cons, and applicable scope based on the scores.

We interviewed E.1 and E.2 in two separate sessions to identify their primary concerns in weather forecast bias analysis and thus ensure that the ontological structure of our approach fits well into the domain tasks. These interviews revealed the need for a highly intuitive, interactive, and customized visual analytics system for grounding the historical ensemble forecast data analysis. Despite the differences in their expectations for this system, certain requirements were expressed across the board.

R.1 Extract bias patterns. As indicated in the previous section, bias patterns are a kind of statistical abstraction of the historical data in a certain spatiotemporal interval. However, methods that aim to extract the potential bias patterns from the reanalysis data are limited.

R.2 Visualize the spatiotemporal distribution of similar bias patterns. E.2 hoped to have a visualization that can well demonstrate the bias patterns in the spatial and temporal dimensions for an overview, which can help him immediately identify the basic properties and distribution of the bias patterns present in the data.

R.3 Reveal the differences and correlations among different spatiotemporal bias patterns. E.2 commented that the current bias analysis methods only focus on a single location, i.e., calibrating precipitation at a single station instead of analyzing its correlation among its neighbors.

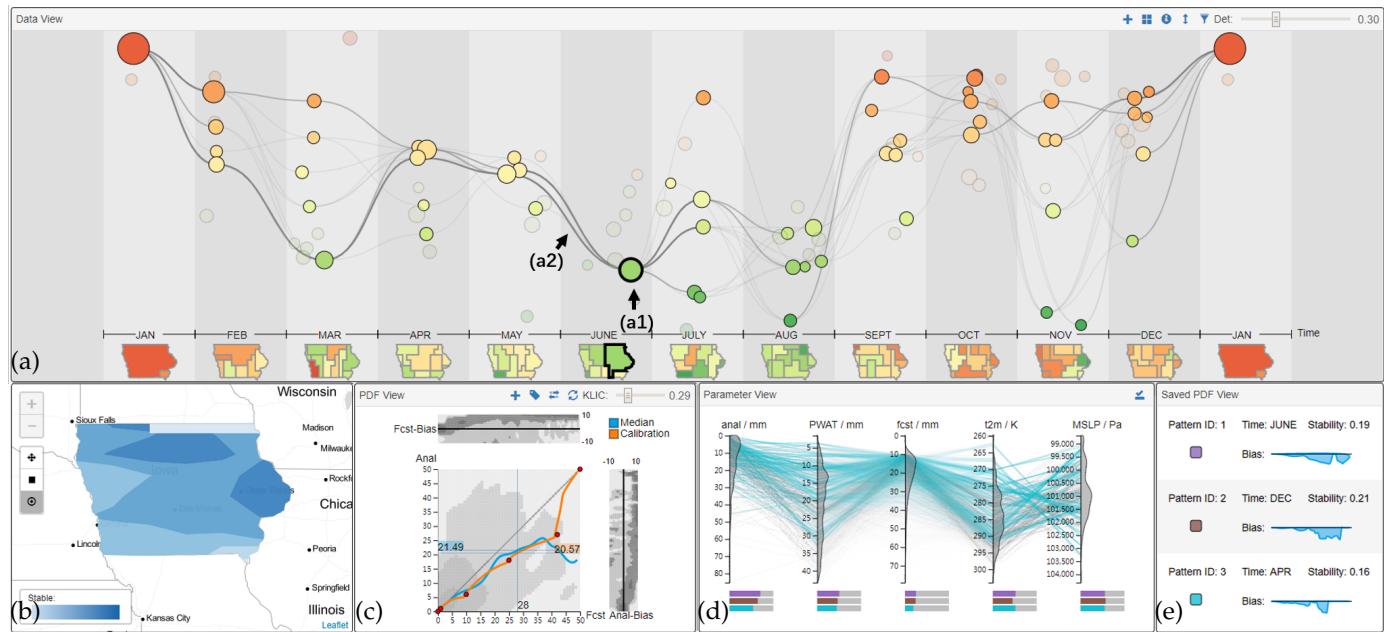


Fig. 2. Interface of *BicaVis*. (a) The spatiotemporal bias pattern view visualizes the spatiotemporal distribution of the bias patterns extracted from the reanalysis data. Each circle (a1) represents an area with a similar bias pattern and the flow (a2) between two circles indicates the shared grid points by the two bias pattern areas. (b) The map view shows an isocontours plot of the stability of each grid point in the space. (c) The forecast-observation PDF view provides an interactive semiautomatic workflow in generating the bias calibration curve. (d) The parameter comparison view shows the overview of the atmospheric parameters of the selected areas through a parallel coordinate plot, which enables the comparison among areas. (e) The saved bias pattern view displays the basic information of the selected bias pattern of interest.

Understanding the stability and uniqueness of the bias patterns can help explore and make informed decisions. Thus, E.2 wanted to observe the difference and correlations among different spatiotemporal bias patterns.

R.4 Select bias patterns of interest. E.1 needed to filter out the bias patterns of interest from the data overview rapidly for a detailed analysis. Thus, a visual interactive mechanism and visual cues should be provided to assist him in selecting the bias patterns of interest.

R.5 Facilitate parameter comparison. To understand the similarities and differences among bias patterns, comparing the atmospheric parameters when generating the corresponding bias patterns is essential to understand the underlying causes of the bias patterns and to further support making informed decisions.

R.6 Support detailed calibrations. According to E.1, the accuracy of weather forecasts largely depends on the subjective judgment of the forecaster in terms of his/her qualification, status, and working conditions, thereby increasing the unreliability of the calibrated results. Therefore, an interactive calibration mechanism is desired to support the combination of domain experts' experience and bias patterns extracted from historical data; in this manner, the accuracy of forecast calibration is improved.

4 APPROACH OVERVIEW

Three tasks were formalized on the basis of the requirements stated by the domain experts. The tasks are as follows: bias pattern extraction (R.1), visualization of the spatiotemporal distribution of bias patterns (R.2, R.3), and detailed visualization of bias patterns (R.4, R.5, R.6). Specifically, we trans-

lated the bias pattern concept into a concrete probability density map and proposed a clustering approach to extract the underlying patterns in the historical data (R.1). Then, we proposed a spatiotemporal bias pattern visual analytics system called *BicaVis* with multiple interactive views to visualize the extracted bias patterns in the reanalysis dataset. As shown in Fig.1, our system consists of two modules, namely, the bias pattern extraction module and the visual analysis module.

In the bias pattern extraction module, given a forecast result in a spatiotemporal interval and the corresponding observation data, the forecast-observation probability density function (PDF) at each grid is initially generated via a kernel density estimation (KDE) (R.1). The distribution field can represent the bias pattern at the grid in a given time range because the probability density distribution of the observation-forecast values can reflect the distribution characteristics of the forecast and the observation data in the current time window. Furthermore, the probability density distribution can derive the distribution characteristics of the bias, which is equal to the difference between the observation and the forecast data. We used a bottom-up hierarchical clustering algorithm to aggregate a single grid point into a connected area on the basis of the similarity of the PDF among the grids. The result of the hierarchical clustering can be considered a clustering tree of the connected areas with similar bias patterns, which are subsequently fed into the following visual analysis module for further exploration.

In the visual analysis module, several coordinated views and intuitive interactions are provided to analyze the bias patterns. Specifically, we proposed a relative position and

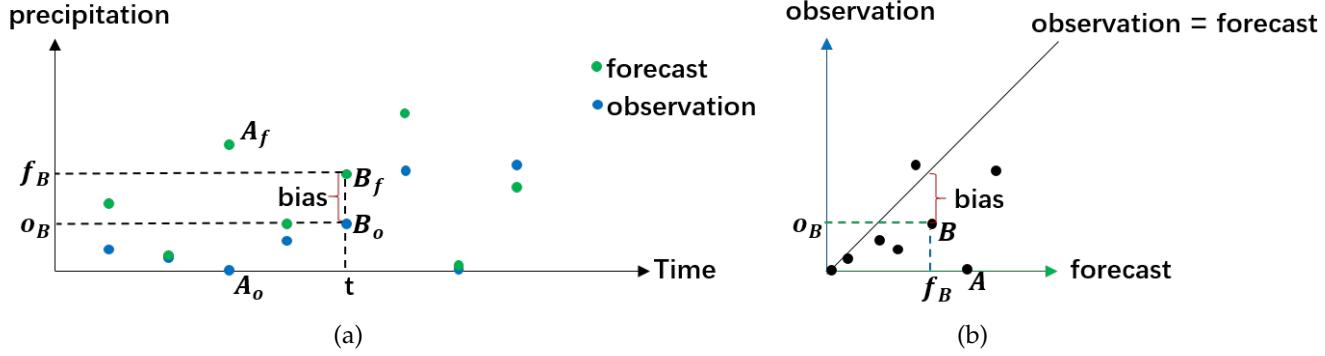


Fig. 3. These data consist of the forecast and observation values at each time step and each grid point. (a) shows the distribution of the data, and (b) shows the reshaped sampling points. The forecast-observation data pair at the same timestamp and grid point is considered a sample point in the forecast-observation domain. For example, the point pair (A_f, A_o) in (a) is transformed to the point A in (b), and (B_f, B_o) in (a) is transformed to the point B .

similarity-encoded scatter plot (Fig. 2(a)) to visualize the spatiotemporal distribution of the bias patterns (R.2) and used a flow metaphor to represent the transition of bias patterns over time (R.3). The users can filter the bias patterns of interest for detailed exploration in Fig. 2(b). Particularly, a parallel coordinate plot (Fig. 2(d)) was used to support the comparison of the atmospheric parameters in different bias patterns (R.5). An interactive semi-automatic calibration curve based on the detailed visualization of a specific bias pattern (Fig. 2(c)) was designed to assist the forecast calibration (R.6).

5 BIAS PATTERN EXTRACTION MODULE

In this section, we first describe the forecast-observation PDF, which transforms the abstract bias pattern into a configurable statistical distribution. Then, we discuss the extraction of the regions with similar bias patterns from the forecast-observation PDF.

5.1 Forecast-Observation Probability Distribution

As discussed in Section 4, bias patterns refer to the generalization of the bias between the forecast and the observation result. We used the forecast-observation PDF to characterize bias patterns. A PDF (denoted as $f_X(x)(X \in D)$) refers to the probability that the value of the random variable X is near a certain point x , where the value domain of X , x is D . The value domain of the forecast-observation PDF $\mathcal{P}_X(x)$ is the domain of all vectors of the form (F_i, O_i) , where F_i and O_i are the results of the i^{th} forecast and the i^{th} observation in the time window T , respectively. The forecast-observation PDF can indicate the joint distribution of prediction and observation results and summarize their distribution characteristics.

Map the original data into the forecast-observation space to formalize the bias pattern. After the bilinear interpolation of the grid in the observation dataset with that in the forecast dataset, each weather forecast result can be formally defined in the time window T on the grid point P as a point pair (t_i, f_i) , (t_i, o_i) , where t_i is the time of the occurrence of i^{th} forecast and i^{th} observation. Fig. 3(a) shows the distribution of the forecast and observation point during the time window T . The green and blue points

represent the forecast and the observation results, respectively. Each of these points share the same position on the time axis. Given that the calibration of weather forecasts based on multiyear historical forecast data is a long-term process, a time window generally spans one or several months. The order of multiple forecasts within a single time window can be ignored statistically. Therefore, the forecast-observation result on the grid point P in the time window T can be considered a set $S_P^T = \{(f_i, o_i) | (f_i, o_i) \in U\}$ of sampling points (as Fig. 3(b)) within a probability field $U_P^T = F_P^T \times O_P^T$, where F and O are the domains of the forecast and the observation result, respectively.

Obtain the statistical definition of the bias pattern. Formally, for any forecast result f_i and its observation result o_i , a unique corresponding $b(f_i, o_i) = f_i - o_i$ is available, and the probability of generating the forecast result f_i and the observation result o_i is determined uniquely by $\mathcal{P}_X(f_i, o_i)$. Meanwhile, the domain of the bias $B_P^T(F_P^T, O_P^T) = F \times O$ is the same as the probability field U_P^T . Thus, $\mathcal{P}_X(f_i, o_i)$ can fully describe the distribution of $b(f_i, o_i)$, that is, the bias pattern. Lastly, the conceptual bias patterns can be converted into concrete PDF $\mathcal{P}_U(f_i, o_i)$.

Each historical weather forecast in the time window corresponds to a sampling of the bias pattern of the current grid point. The KDE method is used to generate the forecast-observation PDF to restore the bias distribution pattern as realistically as possible on the basis of the sampling results. KDE is a nonparametric method of PDF estimation. In contrast to parameter estimation, nonparametric estimation does not add any prior knowledge; nevertheless, it fits the distribution according to the characteristics and properties of the data itself, thereby obtaining a more suitable model than the parameter estimation method. For a given grid point P and time window T , $\mathcal{P}_U(s)$ is estimated on the basis of the following function:

$$\mathcal{P}_U(s) = \frac{1}{|S_P^T|} \sum_{s_i \in S_P^T} K\left(\frac{s - s_i}{h}\right) \quad (1)$$

where $K(x)$ is a kernel function (we use the Gaussian kernel), and h is the bandwidth.

The KDE-fitted forecast-observation PDF fully indicates the bias pattern of grid points and restores the probability

distribution of the forecasts and bias. This approach lays the foundation for the system to extract the regions with similar bias patterns and helps forecasters analyze and understand the bias pattern represented by the PDF for further decision-making.

5.2 Extraction of Areas with Similar Bias Patterns

The forecast-observation PDF on a single grid contains the temporal characteristics of the forecast bias but ignores the spatial influence of the relationship between this bias and other grid points in its neighborhood. The forecast result in the neighborhood of a grid point is also a reference for calibrating its forecast result. For example, a group of grids on one side of a mountain may share similar forecast results in a monsoon climate. If the forecast at one grid point is remarkably higher than the others, then calibrating the forecast value with that of the others is reasonable. Therefore, extracting the bias pattern within a spatial domain is necessary for calibration.

The forecast-observation PDF $\mathcal{P}_U(f_i, o_i)$ can characterize bias patterns; thus, we iteratively merge the two regions with the most similar \mathcal{P}_U to extract regions with similar bias patterns. Our method extracts the areas by using the forecast-observation PDF of a single variable. Bias $b(f_i, o_i) = f_i - o_i$ is a function defined in the space $U = F \times O$; thus, the data involve two dimensions, i.e., the forecast dimension F and the observed dimension O of the single variable.

Confirm the feasibility and time complexity of area extraction. A probability density field [44] $\mathcal{F}(p)$ is generated by estimating the PDF on all grid points in the spatial range; this field is called the forecast-observation probability density field. This approach assigns each point in the given spatial range P with a forecast-observation PDF $\mathcal{P}_U(s)$. For given points $P_1, P_2 \in P$, the forecast-observation PDF $\mathcal{P}_{U_{\{P_1, P_2\}}}(s)$, which is defined on the region $\{P_1, P_2\}$, is considered. Given that $U_{P_1} = U_{P_2} = U_P = U_{\{P_1, P_2\}}$, the forecast-observation PDF can be expressed as follows:

$$\begin{aligned} \mathcal{P}_{U_{\{P_1, P_2\}}}(s) &= \frac{1}{|S_{P_1}^T| + |S_{P_2}^T|} \sum_{s_i \in S_{P_1}^T \cup S_{P_2}^T} K\left(\frac{s - s_i}{h}\right) \\ &= \frac{|S_{P_1}^T|}{|S_{P_1}^T| + |S_{P_2}^T|} \mathcal{P}_{U_{P_1}}(s) + \frac{|S_{P_2}^T|}{|S_{P_1}^T| + |S_{P_2}^T|} \mathcal{P}_{U_{P_2}}(s) \end{aligned} \quad (2)$$

The forecast-observation PDF of P_1, P_2 is a linear expression with the PDF of two grids. The forecast-observation PDF of any region can be generated by Equation (2). Therefore, the region extraction of similar bias patterns and the generation of the forecast-observation PDF of the region can be performed within an acceptable period.

Ensure a measurable similarity of bias patterns. Two areas are considered to have similar bias patterns when their probability of bias occurrence is similar in the probability field B_P^T . This condition indicates that the similarity of their forecast-observation PDF can measure the similarity of bias patterns. The computation method for the similarity between two distributions at different grid points is Jensen-Shannon divergence (JS divergence), which is an improved

method based on Kullback-Leibler divergence (KL divergence). This method measures the similarity of two possibility distributions and overcomes the disadvantage of KL divergence (i.e., it does not satisfy the triangle inequality). Therefore, the two most similar areas (i.e., the two areas with the smallest JS divergence for clustering) are combined for each iteration.

Extract regions with similar bias patterns. The regions are aggregated by the bottom-up hierarchical clustering algorithm. Without specifying the number of clusters, the algorithm can recursively merge the two most similar adjacent regions in each iteration via the distance matrix. The grid points are organized into a binary tree through hierarchical clustering, where each sub-tree represents a similar bias pattern area. In addition, the tree logs the extraction. Therefore, the clustering result can be dynamically adjusted through the hyperparameters set via user interaction.

Represent similar bias pattern areas. After the forecast-observation PDF-based hierarchical clustering of each grid point, a cluster tree, namely, the bias area tree, is constructed to represent similar bias pattern areas. The difference among different branches of the same subtree decreases as the depth increases. The bias area tree must be compressed, and the branches with low representativeness must be merged to construct additional representative clusters of the similar bias pattern area. At the same depth, when the representativeness of a node is low, the similarity among its child nodes is high. The breadth-first search of the bias area tree starts from the root. For each nonleaf node, if the similarity between its child nodes is lower than a threshold T_d with depth attenuation, then the nonleaf node is deleted and replaced with its child nodes. The threshold attenuated by depth can contribute to the uniqueness of the reserved node from its siblings and to the similarity to its parent. The threshold is defined as follows: $T_d = \frac{1}{\alpha^d}$, where α is a hyperparameter, and d is the depth of the node. If the similarity is higher than the threshold T_d , then the node is retained, and the breadth-first search is continued for its child nodes. Compressing the clustering tree saves query and storage costs while preserving areas of representative similarity bias patterns.

6 VISUAL ANALYSIS MODULE

The basic design principle behind *BicaVis* is to leverage or augment familiar visual metaphors and thus enable domain experts to focus on the analysis [45]. We strictly followed the mantra “*overview first, zoom and filter, then details-on-demand*” [46]. On the basis of these principles and the preceding requirements mentioned, we developed four major visualizations that help domain experts who explore the weather forecast ensemble data. Specifically, we designed a spatiotemporal bias pattern view to obtain an overview of the areas with similar bias patterns, a map view to visualize the spatial distribution of similar bias pattern areas and grid points, a forecast-observation PDF view to analyze the bias pattern in detail, and a parameter comparison view to analyze the correlation between the distribution of the data dimensions and the bias.

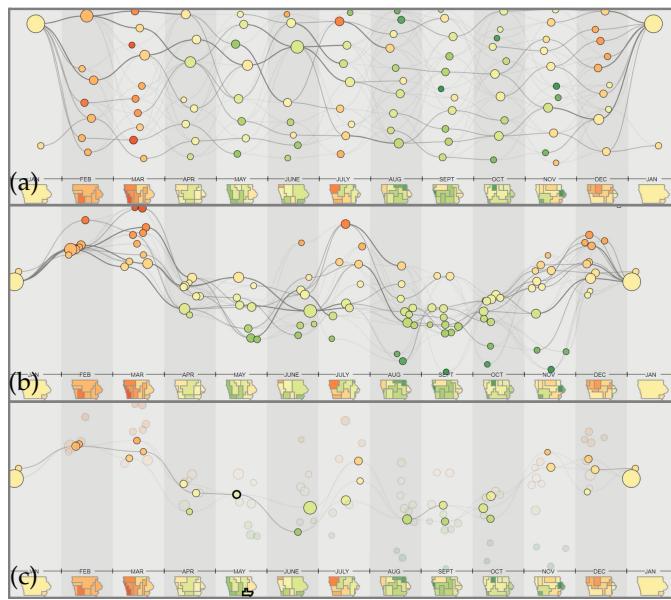


Fig. 4. Spatiotemporal Bias Pattern View. (a) Relative spatial position layout. (b) Similarity layout. (c) Similarity layout with a selected similar bias pattern area.

6.1 Spatiotemporal Bias Pattern View

The spatiotemporal bias pattern view (Fig. 4(a)) provides an overview of the areas with similar bias patterns extracted from the reanalysis data in each time window (R.2). It also visualizes the transition of grid points between areas with similar bias patterns in the adjacent time windows (R.3).

Visual encoding and interaction. A scatter plot-based visualization is used to present a comprehensive overview and transition of the distribution of the bias patterns in spatiotemporal dimensions. In Fig. 4(a), we arrange a set of time windows horizontally in chronological order and set the background color of the adjacent time window to different brightness levels to enhance distinguishability. The circles in each time window represent areas with similar bias patterns in the highest layer of the bias area tree, which corresponds to the same colored areas in the small map view below each time window. The radius of each circle encodes the number of grid points contained in the area, and the color indicates similarity. The normalized value of the similarity is generated by characterizing the PDFs of all similar bias areas into high-dimensional vectors. Then, the dimensions are reduced into a 2D space via using multidimensional scaling (MDS). Such a method can preserve the relative distance between data items in the high-dimensional space. A flow between two circles indicates the shared grid points of the two bias pattern areas. We divide the flow's thickness into three levels according to the amount of the grid points. The users can further filter similar bias pattern areas to explore the bias patterns through interaction further, thereby observing the spatiotemporal characteristics of the bias patterns in conjunction with their domain expertise and interests (R.4). For instance, we can click on a similar bias pattern area, and the corresponding areas in each time window would be highlighted automatically (Fig. 4(c)). We also support similarity-based filtering interaction. The users can directly

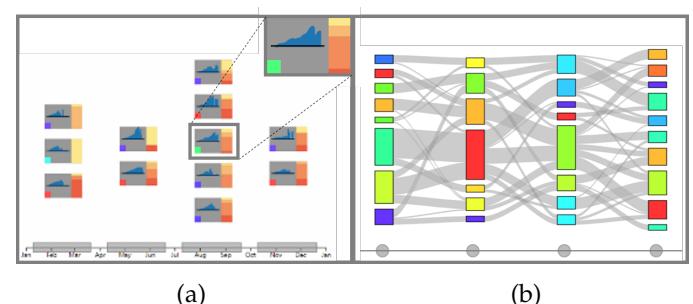


Fig. 5. Design Alternatives. (a) Bias patterns based on a side-by-side label. (b) Bias patterns based on the Sankey graph.

input a similar threshold, and the areas with a similarity higher than the threshold would be filtered.

Two types of layouts, namely, the relative spatial position and similarity layouts, of similar bias pattern areas are provided (Figs. 4(a) and (b)). In the first relative spatial position layout, the x-axis of the center of the circle encodes the relative longitude coordinates in the entire spatial range of the center of the region it represents, and the y-axis is sorted from the top to the bottom on the basis of the relative latitude coordinates to avoid the occlusion of the circles. The relative spatial position layout also avoids occlusion between similar areas and minimizes flow occlusion. In the second similarity layout, the x- and y-axes encode the result of MDS. The dimension reduction results of all bias patterns are shown in the time windows that correspond to each bias pattern. In this layout, the similarity characterized by the relative position can be easily perceived by the analysts, who can remarkably assist in filtering areas with similar bias patterns.

Design alternatives. During the development of *BicavVis*, several design alternatives of the spatiotemporal view are proposed to display the overview of bias patterns. In Fig. 5(a), bias patterns are listed side-by-side in each time window. Each similar bias area is represented by a label similar to that in the upper right of Fig. 5(a). A line chart in the middle shows a simplified bias distribution. The gradient-filled rectangle in the right reveals the inner uniformity of the bias area. The colored bar in the bottom left corner indicates the similarity of the area. Each bar on the time axis below the labels represents a time window. Although this design intuitively expresses the temporal distribution and the internal characteristics of each bias pattern, this design still has many problems, such as (a) the lack of effective representation of spatial relationships between bias patterns, (b) the weak visual connection between labels and the bias patterns shown in other views, (c) excessive details in the overview view that the users ignore. Therefore, this visual design is unacceptable. Fig. 5(b) illustrates another visual design based on the Sankey graph. Instead of using a complex label to encode the detailed information of an area, we use a rectangle to represent the bias area. The height of the rectangle encodes the number of grid points covered by the area, and the color indicates the similarity. The time windows represented by circles lie on the time axis. The flow that connects the bias area in adjacent time windows represents the intersection of grid points of the two regions. The flow width encodes the size of the intersection. This

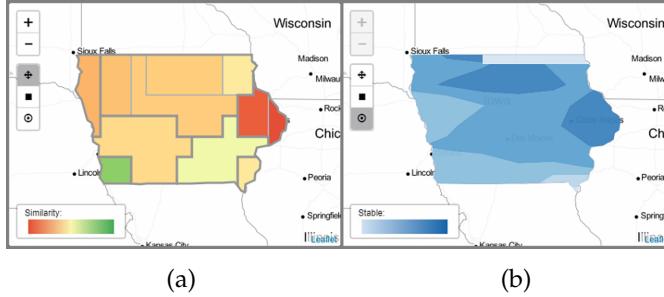


Fig. 6. Map view. (a) Visualization of the space distribution of similar bias pattern areas in the selected time window. The similarity between bias patterns in each area is shown by the closeness of the color filled in each bias pattern area. (b) Visualization of the stability at each grid point. The bias pattern of the grid points in the dark region is closer to the pattern in their neighborhood.

design reflects the spatiotemporal connection between the bias areas and simplifies unnecessary details. However, this design was not adopted because of (a) a lack of sufficient visual elements to establish correspondences between rectangle and areas, (b) visual occlusion caused by numerous translucent flow overlaps, and (c) visual confusion exacerbated by the disorderly arrangement of the areas.

6.2 Map View

The map view was designed to visualize the spatial distribution of the similar bias pattern areas and grid points (R.2); it also corresponds to the interaction conducted in the spatiotemporal bias pattern view. Specifically, a contour map is used in the map view to convey the stability of each grid point in the space (Fig. 4(b)) (R.3). The stability $S(p)$ of a grid point p describes its capability to share a similar bias pattern as other grid points in the neighborhood over all time windows; it is defined as follows:

$$S(p) = \frac{1}{|T| \times (|P| - 1)} \sum_{t \in T, q \in P, q \neq p} \delta_t(p, q), \quad (3)$$

$$\delta_t(p, q) = \begin{cases} 1, & A(p) = A(q) \\ 0, & A(p) \neq A(q) \end{cases}$$

where T is the set of all time windows, P is the set of all grid points, and $A(p)$ is the area of the grid point p . Analysts can efficiently understand the spatial similarity of the grid points through the stability contour map view.

6.3 Forecast-Observation PDF View

The forecast-observation PDF view provides detailed information on the bias pattern and assists analysts in drafting the calibration curve on the basis of the bias patterns (R.5). In Fig. 7(a), the forecast-observation PDF view comprises a line chart, a probability density map, and two bias density maps. The x-axis of the chart presents the forecast value, and the y-axis conveys the observation value. The probability density map behind the lines (Fig. 7(a)-I) shows the forecast-observation PDF of the selected bias pattern in five gray levels. The bias density map (Fig. 7(a)-II) displays the distribution of the bias at each forecast value, and the map on the right (Fig. 7(a)-III) shows the distribution at each observation value. The blue line (Fig. 7(a)-IV) visualizes the

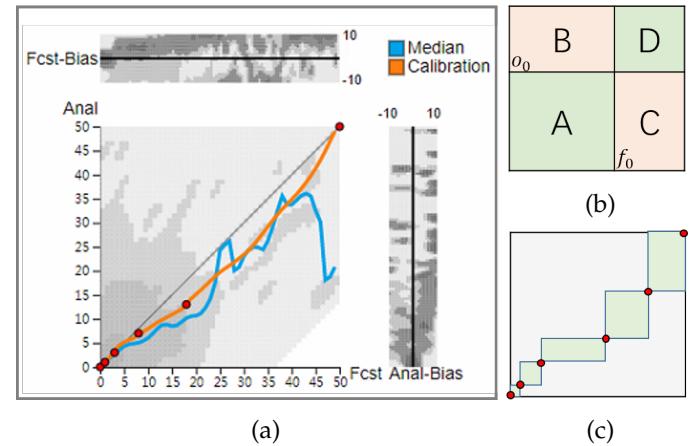


Fig. 7. Forecast-Observation PDF View and the Divide-and-Conquer BRF algorithm. (a) Forecast-observation PDF view. (b) The point (x_1, y_1) splits the domain into four segments. (c) A series of segmentation points divides the entire forecast-observation domain into a series of valid intervals.

median of the observation value for every given forecast value, and the orange line (Fig. 7(a)-V) denotes the calibration curve based on the bias patterns.

The BRF method can automatically generate the calibration value for a given forecast result. A forecast value f is corrected to a calibrated value c by $c = \mathcal{C}_O^{-1}(\mathcal{C}_F(f))$, where \mathcal{C}_F is the CDF of the forecast value in historical data, and \mathcal{C}_O is the CDF of the observation value. The idea of the divide-and-conquer algorithm is combined to extend the method into the forecast-observation domain $F \times O$ and thus generate the bias calibration curve. The calibration curve $C(f)$ is a function where $C(f)$ is the calibrated value obtained by the BRF method for every forecast value $f \in F$ based on $\mathcal{C}_{F=f} = \int_{x \in O} \mathcal{P}_X(f, x) dx$. The fitted PDF produces a probability at every $o \in O$, and the historical data are distributed in a subset of O . This inequality of the definition domain could lead to errors. A value function $W(f, o)$ is designed to define the domain of $\mathcal{C}_{F=f}$ accurately as follows:

$$W(f_0, o_0) = \frac{P(A) + P(D)}{1 - \ln(1 - |P(A) - P(D)|)}, \quad (4)$$

$$P(A) = P(0 < f < f_0, 0 < o < o_0),$$

$$P(D) = P(f_0 < f < f_{\max}, o_0 < o < o_{\max})$$

where f_{\max} and o_{\max} are the maximum values of F and O , respectively. The segment point $P(f, o)$ separates the domain into four parts (Fig. 7(b)). The value function W estimates the ability to refine the domain of the segmentation generated by P . The numerator of W measures the proportion of the historical data contained in segmentations A and D . The denominator measures the uniformity of data that fall within A and D . When the data are uniform, the segmentation is accurate. After separating the domain by the point $P_0(f_0, o_0)$ that maximizes W , the segmentations A_0 and D_0 are uniform and contain most of the historical data, whereas B_0 and C_0 are segmentations that contain almost no historical data. When CDF is only calculated in the definition domain, A_0 and D_0 can retain the highest

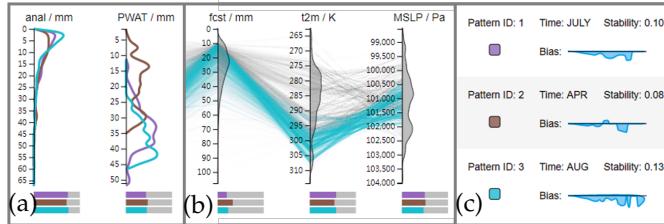


Fig. 8. Parameter Comparison View. Users can switch between (b) PCP plots and (a) line graphs to obtain an overview of each data in the selected area or compare the difference between areas. The saved bias patterns view (c) lists the basic information of the selected areas. Each information panel consists of five parts: the pattern ID, the time window name, the average stability, the representing color in parameter view and the simplified bias distribution.

amount of historical information and remove the useless domain as much as possible, respectively.

On the basis of the divide-and-conquer concept, a group of segment points can be generated by recursively searching the point that maximizes W in the regions A and D (Fig. 7(c)). The BRF method calculates a calibration value for each forecast value in accordance with the CDF generated between adjacent segment points. The calibration curve is generated by all forecast values on the defined domain F and their corresponding calibration values.

However, separation in subareas where the data are evenly and densely distributed results in oversegmentation. The number of segment points is interactively set by the information gain threshold. When the information gain of the new segment point is below the threshold, further iterations are stopped. The red points in Fig. 7(a) correspond to the control points of the calibration curve. The analyst can move and add control points to the appropriate location with reference to other visualization elements and their own knowledge. The calibration curve is immediately recalculated on the basis of the new control point (as segment point). The analyst can then perform weather forecast calibration and make further decisions based on the calibration curve obtained from the interaction.

6.4 Parameter Comparison View

The interviews revealed that the experts consider the correlation between the distribution of these atmospheric parameters (variables) and the forecast deviation (bias) more important than the correlation between these atmospheric parameters. Therefore, the design of the parameter comparison view focused on the correlation between the distribution of the data dimensions and the bias, thus enabling the analysis and comparison of atmospheric parameters generated by the forecast ensemble (R.5). That is, it visualizes the data distribution in the five dimensions in a bias pattern and the correlation between the data distribution and the bias of the accumulated precipitation instead of visualizing the correlation among these dimensions.

A parallel coordinate plot (Fig. 8(b)) presents the overview of each data in the selected area; the plot visualizes the parameter distribution of the selected bias patterns. Each axis of the PCP plot represents a parameter of the

atmospheric state. The gray area on the axes visualizes the distribution of the corresponding parameters in all saved bias patterns. The correlation coefficient between the distribution of the atmospheric and bias parameters is depicted in the rectangular bar below each axis. Each row represents an area of the same color. The length of the colored bar represents the correlation coefficient between the corresponding atmospheric and bias parameters. The users can switch between PCP plots and line graphs through interaction to enable further comparison. The line chart (Fig. 8(a)) shows the distribution of atmospheric parameters in each saved pattern. In Fig. 8(c), the users can observe the basic information and properties, including pattern ID, timestamp, average stability, and simplified bias distribution, of a saved bias pattern of interest. We leverage a line chart to show the simplified bias distribution. The axis represents the forecast domain, and the bias curve, which briefly conveys the bias pattern, encodes the average bias at each forecast value in the bias pattern area.

6.5 Interactions among the Views

In addition to the most defining capabilities of *BicaVis*, rich interactions were integrated to catalyze an efficient in-depth analysis. (1) *Linking and Highlighting*. After the users select a pattern by clicking on a pattern node, all the patterns that are associated with the selected one are highlighted, thereby emphasizing their spatiotemporal relationship. (2) *Filtering*. Bias patterns can be further filtered by enabling a similarity slider that controls the similarity threshold. During the filtering interaction, the users can focus on the patterns that are closely related and similar to the selected ones. (3) *Dragging and Dropping*. In the detailed exploration of the patterns, several control points are generated to assist the users in drafting the calibration curve. These control points enable the users to change the segmentation results by dragging and dropping to fine-tune the calibration line.

7 EVALUATION

In this section, we verify the efficacy of *BicaVis* through two cases with real-world reanalysis data. The first case demonstrates the common workflow of our system to assist domain experts in the weather forecast calibration. The second case illustrates how our system supports abnormal bias pattern detection and analyzes the underlying causes.

7.1 Dataset Description

The data used in the cases comprise two types of datasets, namely, the forecast and observation data. The **forecast data** are produced by the Global Ensemble Forecast System (GEFS) global ensemble from the National Centers for Environmental Prediction in the United States. The dataset consists of 11 ensemble members at different convection parameters. Each member has approximately 8000 grids. The ensemble runs and generates forecast atmospheric parameters daily from 1985, covering an area of $[25.0436^\circ N, 53.1299^\circ N][66.094^\circ W, 125.625^\circ W]$ (latitude-longitude) with a resolution of 0.5° . The data have approximately 150 million forecast records (forecast average value of four output parameters during a 10-year

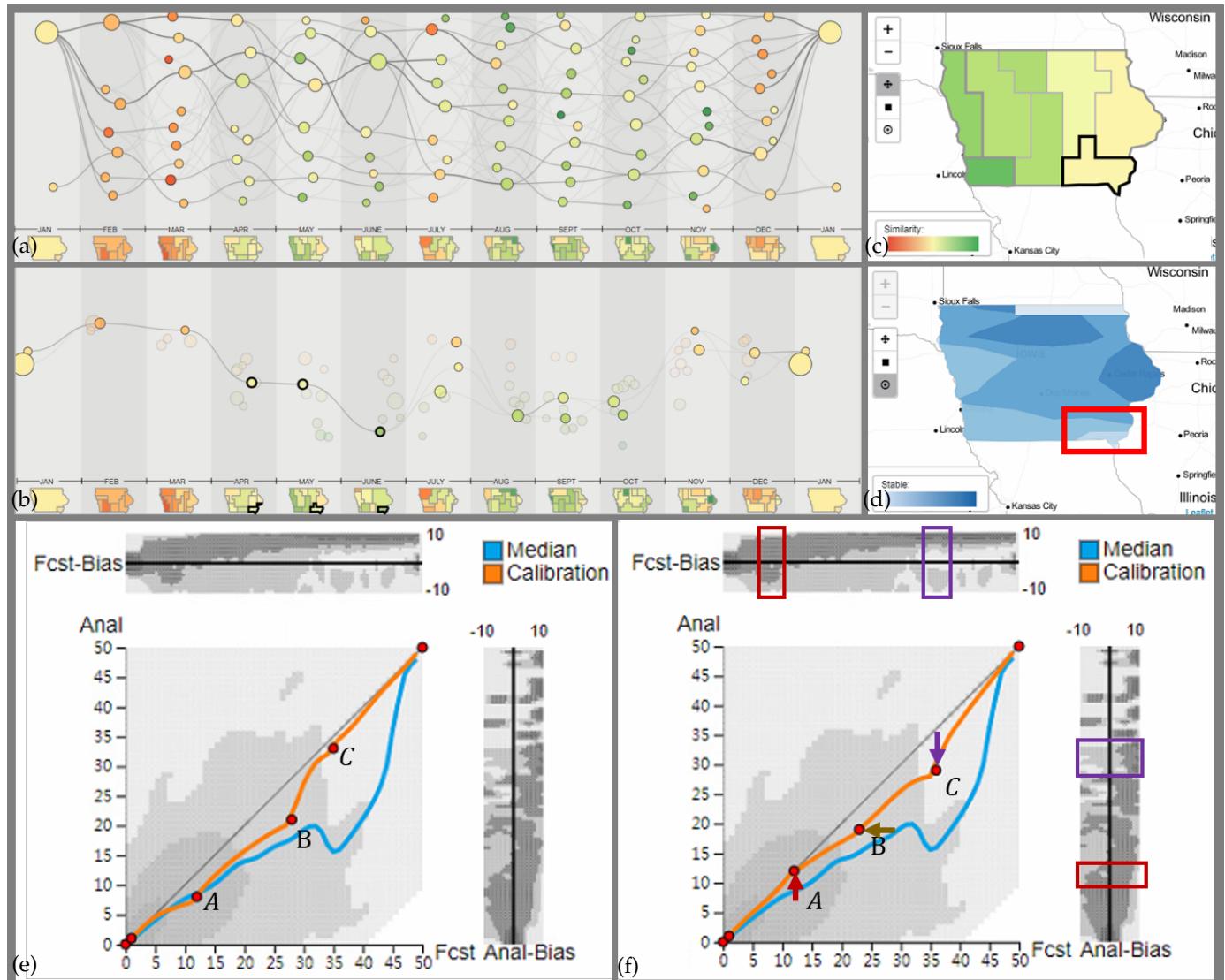


Fig. 9. Case one shows the exploration of similar bias patterns and the adjustment of the calibration curve. (a) Overview of all bias area in Iowa in the relative spatial position layout of spatiotemporal view. (b) Bias areas used to generate the joint forecast-observation PDF after filtering by the experts. (c) Spatial distribution of bias area in the time window "May" in the map view. (d) Stability of the grid points in the time window "May". (e) Joint forecast-observation PDF with detailed bias information and the automatically generated calibration curve in the forecast-observation PDF view. (f) Control points adjusted by the experts on the basis of their domain knowledge and the detailed information.

period). Additional details of the GEFS forecast dataset can be found in [5]. The **observation data** are produced by the climatology-calibrated precipitation analysis (CCPA), which records the observed precipitation in the same area throughout the US every six hours from 2002 with a resolution of 0.125° . These data are considered the ground-truth state of the actual weather state. The size of the data used is approximately 30 million observation records (observed precipitation at each interpolated grid point during a 10-years period). Additional details of the CCPA observation dataset can be found in [47].

We conducted a case study to explore the bias pattern of the total accumulated precipitation in 12 time windows that represent each month. We followed the work of Hamill et al. [6] and adopted five variables, namely, forecast-accumulated precipitation (FCST), observed accumulated precipitation (ANAL), two-meter temperature

(T2M), precipitable water (PWAT), and mean sea-level pressure (MSLP). The data used in our cases included the precipitation reanalysis data from the year of 2002 to 2013.

7.2 Case One: Exploring Similar Bias Patterns and Adjusting the Calibration Curve

The first case aims to demonstrate how *BicaVis* assists E.1 in exploring similar bias patterns and adjusting the calibration curve. After initializing the data, an overview of the similar bias pattern area in Iowa is displayed in the spatiotemporal bias pattern view (Fig. 9(a)). The weather forecast to be calibrated falls in the time window of "May". Thus, E.1 directly selected the time window of "May" by clicking on the text "May" laying on the time axis name. The distribution of the grid points with similar bias patterns in "May" throughout the area is shown in the map view (Fig. 9(c)). From the

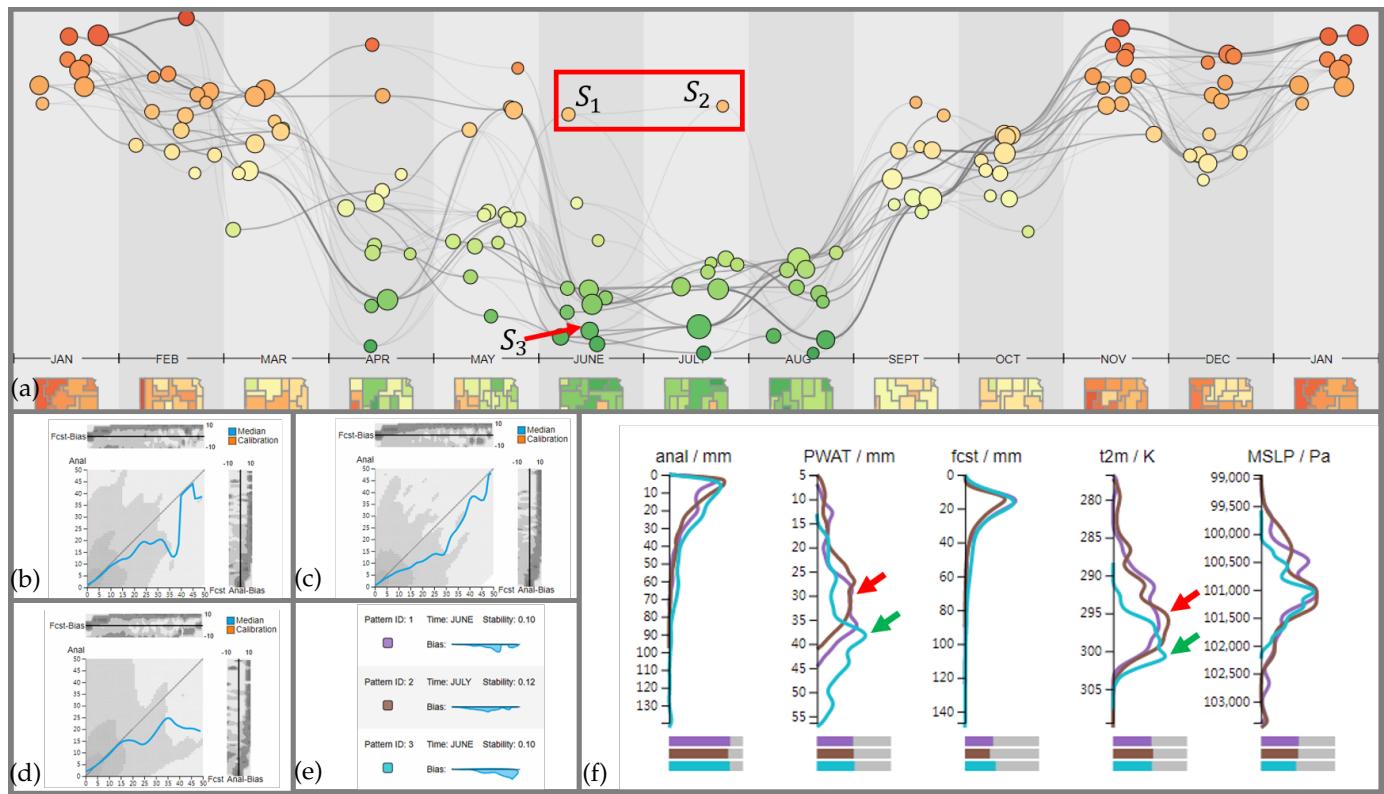


Fig. 10. Case two shows the process of outlier bias pattern detection. (a) Overview of all bias area in Kansas in the similarity layout of the spatiotemporal view. (b)(c)(d) Forecast-observation PDF of three selected areas \$S_1\$, \$S_2\$ and \$S_3\$. (e) Basic information in the saved view. (f) Distribution of atmospheric parameters and the correlation coefficient between them and the distribution of the bias parameters in the parameter comparison view. The red and green arrows indicate the difference between \$S_3\$ and the other two areas, respectively.

map view, E.1 observed that the southeastern part of Iowa is grouped into a single area with a similar bias pattern (the area with a bold outline), which is called \$S_5\$.

Plotting the calibration curve on the basis of the bias pattern within a single time window may result in the loss of temporal characteristics. Thus, considering the similarity of bias patterns on the same grid across other time windows is necessary to determine the cross-time window bias patterns for the calibration curve. Then, E.1 selected \$S_5\$, and the spatiotemporal view automatically filtered the areas in each time window that shares intersection grid points with \$S_5\$ and showed the transition of the grid points in \$S_5\$ across different time windows. A small similarity threshold was set to filter the bias pattern in the remaining time windows because of the low stability of \$S_5\$ (indicated by the red rectangle in Fig. 9(d)). Fig. 9(b) shows the filtered results with only a few similar bias pattern areas that intersect with \$S_5\$ retained; The results included “April” (\$S_4\$) and “June” (\$S_6\$), and they were selected because they are similar and close to \$S_5\$.

Subsequently, our system generated the cross-time window bias pattern on the basis of the selected areas \$S_4\$, \$S_5\$, and \$S_6\$. The detailed information of the bias pattern is illustrated in the forecast-observation PDF view (Fig. 9(e)). Specifically, the orange curve presents the preset calibration curve generated in accordance with the algorithm discussed in Section 6.3 with an information gain threshold of 0.24 set by E.1. Then, E.1 witnessed that the curve has six red

control points that represent the interval boundary for the BRF method and control the shape of the preset calibration curve by dragging the control points. In Fig. 9(f), the forecast and observation biases near control point A (indicated in the red rectangle) are evenly distributed on both sides. Control point A should be closer to the ideal prediction (i.e., the forecast is equal to the observation value and is illustrated as a straight line of 45 degrees in the forecast-observation PDF view). Thus, E.1 moved to point A, as indicated by the red arrow. E.1 also observed that the bias distribution near control point C is concentrated at +10 mm. Thus, control point C must be moved to a position 10 mm below the ideal prediction (i.e., the movement track of control point C is indicated by the purple arrow). E.1 further observed that the bias near control point B presents a different distribution between the forecast and observed values given that the bias distribution on the specific predicted value is inconsistent with the overall bias distribution. In Fig. 9(f), the final calibration curve, which is generated by a semi-automatic process combining interaction and calibration algorithms is illustrated. For the calibration of the target weather forecast, E.1 can obtain the corresponding calibration value on the calibration curve by moving the cursor to the original predicted value that corresponds to the \$x\$-axis.

7.3 Case Two: Outlier Bias Pattern Detection

The second case aims to explore the abnormal bias patterns in Kansas. The similarity layout of the spatiotemporal bias

pattern view is selected to visualize the similarity distribution of bias patterns in Kansas. In Fig. 10(a), two patterns, namely, S_1 and S_2 (indicated by red rectangles), appear above the time window of "June" and "July", and other bias patterns are located below the time window, thereby indicating that S_1 and S_2 are similar to and different from other patterns, respectively. Therefore, S_1 and S_2 are abnormal patterns that deserve further analysis.

According to E.2, in the rainy season, precipitation forecasts often have a large bias from the observations. S_1 and S_2 , which are different from other conventional bias patterns, tend to present the characteristics of the rainy season precipitation forecast bias. Therefore, three patterns, namely, S_1 , S_2 , and S_3 , are selected for further comparison. Figs. 10 (b), (c), and (d) visualize the forecast-observation PDF of the three patterns. Fig. 10(e) shows the basic information about the three patterns. The comparison of the median curve in the forecast-observation PDF view indicates that the forecast bias of the precipitation in S_1 and S_2 is small when the forecast rainfall is within the range of [0, 10 mm] or [40 mm, 50 mm]; moreover, the forecast bias is large when the precipitation is within [20 mm, 30 mm]. The bias in S_3 increases with the forecast value.

To identify the reason for the small bias of S_1 and S_2 in forecasting large precipitation, E.2 compared the atmospheric parameters in S_1 , S_2 and S_3 through the parameter comparison view. The correlation coefficient matrix visualized in Fig. 10(f) indicates that the parameters with high correlation to the bias distributions are T2M, ANAL, and PWAT. The detailed atmospheric parameter distribution is also shown in Fig. 10(f). The T2M in S_3 is concentrated at 300 K, and the distributions of S_1 and S_2 are concentrated at 295 K. In addition, through the PWAT distribution illustrated in Fig. 10(f), the value in S_3 is concentrated at 40 mm, and the values in S_1 and S_2 are evenly distributed on [30 mm, 40 mm]. Therefore, E.2 can infer that the forecast bias of the large precipitation generated at T2M = 300 K is smaller than that at T2M = 295 K, and the bias pattern in which PWAT evenly distributes within [30 mm, 40 mm] causes less bias in forecasting large precipitation.

8 DISCUSSION AND LIMITATION

We conducted a half-hour semistructured interview with the experts (E.1 and E.2) to evaluate our system and check whether our approach facilitates bias calibration in the weather forecasting.

System usability. All experts appreciated the ability of our system to support the interactive extraction and exploration of the bias patterns in the historical forecast data. E.1 stated that "*Unlike previous rigid statistical calibration method, BicaVis provides an intent-based visual exploration to detect bias patterns of interest*". BicaVis enables the human-in-the-loop calibration generation process, which can "*significantly decrease the error of the calibration curve line*" by combining the algorithm performance with the user domain expertise in making complex decisions.

Algorithm complexity. In the similar bias pattern area extraction algorithm, the bias patterns on each grid point are clustered into a similar bias pattern area via hierarchical clustering and JS divergence. We also attempted to use other

similarity measure methods and clustering algorithms such as Euclidean distance and K-means clustering. Although the performance of hierarchical clustering and JS divergence is slightly inferior in terms of computational efficiency, their results can reflect the composition of similar bias pattern areas better than other approaches. In addition, the hierarchical clustering does not require an exact number of clusters, and the resulting clustering tree can be compressed flexibly.

Visual design. The experts were satisfied with the visual designs and interactions. We deliberately selected familiar visual metaphors (e.g., isocontours plots, curves, and map) to help experts immediately familiarize themselves with our system. After introducing the basic views and functions of our system, the experts developed a customized path when using the system for exploration purposes. E.1 prefers the process of generating the calibration curve line, whereas E.2 appreciates the clear presentation of the spatiotemporal relationship between the bias patterns shown in the spatiotemporal bias pattern view.

Generalizability. We discussed with the experts which components of *BicaVis* can be directly deployed to other scenarios and which ones require further customization. We found that the system can also be applied to the calibration of other scalar field data, such as the temperature data. In addition, the KDE method can estimate the forecast-observation PDF of any scale in a continuous domain. Thus, this algorithm can support the extraction of bias patterns based on the PDF estimated from historical forecast and observation results with the same resolution.

Scalability. In *BicaVis*, for data with a different scope, adjusting the number of the displayed bias patterns is necessary to achieve a remarkable visualization result. Specifically, the hierarchical clustering algorithm regulates the clustering accuracy by adjusting the hyperparameter α , thereby affecting the resolution of the bias pattern area shown in the spatiotemporal view. Therefore, unacceptable operation delays may occur when the data size is relatively large because of the algorithm complexity of the hierarchical clustering algorithm.

Limitation. First, generalizing the bias pattern for all historical data cannot eliminate the interference caused by extreme weather. The low frequency and high inaccuracy of the forecast bias under some extreme weather conditions may become noise data in the bias pattern under general weather conditions. Second, currently, we only focused on five selected variables by following T. M. Hamill et al.'s practice [6] in conducting the causal analysis. However, given that the weather is a chaotic system, bias causal analysis can be complicated. Additional data and dimensions must be considered in our future work. Third, we only worked with two experts in the evaluation. Therefore, we could not provide a quantitative assessment of the system.

9 CONCLUSION AND FUTURE WORK

In this study, we present an effective approach to explore and analyze bias patterns in reanalysis data. We initially describe bias patterns as forecast-observation PDFs and aggregate them to a bias area in accordance with their similarities. Subsequently, a novel visual analytics system

called *BicaVis* is proposed to assist forecasters in calibrating the weather forecast and summarizing the bias distribution patterns. Multiple coordinated views with intuitive interactions are provided to support top-down spatiotemporal bias pattern exploration and an interactive semi-automatic calibration curve generation. Two case studies demonstrate the efficacy of our approach to bias pattern extraction and weather forecast calibration. In the future, we plan to design a flexible similar bias pattern area partitioning and an effective calibration curve control point recommendation method. We will also investigate joint distributions and associated patterns to analyze atmospheric state parameters. In addition, we plan to extend our work to explore and visualize vector bias patterns.

ACKNOWLEDGMENTS

We are grateful for the valuable feedback and comments provided by the anonymous reviewers. This research is partially supported by the National Natural Science Foundation of China (Grant Nos. 61972221, 61572274) and NNW2018-ZT6B12 (National Numerical Windtunnel project).

REFERENCES

- [1] T. Lim, "Rationality and analysts' forecast bias," *The journal of Finance*, vol. 56, no. 1, pp. 369–385, 2001.
- [2] D. P. Dee and A. M. Da Silva, "Data assimilation in the presence of forecast bias," *Quarterly Journal of the Royal Meteorological Society*, vol. 124, no. 545, pp. 269–295, 1998.
- [3] H. R. Glahn and D. A. Lowry, "The use of model output statistics (mos) in objective weather forecasting," *Journal of applied meteorology*, vol. 11, no. 8, pp. 1203–1211, 1972.
- [4] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.
- [5] T. M. Hamill, G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau Jr., Y. Zhu, and W. Lapenta, "Noaa's second-generation global medium-range ensemble reforecast dataset," *Bulletin of the American Meteorological Society*, vol. 94, no. 10, pp. 1553–1565, 2013.
- [6] T. M. Hamill and J. S. Whitaker, "Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application," *Monthly Weather Review*, vol. 134, no. 11, pp. 3209–3229, 2006.
- [7] E. E. Ebert, "Ability of a poor man's ensemble to predict the probability and distribution of precipitation," *Monthly Weather Review*, vol. 129, no. 10, pp. 2461–2480, 2001.
- [8] T. M. Hamill, M. Scheuerer, and G. T. Bates, "Analog probabilistic precipitation forecasts using gefs reforecasts and climatology-calibrated precipitation analyses," *Monthly Weather Review*, vol. 143, no. 8, pp. 3300–3309, 2015.
- [9] T. M. Hamill, E. Engle, D. Myrick, M. Peroutka, C. Finan, and M. Scheuerer, "The us national blend of models for statistical postprocessing of probability of precipitation and deterministic precipitation amount," *Monthly Weather Review*, vol. 145, no. 9, pp. 3441–3463, 2017.
- [10] H. Liao, Y. Wu, L. Chen, T. M. Hamill, Y. Wang, K. Dai, H. Zhang, and W. Chen, "A visual voting framework for weather forecast calibration," in *2015 IEEE scientific visualization conference (SciVis)*. IEEE, 2015, pp. 25–32.
- [11] C. Gong, L. Chen, and Z. Zhu, "A visualization system for calibrating multimodel ensembles in weather forecast," *Journal of Visualization*, vol. 19, no. 4, pp. 769–782, 2016.
- [12] H. Liao, Y. Wu, L. Chen, and W. Chen, "Cluster-based visual abstraction for multivariate scatterplots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 9, pp. 2531–2545, 2018.
- [13] M. Rautenhaus, M. Böttger, S. Siemen, R. Hoffman, R. M. Kirby, M. Mirzargar, N. Röber, and R. Westermann, "Visualization in meteorology—a survey of techniques and tools for data analysis tasks," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 12, pp. 3268–3296, 2018.
- [14] L. A. Treinish, "Visualization of scattered meteorological data," *IEEE Computer Graphics and Applications*, vol. 15, no. 4, pp. 20–26, 1995.
- [15] S. Hankin, D. Harrison, J. Osborne, J. Davison, and K. O'BRIEN, "A strategy and a tool, ferret, for closely integrated visualization and analysis," *The Journal of Visualization and Computer Animation*, vol. 7, no. 3, pp. 149–157, 1996.
- [16] P. Wessel, W. H. Smith, R. Scharroo, J. Luis, and F. Wobbe, "Generic mapping tools: improved version released," *Eos, Transactions American Geophysical Union*, vol. 94, no. 45, pp. 409–410, 2013.
- [17] B. Hibbard and D. Santek, "The vis-5d system for easy interactive visualization," in *Proceedings of the First IEEE Conference on Visualization: Visualization90*. IEEE, 1990, pp. 28–35.
- [18] I. Russell, S. Siemen, F. Ii, S. Kertész, S. Lamy-Thépaut, and V. Karhila, "Metview 4–ecmwf's latest generation meteorological workstation," *ECMWF Newsletter*, vol. 126, pp. 23–27, 2010.
- [19] M. Rautenhaus, M. Kern, A. Schäfer, and R. Westermann, "Three-dimensional visualization of ensemble weather forecasts—part 1: The visualization tool met. 3d (version 1.0)," *Geoscientific Model Development*, pp. 2329–2353, 2015.
- [20] C. Helbig, L. Bilke, H.-S. Bauer, M. Böttger, and O. Kolditz, "Meva—an interactive visualization application for validation of multifaceted meteorological data with multiple 3d devices," *PloS one*, vol. 10, no. 4, p. e0123811, 2015.
- [21] Y. Wang, C. Fan, J. Zhang, T. Niu, S. Zhang, and J. Jiang, "Forecast verification and visualization based on gaussian mixture model co-estimation," in *Computer Graphics Forum*, vol. 34, no. 6. Wiley Online Library, 2015, pp. 99–110.
- [22] A. Dasgupta, H. Wang, N. O'Brien, and S. Burrows, "Separating the wheat from the chaff: Comparative visual cues for transparent diagnostics of competing models," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 1, pp. 1043–1053, 2019.
- [23] H. Doraiswamy, V. Natarajan, and R. S. Nanjundiah, "An exploration framework to identify and track movement of cloud systems," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2896–2905, 2013.
- [24] J. Bernard, M. Steiger, S. Widmer, H. Lücke-Tieke, T. May, and J. Kohlhammer, "Visual-interactive exploration of interesting multivariate relations in mixed research data sets," in *Computer Graphics Forum*, vol. 33, no. 3. Wiley Online Library, 2014, pp. 291–300.
- [25] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz, "Overview and state-of-the-art of uncertainty visualization," in *Scientific Visualization*. Springer, 2014, pp. 3–27.
- [26] K. Pothkow and H.-C. Hege, "Positional uncertainty of iso-contours: Condition analysis and probabilistic measures," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 10, pp. 1393–1406, 2010.
- [27] K. Pothkow, B. Weber, and H.-C. Hege, "Probabilistic marching cubes," in *Computer Graphics Forum*, vol. 30, no. 3. Wiley Online Library, 2011, pp. 931–940.
- [28] M. Mirzargar, R. T. Whitaker, and R. M. Kirby, "Curve boxplot: Generalization of boxplot for ensembles of curves," *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, pp. 2654–2663, 2014.
- [29] R. T. Whitaker, M. Mirzargar, and R. M. Kirby, "Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2713–2722, 2013.
- [30] F. Ferstl, K. Bürger, and R. Westermann, "Streamline variability plots for characterizing the uncertainty in vector field ensembles," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 767–776, 2016.
- [31] F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann, "Visual analysis of spatial variability and global correlations in ensembles of iso-contours," in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 221–230.
- [32] ———, "Time-hierarchical clustering and visualization of weather forecast ensembles," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 831–840, 2017.

- [33] T. Pfaffelmoser, M. Reitinger, and R. Westermann, "Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields," in *Computer Graphics Forum*, vol. 30, no. 3. Wiley Online Library, 2011, pp. 951–960.
- [34] T. Pfaffelmoser, M. Mihai, and R. Westermann, "Visualizing the variability of gradients in uncertain 2d scalar fields," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 11, pp. 1948–1961, 2013.
- [35] 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*. IEEE, 2009, pp. 233–240.
- [36] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead, "Noodles: A tool for visualization of numerical weather model ensemble uncertainty," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1421–1430, 2010.
- [37] J. Wang, X. Liu, H.-W. Shen, and G. Lin, "Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 81–90, 2017.
- [38] D. S. Wilks and T. M. Hamill, "Comparison of ensemble-mos methods using gfs reforecasts," *Monthly weather review*, vol. 135, no. 6, pp. 2379–2390, 2007.
- [39] D. S. Wilks, "Comparison of ensemble-mos methods in the lorenz'96 setting," *Meteorological Applications*, vol. 13, no. 3, pp. 243–256, 2006.
- [40] X. Wang and C. H. Bishop, "A comparison of breeding and ensemble transform kalman filter ensemble forecast schemes," *Journal of the atmospheric sciences*, vol. 60, no. 9, pp. 1140–1158, 2003.
- [41] T. Gneiting, A. E. Raftery, A. H. Westveld III, and T. Goldman, "Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation," *Monthly Weather Review*, vol. 133, no. 5, pp. 1098–1118, 2005.
- [42] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, "Using bayesian model averaging to calibrate forecast ensembles," *Monthly weather review*, vol. 133, no. 5, pp. 1155–1174, 2005.
- [43] R. J. Hogan, C. A. Ferro, I. T. Jolliffe, and D. B. Stephenson, "Equitability revisited: Why the "equitable threat score" is not equitable," *Weather and Forecasting*, vol. 25, no. 2, pp. 710–726, 2010.
- [44] W. He, X. Liu, H.-W. Shen, S. M. Collis, and J. J. Helmus, "Range likelihood tree: A compact and effective representation for visual exploration of uncertain data sets," in *2017 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 2017, pp. 151–160.
- [45] Q. Li, K. S. Njotoprawiro, H. Haleem, Q. Chen, C. Yi, and X. Ma, "Embeddingvis: A visual analytics approach to comparative network embedding inspection," *arXiv preprint arXiv:1808.09074*, 2018.
- [46] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proceedings 1996 IEEE symposium on visual languages*. IEEE, 1996, pp. 336–343.
- [47] D. Hou, M. Charles, Y. Luo, Z. Toth, Y. Zhu, R. Krzysztofowicz, Y. Lin, P. Xie, D.-J. Seo, M. Pena et al., "Climatology-calibrated precipitation analysis at fine scales: Statistical adjustment of stage iv toward cpc gauge-based analysis," *Journal of Hydrometeorology*, vol. 15, no. 6, pp. 2542–2557, 2014.



Renpei Huang received Bachelor's Degree from Tsinghua University, China, 2013. His is now a student with the institute of Computer Graphics and Computer Aided Design, School of Software, Tsinghua University, China. His research interests include high dimensional data visualization and geo-based data visualization.



Quan Li received the PhD degree from the Hong Kong University of Science and Technology. He is now a research scientist at AI Group, WeBank, Shenzhen, and leading a team at WeBank AI to conduct research and projects in data science, data visualization and visual analysis, interpretable machine learning, and human-computer interaction techniques. He obtained his Bachelor's Degree from Wuhan University and Master's Degree from Tsinghua University.



Li Chen received the PhD degree in visualization from Zhejiang University, China, in 1996. She is currently an associate professor with the Institute of Computer Graphics and Computer Aided Design, School of Software, Tsinghua University, China. Her research interests include data visualization, image processing, and parallel algorithm.



Xiaoru Yuan received the BS degree in computer science and the BA degree in law from Peking University, in 1997 and 1998, respectively. In 2005 and 2006, he received the MS degree in computer engineering and the PhD degree in computer science from the University of Minnesota, Twin Cities. He is now a professor at Peking University, in the Laboratory of Machine Perception (MOE). His primary research interests lie in the field of scientific visualization, information visualization and visual analytics with emphasis on large data visualization, high dimensional data visualization, graph visualization and novel visualization user interface. He is a senior member of the IEEE.