Forecast Verification and Visualization based on Gaussian Mixture Model Co-estimation

Y.H. Wang[1] and C.R. Fan[2] and J. Zhang[3] and T. Niu[4] and S. Zhang[5] and J.R. Jiang[3]

[1]Shenzhen VisuCA Key Lab/SIAT, China
[2]East China Normal University, China
[3]Computer Network Information Center, Chinese Academy of Sciences, China
[4]Chinese Academy of Meteorological Sciences, China
[5]Mississippi State University, USA

Abstract
Precipitation forecast verification is essential to the quality of a forecast. The Gaussian Mixture Model can be used to approximate the precipitation of several rain bands and provide a concise view of the data, which is especially useful for comparing forecast and observation data. The robustness of such comparison mainly depends on the consistency of and the correspondence between the extracted rain bands in the forecast and observation data. We propose a novel co-estimation approach based on Gaussian Mixture Model in which forecast and observation data are analyzed simultaneously. This appoach natually increases the consistency of and correspondence between the extracted rain bands by exploiting the similarity between both forecast and observation data. Moreover, a novel visualization and exploration framework is implemented to help the meteorologists gain insight from the forecast. The proposed approach was applied to the forecast and observation data provided by the China Meteorological Administration. The results are evaluated by meteorologists and novel insight has been gained.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Weather Visualization —Verification

1. Introduction

Precipitation is one of the most important atmospheric variables that directly affects human activities and daily life. Precipitation forecasts play a key role in decision making where the occurrence of precipitation is sensitive, such as the issuing of flood warnings, the management of water sources, the design of agricultural policies and so on. Meteorologists are particularly interested in the shape and size of the rain bands: clouds and precipitation structures associated with areas of rainfall [CWS*08]. Verification of precipitation forecasts is a critical component of the development and application of a forecasting systems. It not only monitors the quality of forecasts but also provides useful feedback to the meteorologists to improve their forecasts.

The traditional and widely used verification approach is a side-by-side pointwise comparison after matching the forecast grid to the observation grid or a set of observation points [SWB89]. Although this kind of comparison provides some statistics to monitor forecast quality, it does not provide specific information regarding how a forecast went wrong or did well. As the spatial and temporal resolutions of forecasts from numerical weather prediction models grow, the need for spatial and temporal verification approaches increases greatly. Although many new strategies [CWS*08] have been proposed including the feature-based approaches and the scale-decomposition approaches, these techniques mainly focus on small precipitation areas.

The Gaussian mixture model (GMM) approach [LK10] provides a high level view of the data and facilitates convenient verification of the simulation. It consists of three stages. Firstly, rain bands represented by Gaussian distributions are extracted from the forecast and observation data individually by maximizing the likelihood of the GMM. Then, the correspondences between rain bands in the forecast and the observation data are determined by a comparison of their associated Gaussians. Finally, differences can be calculated using the parameters of the Gaussians for the matched pairs of rain bands. However, this approach faces several challenges when applied to complex

forecast and observation data over a long time period on a vast region such as North America or mainland China. In the rain band extraction stage, the number of rain bands varies with time. Hence a mechanism to automatically determine the number of rain bands is desirable. The robustness of the optimization algorithm for rain bands extraction is also important. If the extraction of rain bands is performed independently on the forecast and observation data, local maximums for the two data sets may lead to inconsistencies of the rain bands between them, which poses difficulties in determining the correspondence between forecast and observation rain bands.

In this paper, we propose a novel co-estimation scheme which integrates the rain bands extraction and correspondence determination. Exploiting the similarity in the forecast and observation data, the co-estimation provides a mechanism to extract rain bands from both data simultaneously. Therefore the correspondences between rain bands are naturally derived from the analysis, avoiding a subsequent step to determine correspondence. As a result, the issue of inconsistency that exists in methods such as [LK10] is alleviated while false rain bands and missing rain bands can also be identified in the process. In addition, we adopt the Bayesian Information Criterion (BIC) type model selector to automatically determine the number of rain bands rather than fixing it empirically [LK10].

To help meteorologists analyze the performance of forecasts in a given period of time, we present a novel framework for exploring the spatial-temporal trend of the forecast. The data we used lack temporal coherency because both the forecast and observation data are sampled on a daily interval. Therefore, we decide to extract rain bands for forecast and observation data at each time step independently. Several error indicators are introduced to measure the forecast performance. A multi-view visualization framework is implemented for exploring the results. With the ThemeRiver [HHWN02] view, the values of different errors in a time range can be easily compared. With a time series curve view, the temporal trend of these errors is revealed. A linked side-by-side comparison view enables the meteorologists to explore the geospatial rainfall distribution in key frames. A blending view shows the rain bands from forecast and observation in a single map. This framework was used to compare the forecasts covering continental China during June - August 2008 and the observations provided by the China meteorological administration. The meteorologists found the tool useful in understanding the differences of the geospatial distributions between forecast and observation and finding when, where and why the forecasts went wrong.

In summary, the main contributions of this paper include:

- We propose a new co-estimation scheme which facilitates not only natural matching of rain bands but also false rain bands and missing rain bands identification through

simultaneous analysis of the forecast and the observation data;
- We applied a set of effective visualization methods including the ThemeRiver view, the time series curve view, side-by-side view, and the blending view to visually verify the rain band forecast. Our evaluation showed that meteorologists gained new insight from the forecast with our visualization system.

2. Related Work

Previous work related to our work is divided into three different categories: verification of precipitation forecasts, comparative visualization, and Gaussian mixture model.

2.1. Verification of Precipitation Forecasts

Verification of precipitation forecasts is one of the most important topics in meteorology [Wil95]. The widely used approaches focus on the calculation of verification scores over a forecast-observation data set based on pointwise comparison. These approaches ignore the spatial distribution of precipitation. To overcome this drawback, many new spatial verification methods have been proposed. A complete review of these methods is beyond the scope of this article, we refer the reader to Casati et al. [CWS*08]. We restrict our discussion to the feature-based verification methods, which first identify features in the forecast and observation field and then compare the attributes associated with each pair of features. Ebert and McBride were among the first to explore the verification of rainfall using labeled contiguous rainfall areas (CRA). The CRA is the union of observed and forecast rainfall areas on which the rainfall exceeds a user-specified threshold. To obtain more contiguous areas, Davis et al. [DBB06] defined features in both forecasts and observations based on a convolution and thresholding procedure. To measure the errors between each paired area, they then associated the forecast and observed objects with geometric shapes. Lakshmanan and Kain proposed a GMM based forecast verification method [LK10], where Gaussians are employed to represent rain bands. Similar to these methods, the feature in our approach is also defined by a contiguous rain area. The difference is that we considered precipitation verification based on a long time period over a vast region. A novel co-estimation algorithm is proposed and other necessary improvements are made.

2.2. Comparative Visualization

Comparative visualization is a very instructive method for verification. According to the starting point of comparison, it can be classified into three levels [SP98]: image level, data level and feature level. The most common comparison approach is image level comparison which compares images (e.g., output of visualization algorithms) using a variety of methods [GAW*11], such as side-by-side, overlay, and

difference images. Side-by-side relies on user's memory to compare images which are shown separately, while overlay shows multiple images on the same space and thus sometimes suffers from visual clutter. Compared to these side-by-side and overlay methods, difference images explicitly represent the difference between objects. In most of visualization systems, these methods are mixed together. Starting from the raw data, data level comparison [PP97] can compare the data sets using a variety of intermediate representations derived from the raw data. The widely used method of verification scores [Wil95] in precipitation forecasts is an example. The main drawback of this approach is that the resulting comparison systems may become application dependent.

Unlike image level and data level comparisons, feature level comparisons at a higher level, incorporating application specific knowledge, are of great interest to users. The features are derived according to application characteristics. In this paper, we extract the rain bands as features and then compare forecasts and observations based on them. To facilitate visual analysis of precipitation distribution, we provide a coordinated view to compare the differences between pairs of rain bands.

2.3. Gaussian Mixture Model

Gaussian Mixture Model (GMM) [Bil98] is a mature clustering method. It models each cluster as a Gaussian, with its mean somewhere in the middle of the cluster, and a standard deviation that measures the spread of that cluster. It has been used in the visualization community [CCM09, WCZ*11]. The most relevant work is [LK10] which also uses GMM for precipitation verification. Our GMM is applied in the context of feature level comparison, where the correspondences between the features extracted from the two data sets have to be established. Rather than defining some metrics to match the individual extracted features, we propose a new GMM based estimation algorithm: co-estimation, which can simultaneously identify features and build feature correspondences.

3. Data Description

While our verification approach can be applied to a variety of forecast and observation data, the present application considers precipitation forecast from the Global and Regional Assimilation Prediction System (GRAPES) [XL07]. This model is used to generate precipitation forecast every 3 hours for China and the surrounding area. Then, it averages 8 ensemble simulation results to get daily forecast 24 hours ahead of the observation. It uses a 500×330 grid with a resolution of $0.15°$ longitude/latitude, which covers the region from latitude $15.0°$N to $64.5°$N, longitude $70.0°$E to $145.0°$E. We plot the map of the China in a longitude/latitude coordinate system used by our meteorologists, as shown in Figure 2.
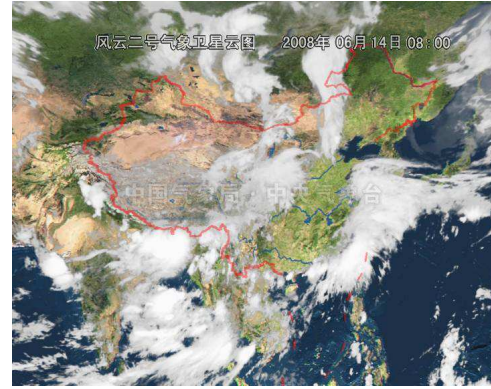
Figure 1: *Satellite image of the strong precipitation process over China on 2008-06-14 08:00:00 GMT+8 [Chi08].*

We chose the observation data provided by the ground-based meteorological observation network of China Meteorological Administration as verification data. In each ground observatory, the amount of rainfall is measured using rain gauges. After obtaining the daily reported rainfalls from 2,400 ground observatories, the Barnes interpolation [Bar64] scheme is used to interpolate this set of observation points into the same grid of the forecast system. All of the ground observatories are located within China. For comparison, we only consider the forecast result within mainland China and Hong Kong. Rainfall intensity is classified by the rainfall amount in 24 hours as [Chi07]:

- light rain − when the rainfall amount is less than 10 mm;
- moderate rain − when the rainfall amount is between 10 to 25 mm;
- heavy rain − when the rainfall amount is between 25 to 50 mm;
- violent rain − when the rainfall amount is larger than 50 mm.

Usually the rainfall amount is between 0 and 25 mm.

To demonstrate the effectiveness of our verification approach, we use the forecast produced by GRAPES during June - August, 2008, because the summer average rainfall in 2008 is the highest recorded rainfall in recent decades and it is spatially uneven [Wu10]. The rainfall in some parts of northern China was only 30 percent to 80 percent of the normal amount, while the rainfall in Shandong province and parts of southern China increased by 30-100 percent from the normal amount. Moreover, there were several strong precipitation processes [Wu10] that caused extensive damage and great economic loss. Figure 1 shows a visible satellite image of a strong precipitation process influencing a large region in China on June 14, 2008. This precipitation process is characterized by two severe frontal rain bands.

4. Rain Bands Exploration

The Gaussian Mixture Model approach [LK10] approximates the distribution of the rainfall using several rain bands, each represented as an elongated ellipse determined by a Gaussian. The parameters of the Gaussians are optimized by maximizing the likelihood and the number of Gaussians is determined by balancing between fidelity and complexity. In this section, we discuss the differences between our algorithm and the algorithm of [LK10] in these two aspects.

4.1. Extraction of Rain Bands

Consider the data as a set of samples: $\{(x_1, h(x_1)), (x_2, h(x_2)) \cdots, (x_n, h(x_n))\}$ where x_i is a 2D vector of latitude and longitude in latitude/longitude grid and $h(x_i)$ is the rainfall amount at the location of x_i. Using the GMM approach [LK10], the probability of a sample x being generated by a rain band is:

$$g(x|\mu, \Sigma) = \frac{1}{2\pi |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)},$$

where μ is the center vector of the Gaussian of the rain band and Σ is the 2×2 covariance matrix. Hence, the rainfall amount at the location of x is represented as the mixture of several rain band distributions:

$$\hat{h}(x|\theta) = \sum_{j=1}^{k} p_j g_j(x|\mu_j, \Sigma_j), \tag{1}$$

where $\theta = (\theta_1, \cdots, \theta_k) = ((p_1, \mu_1, \Sigma_1), \cdots, (p_k, \mu_k, \Sigma_k))$ contains the parameter set of the k Gaussians, g_j is the probability of x being generated by a particular rain band j. Since each Gaussian function integrates to one, the mixing probability p_i must satisfy the following condition:

$$\sum_{j=1}^{k} p_j = 1 \text{ and } p_j \geq 0.$$

This density estimation problem can be defined as specifying the rain bands from which the samples $\{(x_1, h(x_1)), (x_2, h(x_2)) \cdots, (x_n, h(x_n))\}$ are most likely to be generated to determine θ. Assuming independence between the samples, the maximum likelihood estimation of θ is

$$\hat{\theta} = \arg\max_{\theta} \prod_{i=1}^{n} \hat{h}(x_i|\theta) = \arg\max_{\theta} \sum_{i=1}^{n} \log \hat{h}(x_i|\theta).$$

The expectation-maximization (EM) algorithm [Bil98] is a proven method for parameter estimation. It iteratively improves the estimation of θ. Starting with an initial parameter set θ, it iteratively performs the following two steps until converging on a maximum of the likelihood function:

- E Step:

$$P(j|x_i) = \frac{p_j g(x_i|\mu_j, \Sigma_j)}{\sum_{j=1}^{k} p_j g(x_i|\mu_j, \Sigma_j)}$$

- M Step:

$$p_j = \frac{1}{n} \sum_{i=1}^{n} P(j|x_i) \quad \mu_j = \frac{\sum_{i=1}^{n} P(j|x_i) x_i}{\sum_{i=1}^{n} P(j|x_i)}$$

$$\Sigma_j = \frac{\sum_{i=1}^{n} P(j|x_i)(x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{n} P(j|x_i)},$$

where $P(j|x_i)$ is the probability of the sample x_i belonging to the jth rain band.

Here, we use the greedy EM algorithm [VVK03], which does not require the user to specify the initial parameters. Instead of starting with a configuration of all Gaussian components and improving upon this configuration with EM, it builds the Gaussian mixture adaptively. Starting with a single component whose parameters are easily computed, inserting a new component and updating the complete Gaussian mixture using EM algorithm are alternatively performed. Figure 2 shows the extracted rain bands from the forecast data on August 2, 2008 using this greedy algorithm. Compared to Figure 2(a), a small rain band located in North China is extracted in Figure 2(b), while Figure 2(c) shows a thin rain band located in Tibet which does not appear in Figure 2(b).

### 4.2. Estimation of the number of Rain Bands

In the greedy EM algorithm, the maximum number of components needs to be specified by the user. Usually the meteorologists have some idea about the maximum number of rain bands over a region. However, it is difficult for them to find the appropriate number of rain bands each day because it varies with time.

In [LK10], a BIC type model selection procedure, Minimum Description Length (MDL) [RHRP98] to be precise, is introduced to determine the number of rain bands. As they mentioned, the optimal number of rain bands produced by this model is usually in the order of hundreds. However, their solution fixes the number of rain bands to three.

A fundamental objective in MDL is to balance model fit versus model complexity. Hence, the description length consists of two parts: the fit of the model to the data and the complexity of the model. For GMM, the former is the log likelihood of the data under the model, while the latter is the number of free parameters. Thus, the estimation of k in Equation 1 is based on the minimization of the expression:

$$MDL(k, \theta) = -\sum_{i=1}^{n} \log \hat{h}(x_i|\theta) + \frac{1}{2}\lambda L \log(2n). \tag{2}$$

In our case of a k-component 2D GMM, the number of
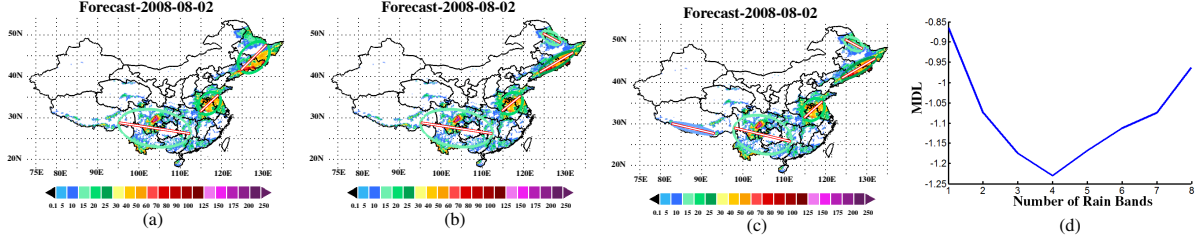
Figure 2: *Using the greedy EM algorithm to extract rain bands with different number of Gaussians and selecting the optimal number of rain bands with weighted MDL for the forecast data on August 2, 2008. (a) 3 rain bands; (b) 4 rain bands; (c) 5 rain bands; (d) the minimal description lengths of 8 mixtures.*

free parameters is $k-1$ for prior probabilities, $2k$ for means and 3k for covariance due to the symmetry of Σ, therefore $L = 6k-1$. λ is the weight between the model fidelity and complexity. With a small λ, there could be hundreds of Gaussians to approximate the distribution and co-estimation would be far more complex. Empirically, much fewer large components (more than three however) are required for forecast verification. This observation is also confirmed by [LK10]. To balance the two needs, we set λ to 350 which results in the number of components between 2 and 6.

With a pre-specified maximum number of rain bands, the greedy EM algorithm sequentially calculates a description length for each trained mixture. Then, the mixture with the minimal MDL is selected. Figure 2(d) shows the minimal description lengths of 8 different mixtures extracted from the forecast data on August 2, 2008. It can be seen that the four rain bands approximation is the optimal as shown in Figure 2(b). This example also demonstrates that the MDL is able to effectively determine the number of rain bands automatically in practice.

4.3. Visualization of Rain Bands

Following the standard of the China Meteorological Administration on the classification of rain gauges [Chi07], we apply the color map provided by meteorologists to the rainfall data (Figure 2). Other choices of the color map (for example ColorBrewer [HB03]) can be easily implemented. For intuitive rain bands exploration, we encode each rain band's associated Gaussian with an ellipse. The center of the ellipse is the mean value μ, and other parameters can be computed by applying the singular value decomposition algorithm to the symmetric matrix Σ^{-1}:

$$\Sigma^{-1} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

$$= \begin{bmatrix} -\cos(\phi) & \sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}^T$$

where $\frac{1}{\sqrt{\sigma_1}}$ and $\frac{1}{\sqrt{\sigma_2}}$ are the radii along the major and minor axes, and ϕ is the angle which rotates the coordinate

axes to the major axis, obtained from the major eigenvector orientation. In other words, ϕ represents the orientation of the rain band. Due to its importance in the analysis of precipitation, we use the white line to highlight the major axis. The color of the ellipse contour is determined by the average rainfall amount in its covered range:

$$\omega = \frac{1}{N} \sum_i h(x_i),$$

where $h(x_i)$ is the rainfall amount of the point x_i located in the ellipse and N is the number of points in the ellipse. Then the color of the ellipse contour can be obtained by mapping this average intensity to the color map.

5. Co-estimation of Rain Bands

Inspired by the co-segmentation based correspondence determination [ČMP09], we propose a co-estimation algorithm which simultaneously extracts rain bands and builds correspondences between them. Figure 3 shows the pipeline of this algorithm.

5.1. GMM-based Co-Estimation

The GMM-based co-estimation consists of three stages. First, the forecast and observation data are merged together by taking their average. Then rain bands are extracted from the averaged data. Through this operation, similar portions of these data are emphasized and bias toward each other is introduced. The resulting extracted rain bands include the ones shared by both and the ones that belong to just one of the data, as illustrated in Figure 3(c). Then, the common and distinctive rain bands are distinguished for each data using again a model fidelity and complexity check. For example, as illustrated in Figure 3(c) and Figure 3(d), the rain band at the northern east China does not contribute to the model fidelity and thus removing it will only reduce the complexity. Hence it is identified as a false rain band. Finally, the remaining rain bands for each data are refined using the EM algorithm to maximize the likelihood, illustrated in Figure 3(f) and Figure 3(g).
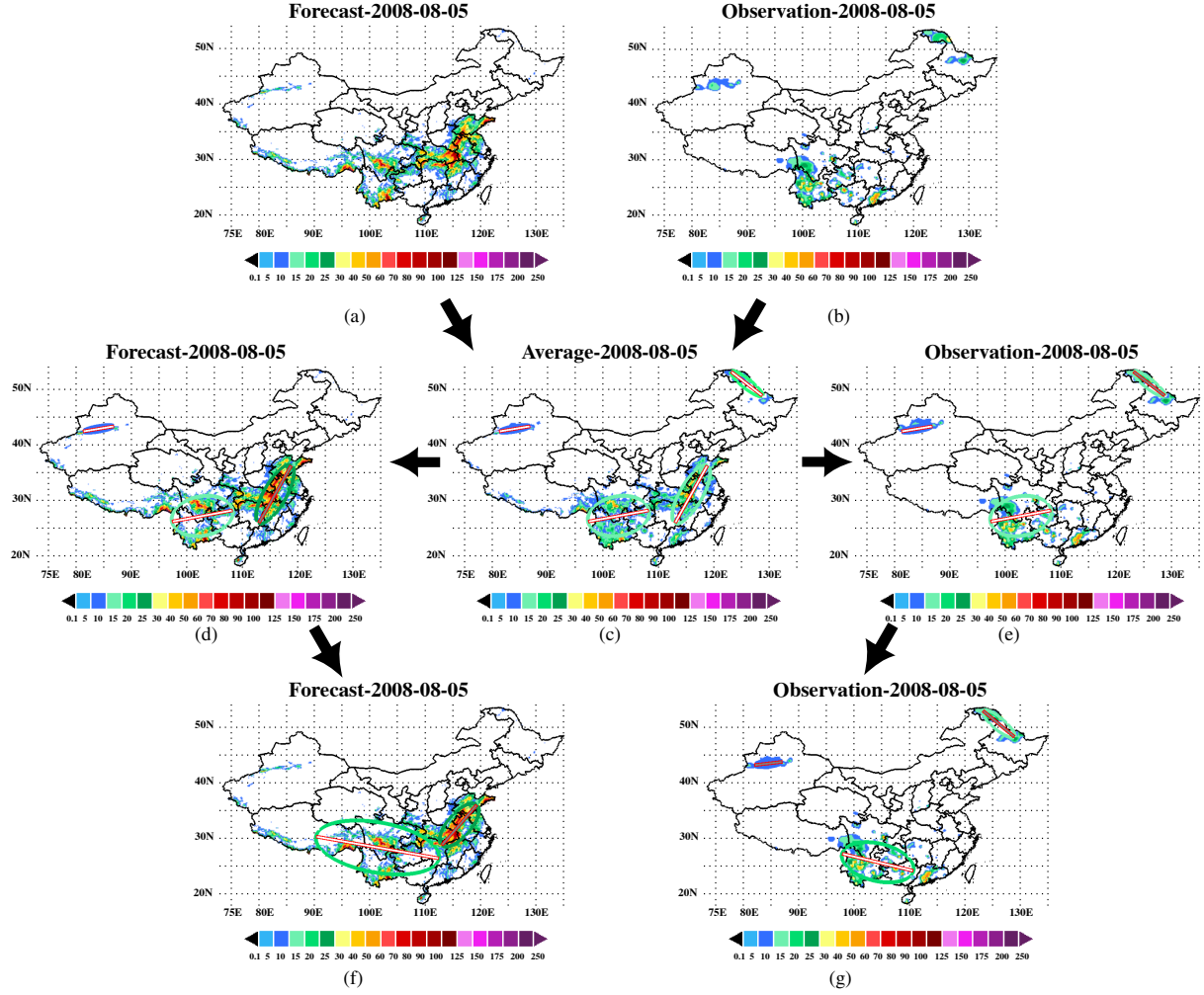
Figure 3: *The pipeline of our co-estimation algorithm illustrated on the forecast and observation data on August 5, 2008. (a,b) The rainfall distribution of the forecast and observation, respectively; (c) the averaged distribution of forecast and observation and its corresponding rain bands; (d,e) the initial rain bands for the forecast and the observation, respectively; (f,g) the refined rain bands of the forecast and observation, respectively. The white major axes indicate matched rain bands and gray unmatched.*

Average GMM To guarantee all rain bands in the forecast $\{h_1^f, \cdots, h_m^f\}$ and observation $\{h_1^o, \cdots, h_m^o\}$, where m is the number of grid points, can be extracted, we first combine them into one data set $\{h_1^a, \cdots, h_m^a\}$ using the average operator:

$$h_i^a = \frac{h_i^f + h_i^o}{2},$$

and then extract several rain bands from this data in the form of the GMM $\{\theta_1^a, \cdots, \theta_k^a\}$:

$$\hat{h}^a(x|\theta^a) = \sum_{j=1}^{k} p_j^a g_j(x|\mu_j^a, \Sigma_j^a). \tag{3}$$

We name this GMM the *average* GMM. Figure 3(c) shows the extracted rain bands.

Initial GMM With this *average* GMM, the initial GMM for each data can be obtained by checking whether the rain band represented by the jth Gaussian in the average GMM appears in this data. After calculating the description length $MDL(k, \theta^a)$ with the average GMM for this data, we then define a new model with parameters $\theta_j = \{\theta_i^a\}_{i \neq j}^k$. With this new model, we calculate a new description length for this data $MDL(k-1, \theta_j)$. If $MDL(k-1, \theta_j) < MDL(k, \theta^a)$, we assume the jth Gaussian does not appear in the data, because its contribution to the term of log likelihood is smaller than the model complexity. Otherwise, this Gaussian is valid.

After testing each Gaussian for forecast and observation data, we obtain the initial Gaussians θ^o and θ^f for them:

$$\bar{h}^o(x|\theta^o) = \sum_{j \in S_o} p_j^a g_j(x|\mu_j^a, \Sigma_j^a),$$

$$\bar{h}^f(x|\theta^f) = \sum_{j \in S_f} p_j^a g_j(x|\mu_j^a, \Sigma_j^a), \qquad (4)$$

where S_o and S_f consist of the indices of the Gaussians whose corresponding rain bands appear in the observation or forecast, respectively. Since both S_o and S_f are subsets of the original indices, the correspondences between the rain bands in the forecast and observation are naturally established. In order to distinguish between the matched and unmatched rain bands, the major axis of the matched rain band is colored in white and the unmatched in gray. Figure 3(d,e) show an example, where the rain band in northeast China is removed in Figure 3(d) and the rain band in eastern China is removed in Figure 3(e).

Refined GMM By individually refining the *initial* GMM of either forecast or observation data with the EM algorithm, a *refined* GMM can be obtained which maximizes the likelihood separation of different rain bands:

$$\hat{h}^o(x|\theta^o) = \sum_{j=1}^{k_o} p_j^o g_j(x|\mu_j^o, \Sigma_j^o)$$

$$\hat{h}^f(x|\theta^f) = \sum_{j=1}^{k_f} p_j^f g_j(x|\mu_j^f, \Sigma_j^f), \qquad (5)$$

where k_o and k_f are the numbers of rain bands in the observation and forecast, respectively. With this refinement, the rain band in South China is enlarged in Figure 3(f) and narrowed in Figure 3(g).

Since the meteorologist is interested in the rain bands which may bring heavy rains, we remove any Gaussian whose average rainfall amount is less than a specified threshold. We also remove the Gaussian ellipses whose covered areas are smaller than a threshold, because the meteorologist assumes the corresponding rain band is caused by local precipitation in these cases and not useful for the analysis of the overall distribution of the rainfall. Due to these two conditions, the small rain band in Figure 3(d) is removed in Figure 3(f).

5.2. Verification of Rain Bands

To quantitatively measure the differences between rain bands, the matched and unmatched rain bands are compared in different ways. Since the parameters that the meteorologists mainly concern about are the centroid, average intensity, and orientation of the rain band, we denote them as $r = \{\mu, \omega, \phi\}$. For k matched rain bands with parameters $r^f = \{\mu_i^f, \omega_i^f, \phi_i^f\}_{i=1}^k$ and $r^o = \{\mu_i^o, \omega_i^o, \phi_i^o\}_{i=1}^k$, we calculate the errors between the parameters of each pair

of rain bands in the forecast and observation:

$$e(\mu_{i_x}) = \mu_{i_x}^f - \mu_{i_x}^o,$$
$$e(\mu_{i_y}) = \mu_{i_y}^f - \mu_{i_y}^o, \qquad (6)$$
$$e(\omega_i) = \max(\omega_i^f/\omega_i^o, \omega_i^f/\omega_i^o) - 1,$$
$$e(\phi_i) = \min(\phi_i^f - \phi_i^o, 180 - \phi_i^f + \phi_i^o).$$

where μ_x and μ_y are defined by latitude and longitude, respectively. These errors measure the deviation between a pair of rain bands.

To give an overall error for a pair of matched rain bands, we weight each error with the constants in [LK10] and sum the absolute values of them as:

$$e(r_i^f, r_i^o) = 0.3 * \min(\frac{e(\mu_{i_x})}{100}, 1) + 0.3 * \min(\frac{e(\mu_{i_y})}{100}, 1)$$
$$+ 0.5 * e(\omega_i) + 0.2 * e(\phi_i)/90. \qquad (7)$$

The overall forecast error of k matched rain bands is computed as the average of all rain band errors weighted by the area of the corresponding rain band in the observation.

If an unmatched rain band appears in the forecast, it corresponds to a false alarm; if it appears in the observation, it corresponds to a missing event. The rain band based false alarm rate is defined as the ratio between the number of valid points in the unmatched rain bands and all the valid points,

$$FAR = \frac{N_{uf}}{N_f}, \qquad (8)$$

where N_{uf} is the number of valid points covered by the unmatched rain bands in the forecast and N_f is the number of all valid points covered by the extracted rain bands in the forecast. Here "valid" means a point is located in the rain band and the rainfall amount at the point is larger than a specified threshold. Similarly, the missing rate can be defined as,

$$MR = \frac{N_{uo}}{N_o}, \qquad (9)$$

where N_{uo} is the number of valid points covered by the unmatched rain bands in the observation and N_o is the number of all valid points covered by the extracted rain bands in the observation.

## 6. Spatial-Temporal Exploration of Forecast Performance

With the co-estimation, several error indicators are calculated such as the errors between matched rain bands, the false alarm ratio (*FAR*), and the missing ratio (*MR*) for a time sequence of forecast and observation data. These indicators can help the user to quantitatively evaluate the forecasts and find abnormal ones to investigate further. We provide a novel framework for visualization and exploration of these indicators.
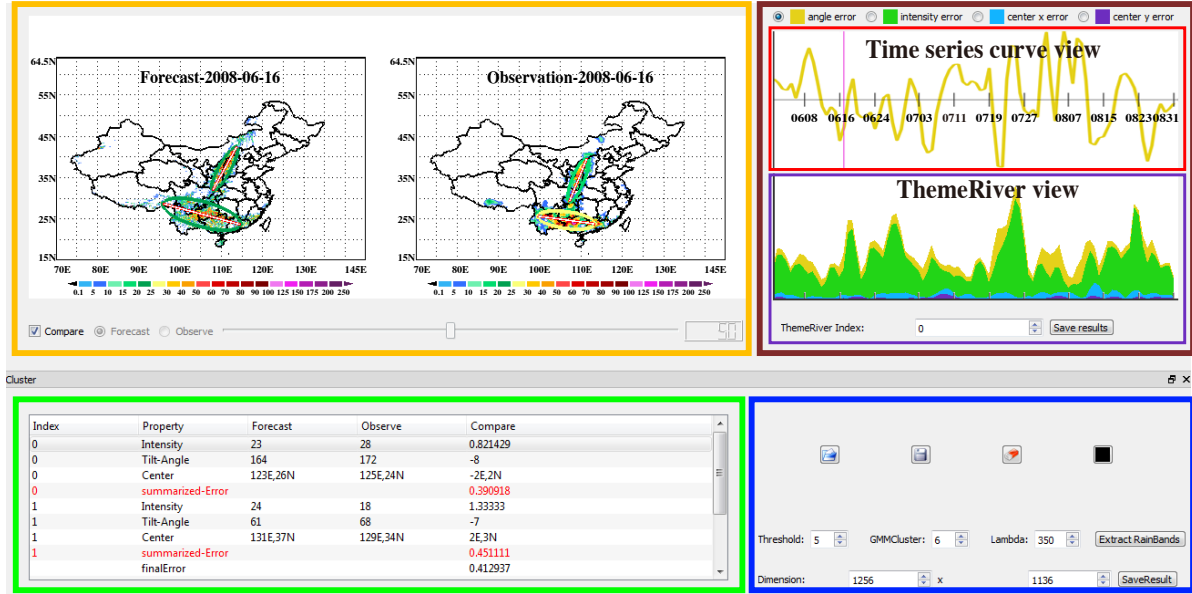
Figure 4: *The user interface of our system. The comparison view is outlined in yellow, the ThemeRiver view in purple, the time series curve view in red, the statistical view in green, and the parameters control area in blue. The verification result of June 16, 2008 in the comparison view is indicated by a pink selection bar in the time series view.*

6.1. User Interface

The proposed rain band based verification facilitates an informative overview as seen in Figure 4: a side-by-side
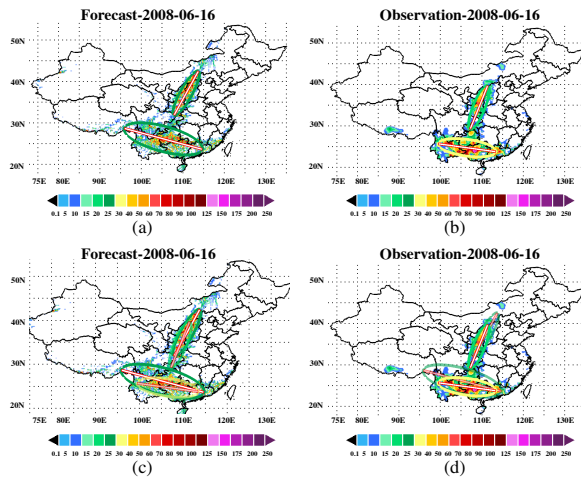
Figure 5: *The side-by-side view and blending view to show comparison of the forecast and observation data on June 16, 2008. (a,b) Side-by-side view; (c) Blending the rain band in observation data to forecast data with the blending ratio 0.7; (d) Blending the rain band in forecast data to observation data with the blending ratio 0.7.*

comparison view (outlined in yellow), a linked trend view (outlined in brown) consisting of a ThemeRiver view (outlined in pink) and a time series curve view (outlined in red), a statistical view (outlined in green), and a parameters control area (outlined in blue). The differences in the rain bands from forecast and observation can also be examined in the blending view. The main idea of combining these views is inspired by Tufte's macro/micro principle [Tuf83] for the purpose of showing the individual verification while conveying an overview. The comparison view and the statistical view are the micro views for displaying the verification result of key frames of the forecast and observation sequences. The parameters control area contains the maximal number of rain bands, the intensity threshold for rain band and the size of the rendered image.

ThemeRiver view We implemented a simplified version of ThemeRiver [HHWN02] that stacks multiple time series plots in one view. This view helps us see not only the trends of individual variables but also the cumulative trend of several variables. Here, there are two kinds of trends, which represent the forecast quality based on matched and unmatched rain bands, respectively. To reveal the overall error between matched rain bands, center error in x and y directions, intenstiy error, and angle error all are represented as the correspondng weighted terms in Equation 7 and are rendered as four themes in one trend. The other trend consists of the metrics of *FAR* and *MR*, which represents the changing of the numbers of false and missing alarms. In Figure 4, the ThemeRiver view shows the four errors

of matched rain bands during June - August 2008. We can easily see that the intensity error is the largest and the center error in y direction is the smallest.

To help the user explore the temporal trend, several interactions are provided. First, the user can reorder the themes to emphasize important metrics. Second, the theme of interest can be highlighted by clicking. Third, the user can focus on a specific time slice with a selection bar.

Time series curve view Although the ThemeRiver is very effective in revealing the cumulative extent of several attributes, it is not effective in showing the trend of individual attribute which is sometimes important for analysis. For this purpose, we provide a time series curve view as shown in Figure 4. To highlight important patterns like the peaks and crests and suppress the aliasing effect, a Gaussian filter is used to smooth the time series curves.

ThemeRiver view focues on revealing the relative differences among four errors, while time series curve view shows the trend of each error. To help the user explore both types of the information simultaneously, these two views are linked together. Once the user selects a label at the top of ThemeRiver view, its time series curve is shown and the corresponding theme in ThemeRiver view is highlighted. When the user selects a theme in ThemeRiver view, its corresponding time series curve is shown. The color of the curve is set to the same as its corresponding theme in the ThemeRiver view.

Side-by-Side comparison view As noted by Robertson et al. [RFF*08], side-by-side comparisons are less sensitive to visual clutter, but require more time to scan all windows and more mental overhead to compare the differences. To alleviate this problem, the diameter of the unmatched rain band is colored in white while the matched are in red. Meanwhile, we highlight the matched one in the other data when the user selects a rain band in the forecast or observation data.

Blending view The side by side view requires users to mentally combine the two images in their minds when they compare the rain bands in forecast and observation data. This could be demanding, time consuming, and inaccurate, especially when many small rain bands appear in the map. Therefore, we also provide an alpha blending of the two kinds of rain bands to get a better observation at some overlapped rain bands when forecast and observation results simultaneously appear in the same view. In a single view, the user can set the background to either forecast or observation and then adjust the blending ratio with a sliding bar. This blending eases the user's task of comparing shape, orientation, and locations of matched rain bands.

6.2. Linking between Views

Using interactive brushing, the time sequences can be filtered into a specific time range or a time slice. When brushing one of the ThemeRiver view or the time series curve view, the other views will be automatically updated. Once a time slice is selected from this range, the comparison view and the statistical view will show the verification result. When the user explores the rain bands in the comparison and blending view, the errors corresponding to the selected pair of rain bands are highlighted in the statistical view. On the other hand, when the user explores the statistical view, the corresponding pair of rain bands will be highlighted when he/she selects one error term.

Figure 4 shows an example where the time slice is indicated by the pink line in the time series curve view. In this example, the verification result for June 16, 2008 is selected, which has the largest forecast error in angle error of rain bands. From the comparison views in Figure 5, we can see that two rain bands are matched but the directions are very different. Selecting this pair of rain bands in the forecast or observation, its corresponding error is highlighted in the statistical view.

7. Expert Evaluation

Two meteorologists used our system to perform visual analysis of the forecast and observation data during June - August, 2008. Before analyzing the forecast performance, the meteorologists conducted several tests to verify our rain bands extraction and matching algorithm. Then, they analyzed the forecast performance on different levels and investigated which factors the forecast system should emphasize or deemphasize.

7.1. Extraction of rain bands

In this section, the meteorologists applied our method on several data sets to verify whether our method can extract reasonable rain bands. We present two examples here.

Figure 6(a) is the observation data on June 16, 2008. The meteorologists, according to their prior knowledge, know there were two rain bands mutually perpendicular to each other. From the orientation of our extracted rain bands, we can see the intersection angle between them is about $90°$ ($86°$ to be exact), which confirms the knowledge of the meteorologists. Figure 6(b) is another example, where the meteorologists thought there was one rain band from the northeast of China to South China while our method separates it into two rain bands. Although our result is not exactly the same as the meteorologists' preconception, it is still reasonable, because there are two centroids of precipitations. The one from the northeast of China to North China can be classified as heavy rain, while the intensity of the other from Hubei province to Guangxi is not so strong. The other two rain bands are also reasonable where the one in the corner of northeast of China is caused by frontal predications and the one in southwest is caused by the ocean circulation.

Through these tests, the meteorologists concluded that our extracted rain bands are reasonable. Moreover, they pointed out that the highlighted major axis of rain bands is useful in getting visual perception of the orientation of rain bands.
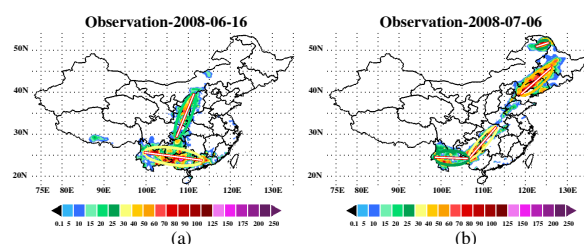
Figure 6: *Two examples of rain bands exploration. (a) two mutually perpendicular rain bands; (b) four smaller rain bands.*
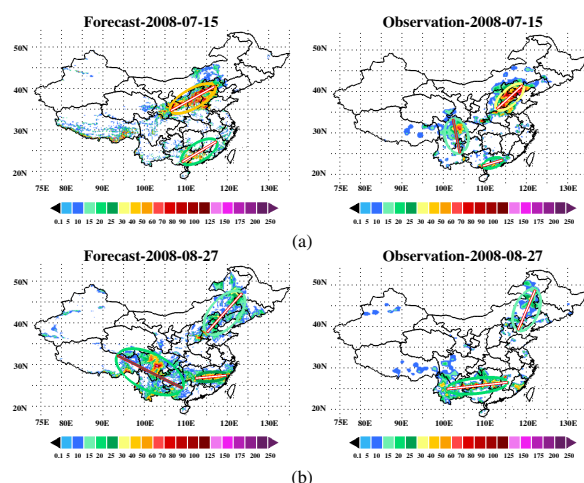
Figure 7: *Two examples of using our rain band based verification of precipitation forecast. (a) a missing rain band with a gray major axis shown in the observation; (b) a false rain band with a gray major axis shown in the forecast.*

7.2. Matching of rain bands

By using our proposed co-estimation algorithm, the correspondence of rain bands in forecast and observation can be built naturally. Correct correspondence can help the meteorologists to compare the formations of rain bands. Here, we present two examples where the meteorologists verified the validity of the correspondences with our method (Figure 7).

Figure 7(a) shows the rain band based verification result on July 15, 2008. Obviously, the rainfall map of the forecast indicates that there is a rain band missing in the forecast.

In this result, we can see that two rain bands are matched but they are much larger in the forecast than the ones in the observation. And one rain band in the Sichuan Basin is missing in the forecast. Therefore, the correspondences are reasonable. Figure 7(b) is a slightly more complex example. Not only does it have a false rain band, but also the rain band in the southwest of China in the observation may match to the false rain band because they are more similar in size. Nevertheless, our co-estimation method builds the correct correspondence and suggests that the center of the matched rain band located at the southwest of China in the forecast should move to the west a little.

After extensively testing this issue on all pairs of the forecast and observation, our co-estimation algorithm gets correct matching in 89 out of 92 tests. Thus, the meteorologists finally concluded that our method gives the right correspondence most of the time, which is particularly useful in helping them analyze the difference between forecast and observation. Besides, they also observed that the blending view ameliorates the task of comparing matched rain bands, making their analysis work more convenient.

### 7.3. Forecast Performance Analysis

By linking time series with spatial distributions of rain bands, the meteorologists can perform spatial-temporal analysis of the forecast sequence. Here, the meteorologists analyzed the forecast performance on two levels by setting two different thresholds of rainfall intensity at 5*mm* and 25*mm*. The former corresponds to the distribution of all kinds of precipitation, while the latter corresponds to the heavy and violent rains.

After setting the threshold to 5*mm*, Figure 9(a,b) shows the errors between matched rain bands and the errors of unmatched rain bands (*FAR*, *MR*), respectively. By observing their corresponding themes in Figure 9(a), we can see that the center error in y direction plays a minor role while the intensity error dominates the ThemeRiver view. It can be seen that they (Figure 8(b) and Figure 8(c)) are both biased toward the positive direction by inspecting the corresponding time series curves (Figure 8). In contrast, most of errors in the x direction (Figure 8(a)) are negative. The meteorologists concluded that the moving speeds of the forecasted fronts are a little off. Further investigation and experiments are needed to identify the reasons.

By analyzing the time series curve, we can clearly identify several outliers, which are more than two standard deviations away from the mean. For example, there are four outliers marked in red in the curves of the intensity error and the center error in y direction. After checking the side by side view of the corresponding data sets, we concluded that this is caused by the definition of our error indicator. The shape of the rain band in this case is close to a circle, no signification
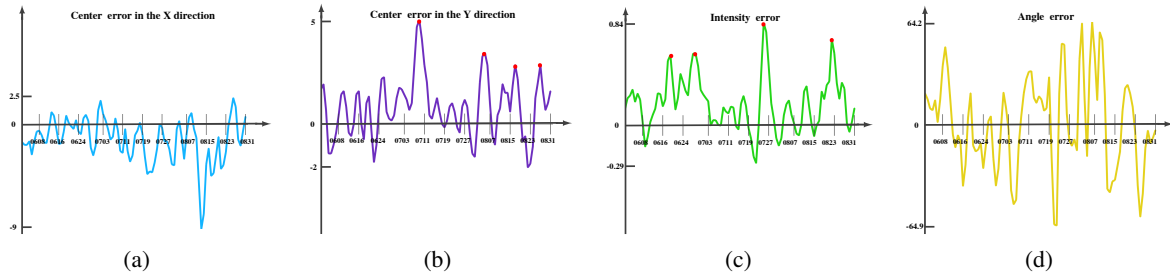
Figure 8: *Time series curves of four error indicators of matched rain bands with the threshold set at 5mm, where outliers in (b,c) are marked in red.*
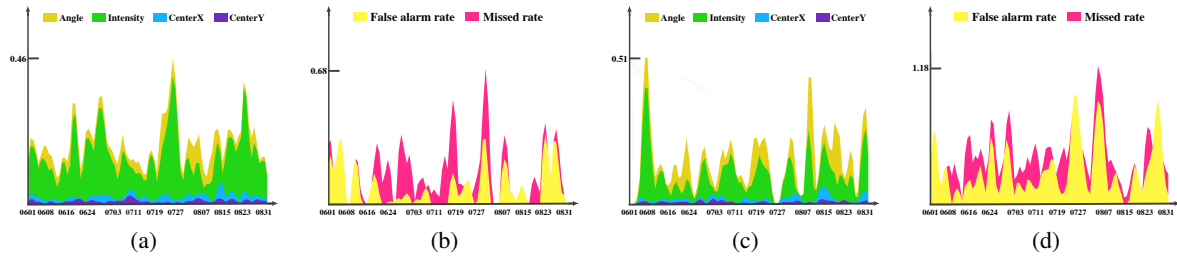
Figure 9: *The ThemeRivers of errors of rainfall intensity between matched rain bands and errors introduced by unmatched rain bands (FAR and MR) with different thresholds. (a,b) The threshold set at 5mm; (c,d) The threshold set at 25mm.*

orientation is shown. However, our error indicator always uses the direction of major axis, giving misleading results in this case. New definition of this error is under consideration.

From Figure 9(b), we can see that the false alarm rate is higher than the missing rate in the first half of June; after that the false alarm rate is close to the missing rate. Figure 9(c,d) show the ThemeRiver views of corresponding errors as the threshold is set to 25mm. From Figure 9(c), we can see the intensity error still dominates the ThemeRiver view while the center error in y direction is the smallest as in Figure 9(a). Since our domain scientists assume the rainfall intensity is less than 5mm can be ignored, we can conclude that many rainfalls are missed by the forecast system.

Combining Figure 9(c) and Figure 9(d) together, we can find the errors between matched rain bands disappear while the false alarm rate is high in some days, for example, June 1, July 3, and August 29. Compared to Figure 9(b), the false alarm rate and especially the missing rate are higher in Figure 9(d). This indicates that the forecast system performs poorly for heavy rain during this time period. This is an unexpected result. After checking with our domain scientists, there are climate anomalies in summer of 2008 in China and several rainstorms are unexpected.

7.4. Expert Feedback

Our framework is iteratively refined based on the requests and the feedback from the meteorologists. After explaining

the visualizations we created for their data, they were able to identify issues. For example, we only designed the time series curve view to show four errors defined in Equation 6 initially. One meteorologist pointed out that the relative differences among different errors are important for them in the diagnosis of the forecast system. Considering this, we linked ThemeRiver view and time series curve view together to simultaneously depict the relative differences and absolute errors.

The informal feedback from meteorologists confirms that our visual analysis framework offers both a good overview of the forecast performance and detailed patterns about rain bands. In particular, they found our side-by-side and blending views are quite useful for them to check the differences between the rain bands in forecast and observation data. The linked views between ThemeRiver view and time series curve view help them examine the trends of the individual errors and the relative frequency of the four errors at the same time.

8. Conclusion and Future Work

This paper introduces a GMM based co-estimation method for verification of the precipitation forecasts. By using the GMM to extract rain bands from the forecasts or observations, the spatial distribution of the rainfall is revealed. A new co-estimation scheme facilitates simultaneous analysis of the forecasts and observations. With a novel visual analysis framework, the spatial-temporal

trend of the verification results can be explored efficiently. Using the forecast and observation data of China during June - August 2008, the effectiveness of our system has been verified by meteorologists.

However, there are still some limitations in our work. The current GMM based rain band extraction is exclusively based on the distribution of rainfall, while the meteorologists sometimes consider other factors, such as the wind. Incorporating such kind of factors into our algorithm is a part of future work. Second, our GMM co-estimation algorithm relies on the assumption that the rainfall distributions in the forecast and observation are similar. However, this assumption is not always true. Future work is needed to accommodate for rain bands of disparate sizes.

Acknowledgements

References

[Bar64] BARNES S.: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor 3*, 4 (1964), 396–409. 3

[Bil98] BILMES J.: *A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models*. Tech. Rep. ICSI-TR-97-02, University of Berkeley, 1998. 3, 4

[CCM09] CORREA C., CHAN Y., MA K.: A framework for uncertainty-aware visual analytics. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology'09* (2009), pp. 51–58. 3

[Chi07] CHINA METEOROLOGICAL ADMINISTRATION: National standard on the classification of rain intentisy. http://www.gov.cn/ztzl/2008tffy/content_1113935.htm, 2007. 3, 5

[Chi08] CHINA METEOROLOGICAL ADMINISTRATION: Weather satellite image. http://www.duststorm.com.cn/Article/ShowArticle.asp?ArticleID=22124, 2008. 3

[ČMP09] ČECH J., MATAS J., PERDOCH M.: Efficient sequential correspondence selection by cosegmentation. *IEEE transactions on pattern analysis and machine intelligence* (2009), 1568–1581. 5

[CWS*08] CASATI B., WILSON L., STEPHENSON D., NURMI P., GHELLI A., POCERNICH M., DAMRATH U., EBERT E., BROWN B., MASON S.: Forecast verification: current status and future directions. *Meteorological Applications 15*, 1 (2008), 3–18. 1, 2

[DBB06] DAVIS C., BROWN B., BULLOCK R.: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Monthly weather review 134*, 7 (2006), 1772–1784. 2

[GAW*11] GLEICHER M., ALBERS D., WALKER R., JUSUFI I., HANSEN C. D., ROBERTS J. C.: Visual comparison for information visualization. *Information Visualization 10*, 4 (2011), 289–309. 2

[HB03] HARROWER M., BREWER C.: Colorbrewer. org: an online tool for selecting colour schemes for maps. *Cartographic Journal, The 40*, 1 (2003), 27–37. 5

[HHWN02] HAVRE S., HETZLER E., WHITNEY P., NOWELL L.: Themeriver: Visualizing thematic changes in large document collections. *IEEE transactions on visualization and computer graphics* (2002), 9–20. 2, 8

[LK10] LAKSHMANAN V., KAIN J.: A gaussian mixture model approach to forecast verification. *Weather and Forecasting 25*, 3 (2010), 908–920. 1, 2, 3, 4, 5, 7

[PP97] PAGENDARM H., POST F.: Studies in comparative visualization of flow features. *Scientific Visualization, Overviews, Methodologies, and Techniques* (1997), 211–227. 3

[RFF*08] ROBERTSON G., FERNANDEZ R., FISHER D., LEE B., STASKO J.: Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics* (2008), 1325–1332. 9

[RHRP98] ROBERTS S., HUSMEIER D., REZEK I., PENNY W.: Bayesian approaches to Gaussian mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20*, 11 (1998), 1133–1142. 4

[SP98] SHEN Q., PANG A.: Data level comparison of wind tunnel and computational fluid dynamics data. In *Proceedings of IEEE Visualization'98* (1998), pp. 415–418. 2

[SWB89] STANSKI H., WILSON L., BURROWS W.: *Survey of common verification methods in meteorology*. World Meteorological Organization, 1989. 1

[Tuf83] TUFTE E.: *The visual display of quantitative information*. Graphics Press, 1983. 8

[VVK03] VERBEEK J., VLASSIS N., KROSE B.: Efficient greedy learning of Gaussian mixture models. *Neural Computation 15*, 2 (2003), 469–485. 4

[WCZ*11] WANG Y., CHEN W., ZHANG J., DONG T., SHAN G., CHI X.: Efficient volume exploration using the gaussian mixture model. *IEEE Transactions on Visualization and Computer Graphics 17*, 11 (2011), 1560–1573. 3

[Wil95] WILKS D.: *Statistical methods in the atmospheric sciences*. Academic Press, London, 1995. 2, 3

[Wu10] WU R.: Subseasonal variability during the South China Sea summer monsoon onset. *Climate dynamics 34*, 5 (2010), 629–642. 3

[XL07] XUE J., LIU Y.: Numerical weather prediction in China in the new century – progress, problems and prospects. *Advances in Atmospheric Sciences 24*, 6 (2007), 1099–1108. 3