CrossMark

Uncertainty visualization for variable associations analysis

Huijie Zhang[1] · **Dezhan Qu**[1] · **Quanle Liu**[1] · **Qi Shang**[1] ·
Yafang Hou[1] · **Han-Wei Shen**[2]

Abstract Uncertainty is inevitable in scientific simulations. As the increase in computing power, ensemble data have been generated for multiple variables. Uncertainty has become a great challenge to the analysis of variable associations for multivariate ensemble data, as the variable associations are very complex and diverse among different ensemble members. In this paper, we propose a novel visualization method to present the uncertain associations between a reference variable and the associated variable for multivariate ensemble data. Considering the huge scale of original ensemble data, Gaussian mixture model (GMM) is exploited to quantify the uncertainty and represent the original data compactly. To reveal the spatial uncertainty of the reference variable, a GMM-based method for extracting uncertainty isosurface is proposed and shows the accuracy advantage over Gaussian-based method. Meanwhile, a data reduction method is proposed to enhance the performance of extracting uncertainty isosurface. By mapping the values of the associated variable onto the uncertainty isosurface of the reference variable, a syncretic rendering method is proposed to show the variable associations intuitively. Besides, the screen space accumulating strategy is introduced to present the uncertainties of the associations. Furthermore, we provide a switchable view for users to obtain the credibility of variable associations. The credible associations can assist users to make reliable decisions. For the regions with not credible associations, the detailed information of the associations in every ensemble member can be explored through animation for further analysis. The effectiveness of our method is demonstrated by synthetic, climate and combustion data sets.

✉ Huijie Zhang
  zhanghj167@nenu.edu.cn

  Dezhan Qu
  qudz862@nenu.edu.cn

  Quanle Liu
  liuql402@nenu.edu.cn

  Qi Shang
  shangq471@nenu.edu.cn

  Yafang Hou
  houyf397@nenu.edu.cn

  Han-Wei Shen
  hwshen@cse.ohio-state.edu

1 School of Computer Science and Information Technology, Northeast Normal University, Changchun, People's Republic of China

2 GRAVITY Group, Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, USA

Keywords Uncertainty visualization · Gaussian mixture model · Uncertainty isosurface · Variable associations analysis

1 Introduction

When performing complex scientific experiments or simulations, uncertainty is inevitable and affects the analysis of scientific data in many applications. At present, the most common method for reducing the influence of this kind of uncertainty is producing ensemble data, which performs a set of simulations with different model parameters. Ensemble data are often multivariate, such as climate data and Weather Research and Forecasting (WRF) data that both have several variables. The analysis of variable associations plays an important role in understanding the multivariable data. The variable associations of single-value multivariate data

sets have been well studied [1,2]. However, most of the correlation visualization methods still ignore the uncertainty. Actually, uncertainty is very necessary to be considered for the visualization and analysis of variable associations in ensemble data, because different simulations in ensemble data can result in different association patterns between two variables. Moreover, the opposite conclusions can even be drawn when the difference between the association patterns of different ensemble members is very large. Therefore, it is difficult to make reliable decisions from the complex and uncertain association patterns.

Due to the spatial heterogeneity, different spatial regions usually lead to different association patterns, and it is necessary to visualize the relationship patterns in space. This means that the associations need to be presented together with the intrinsic spatial information. However, this results in difficulties when showing the uncertainty information since a new visual dimension is usually needed. Therefore, how to visualize and analyse the uncertain associations between different variables in ensemble data is a non-trivial problem.

To address this challenging problem, in this paper, we propose a novel visualization method that can effectively analyse the uncertain associations between a scalar value of a reference variable and the associated variable in uncertain field. The associations between the specific scalar values and variables are usually very concerned by scientists. To show the spatial uncertainty of the scalar value of the reference variable, uncertainty isosurface where the scalar value may exist is extracted. Compared with visualizing the whole volume space, uncertainty isosurface can usually denote some salient features or interesting regions. To understand the associations between the reference variable and the associated variable, the uncertain values of the associated variable are mapped onto the uncertainty isosurface of the reference variable. Then, the associations between different variables of ensemble data can be intuitively recognized. Furthermore, uncertainty measurement is utilized to evaluate the credibility of the variable associations for different regions. For the credible regions, scientists can make predictions directly. For the regions with not credible associations, animation is provided for scientists to explore the variable associations of each ensemble member in detail.

For the representation of uncertainty in ensemble data, modelling distributions for each point using the values in different simulations is a prevalent method in recent years. Generally, each point in ensemble data is considered as a random variable. Then, the discrete scalar values can be represented by a Probability Density Function (PDF) continuously. Gaussian distribution [3–5] has been widely used in uncertainty visualization for science and engineering simulations, due to its fast computation with little space cost. However, for some fields that the data do not conform to Gaussian naturally, more flexible models need to be utilized

to better fit the distributions. Some nonparametric models such as histogram and Kernel Density Estimation (KDE) have been used [6–8]. However, since the space costs of nonparametric models are huge, these methods cannot be applied for large-scale data sets. GMM is a relatively compact parametric model that can fit a wide range of distributions and has been used for uncertainty volume rendering [9]. Hence, GMM is exploited in this paper to represent the uncertainty of both the reference variable and the associated variable.

For the extraction of uncertainty isosurface, we propose a method inspired by the Probabilistic Marching Cubes (PMC) method proposed by Pöthkow et al. [10]. The advantages of PMC are that it can quantize the uncertainty by crossing probability values and make the uncertainty isosurface more accurate through considering spatial correlation as well. However, this method uses Gaussian to model the distributions, which cannot be applied to all kinds of data sets. Furthermore, PMC method is of low efficiency for large-scale data sets, because it uses a time-consuming algorithm for all the data serially. In this paper, on the basis of preserving the advantages of PMC, we model the uncertainty information with GMM to get more accurate results for synthetic and real ensemble data sets. To address the performance issue, a data reduction method is suggested to significantly decrease the processed data size and the parallel algorithm is designed to improve computation efficiency. Hence, our method is capable of extracting the uncertainty isosurface for large-scale data sets.

For mapping the associated variable onto the uncertainty isosurface of the reference variable, we propose a syncretic rendering method to integrate the uncertainty information of the reference variable and the associated variable. The uncertain values of the associated variable are blended using the screen space accumulating strategy [9]. The opacity of transfer function is set as the crossing probabilities of the reference variable. In this way, the regions where the associations more likely to exist are highlighted, and users will be able to easily identify the important association patterns through visual perception. This screen accumulating view can provide an overview of the variable associations.

For the exploration of variable associations, a switchable view composed of a mean view and a standard deviation view is provided to help users interactively observe the variable associations and their corresponding credibility. Animation is provided to understand the details of ensemble members in the regions with relatively uncertain associations. The effectiveness and usefulness of this approach is demonstrated by analysing three multivariate ensemble data sets.

In summary, the contributions in this work are threefold:

1. We propose a data reduction method for extracting the uncertainty isosurface. Through the data reduction and

parallel implementation, uncertainty isosurface can be extracted for large-scale data sets efficiently.

2. Multidimensional GMM considering spatial correlation is exploited in the extraction of uncertainty isosurface to improve the accuracy of uncertainty isosurface. Compared with Gaussian, GMM can better conform to the real distributions of simulations. The uncertainty isosurface extracted by the GMM-based method is more accurate and reliable than the uncertainty isosurface extracted by Gaussian-based method.

3. We propose a visualization method to present variable associations in multivariable uncertain field. The general association patterns among different ensemble members can be intuitively shown, which provides scientists an overall understanding of the variable associations. The credibility of the associations can be shown at the same time, which can help scientists draw reliable conclusions in scientific experiments. The associations that are not credible enough can be further analysed by browsing the associations among different ensemble members.

The rest of this paper is organized as follows: firstly, the related research works are summarized in Sect. 2. Then, the overview of methods and the workflow are given in Sect. 3. The details of our methods are demonstrated in Sects. 4, 5 and 6. Results are illustrated in Sect. 7. The selections of isovalues, the parameters setting, the performance and the evaluation of our method are discussed in Sect. 8, followed by the conclusion and future work in Sect. 9.

2 Related work

2.1 Uncertainty quantification

The uncertainty caused by the different results of multiple runs in an experiment can usually be represented by probability density function statistically [11,12]. As a concise model, Gaussian model was generally used in the uncertainty visualization for diverse applications [3–5]. However, for those simulations that do not fit Gaussian distribution, the accuracy of these methods will decrease. Liu et al. [9] took advantage of GMM which can compactly model relatively complex distributions to summarize the large ensemble data for volume visualization. Pöthkow and Hege [7] compared the effects of the nonparametric models with Gaussian when extracting the uncertainty isocontours, including empirical distribution, histogram and kernel density estimation. Nonparametric models were observed with good feasibility for various data sets. However, the storage costs are expensive. To balance the accuracy and storage cost, GMM is exploited in this paper to represent the uncertainty in the ensemble data.

2.2 Visualization of uncertainty

Uncertainty visualization is one of the most challenging topics in scientific visualization. Uncertainty will bring troubles to the visualization, especially for 3D or higher dimensional data sets. This is because the uncertainty information usually needs to be encoded by another visual dimension. Various visualization methods have been proposed to show the uncertainty. In 1997, Pang et al. [12] summarized the early methods of uncertainty visualization. In recent years, the visualization of uncertainty has been concerned by increasing number of researchers. The up-to-date overviews of uncertainty visualization given by Bonneau et al. [13] and Brodlie et al. [14] described the taxonomy of uncertainty and the state-of-the-art visualization techniques.

Animation has been utilized to convey the uncertainty in several works [15–17]. Regions with high frequency of variations usually have high uncertainties and should be further explored. However, the flickers can exist in different parts of the image. Therefore, it is not easy to catch the overall uncertainty in ensemble data.

In order to overcome the limitations of animation, Hengl [18] encoded the uncertainty using HIS colour space, in which the luminance was determined according to the uncertainty. However, HIS colour space has too few available colours to encode the variations in detail. For this issue, high dynamic range (HDR) volume rendering was used by Dinesha et al. [19] for uncertainty visualization. Liu et al. [9] presented a screen space integration technique to show a fuzzy rendering result that took uncertainty into account.

Taking advantage of opaqueness is another effective way to demonstrate the uncertainty. In the works of Pöthkow et al. [10,20], the positional uncertainty of isosurface was conveyed by combing the opaque isosurface of the ensemble mean with the ray-casting result of uncertainty isosurface. Glyph [21,22] is commonly used to encode the uncertainty information, because it can show the multiple information simultaneously. Recently, Hao et al. [23] utilized glyph-based visualization to visually compare the data across multiple ensemble members and explore the temporal ensemble data sets.

In this paper, we introduce the screen space integration scheme [9] and propose a transfer function setting strategy to give an overview of the associations between different variables among all ensemble members.

2.3 Uncertainty isocontouring

Extraction of uncertainty isocontours is an effective way to visualize the uncertainty features. Approaches of uncertainty isocontouring can be classified to the value uncertainty and the positional uncertainty [14]. The value uncertainty is often indicated by combing the mean contour with the metaphor

of uncertainty. The positional uncertainty usually shows the possible positions where contours may exist. Sanyal et al. [24] described an uncertainty ribbon, the mean contour lines with thickness, to indicate the uncertainty. A probability-based method of extracting uncertainty isocontours was proposed by Pöthkow and Hege [20]. This method assumed that the voxels are independent of each other, but this often does not conform to the real case. By further considering the correlation in the data, more reliable uncertainty isosurfaces have been obtained in ensemble data sets [10,25]. These methods are all based on Gaussian distribution. Athawale et al. [8,26] analysed the positional uncertainty of the isocontours using nonparametric models. Nevertheless, the storage cost is huge, and the computation complexity is relatively high. In this paper, we propose an efficient GMM-based method which can support the analysis of nearly arbitrary ensemble data sets with low memory cost.

2.4 Visualization of variable associations

Analysis of variable associations is an essential task in the visualization of multivariate data. Gosink et al. [27] mapped a correlation field onto an isosurface to explore multivariate data sets. Guo et al. [28] designed a multidimensional transfer function, which could intuitively show the associations between different variables, through embedding scatter plots projected by multidimensional scaling (MDS) into the parallel coordinates plot (PCP). Biswas et al. [29] proposed an exploration framework for multivariate data by utilizing the metrics in information theory. Their method can analyse the relationship between the scalar value of a selected variable and another variable. Recently, informativeness and uniqueness metrics were introduced by Liu and Shen [30] to measure the information flow and explore the associations between scalar values of different variables. Zhang et al. [31] proposed a correlation metric for the voxels in multivariate time-varying data. In their work, the correlation patterns have been revealed by considering the time-varying trend of multiple variables and the information of spatial correlation.

However, the above works are focused on the single-value multivariate data. For ensemble data, only the spatial correlations between voxels were studied by Pfaffelmoser et al. [32] while studies on visualization methods of associations between different variables of ensemble data are rare. In our work, the uncertainty of variable associations is considered and is analysed through several views.

3 Overview

In this section, we provide an overview of our approaches. In order to analyse the uncertain associations between the selected reference variable and an associated variable, we

extract uncertainty isosurface of the reference variable and colour it with uncertain values of the associated variable. For the reference variable, the uncertainty isosurface is extracted to get the regions focused by users. According to the user's interest, the isovalues can be selected in the field of ensemble mean for reference variable. In order to consider the spatial correlation for extracting uncertainty isosurface, every 8 adjacent points in a cube are taken as a cell. For each cell of the reference variable, multidimensional GMM is used to model the distributions of ensemble members. Through Monte Carlo sampling, the probability of crossing the isosurface for each cell can be obtained. To enhance the performance, we perform a data reduction that can cut the cells have zero or very little probabilities of crossing the isovalue. For the associated variable, per-voxel GMM is used to model the distribution of ensemble members. Through Monte Carlo sampling based on GMM and our syncretic rendering method, the screen accumulating view presents the overview of uncertainty associations between the associated variable and the reference variable's scalar value. To support the in-depth analysis of uncertain associations, we provide a switchable view composed of the mean field and the standard deviation field for the associated variable on the uncertainty isosurface to observe the credibility of the variable associations. Furthermore, animation is utilized to reveal the details of uncertainty information. Through the proposed methods, the variable associations in uncertain field and the credibility of the associations can be intuitively shown. The workflow of the uncertainty visualization for variable associations is presented in Fig. 1.

4 Uncertainty quantification

Since GMM can approximate most of distributions in a compact way, we exploit it to quantify the uncertainties of ensemble data. For the reference variable, we use GMM to model the multidimensional distribution of each cell, and further compute the crossing probability for each cell using Monte Carlo sampling. For the associated variable, GMM is used to fit the distribution of each voxel and colour the uncertainty isosurface of the reference variable by Monte Carlo integrating in screen space.

Gaussian mixture model is a commonly used parametric probabilistic model. It is composed of several Gaussian distribution components combined through a linear weighted sum to approximate the entire distribution. The PDF approximated by GMM with K components is defined as:

$$p(x) = \sum_{i=1}^{K} \pi_i N(x|\mu_i, \Sigma_i), \tag{1}$$

in which μ_i and Σ_i denote the mean and covariance matrix for each component i and π_i is the weight of compo-

Fig. 1 Workflow of uncertainty visualization for variable associations
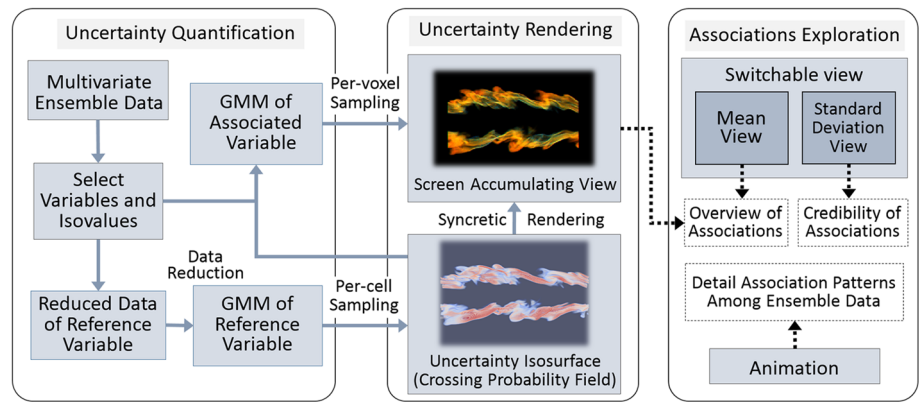
nent i. Usually, expectation–maximization (EM) algorithm is exploited to compute these parameters. Given the initial parameters, a value of the likelihood function can be computed, and the parameters can be updated iteratively. Iterations are stopped when the parameters maximize the likelihood function.

A good initialization of parameters can accelerate the iterating process of EM algorithm. A common-used initialization method is using the fast K-means algorithm to divide the data into K clusters in advance. Then, the mean and covariance matrix of each cluster's data are used as the initial parameters of each Gaussian component. The number of data in each cluster is used to compute the initial weight of each component. Performing the initialization for multiple times can also avoid trapping in local optimum during the computation of EM algorithm.

GMM is a compact and effective way to model almost arbitrary distributions. Compared with Gaussian, GMM can better fit the complex multimodal distribution, which is illustrated in Fig. 2. As a compact form, GMM stores a distribution with only $K \times 3$ parameters for univariate data, and the storage cost is significantly reduced by comparing with the original ensemble data.

In order to best fit GMM with the real distributions using least storages and computation time, Bayesian information criterion (BIC) is employed to determine the number of components. BIC can evaluate the performance of the model for fitting the data distribution. The lower BIC score means that the model better approximates the data distribution with lower possibility of over-fitting. The BIC scores are computed for GMM with different component numbers from 1 to 4 in the pre-processing, and perform GMM fitting with the component number corresponding to the minimum BIC score.

5 Extraction of uncertainty isosurface

In this section, we present the extraction of uncertainty isosurface for the reference variable. Since ensemble data are

very large and the computations will be time-consuming, we first describe the data reduction of the reference variable according to the isovalue for increasing the efficiency. Then, we describe how the uncertainty is modelled by multi-dimensional GMM and present the computation of crossing probability field. Finally, the results of our method are compared with the results of the ensemble mean and PMC method [10] using a synthetic data set.
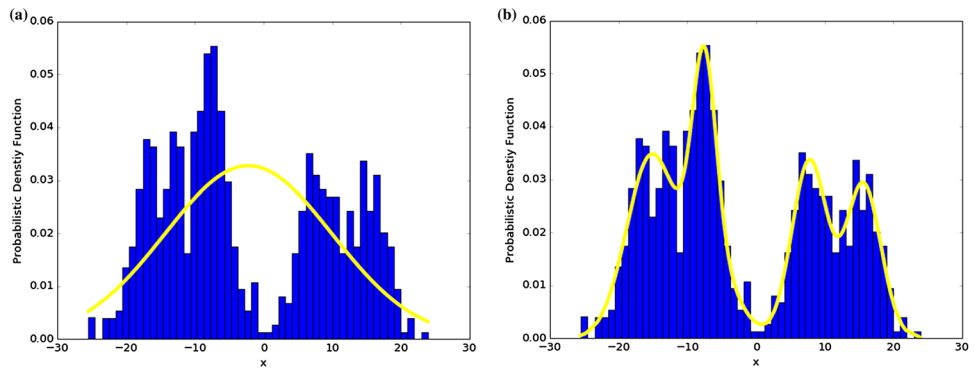
5.1 Data reduction

We define that the ensemble volume data have m ensemble members and each ensemble member has n voxels. The sizes of spatial dimensions are, respectively, Dim_x, Dim_y and Dim_z. Similar to PMC [10] method, the 8 neighbouring points in a cube is built as a cell. Obviously, ensemble data have a large data size that is usually difficult to be loaded in memory directly. Moreover, the multidimensional GMM fitting for all the cells is also a relatively time-consuming process. Therefore, data reduction of the original ensemble data is necessary to accelerate the computation process fundamentally.

By comparing the given isovalue v_{iso} with the data values in each cell, we cut the cells that are impossible to intersect with the uncertainty isosurface. For ease of exposition, if the cell meets the constraint conditions, we reserve it and call it a valid cell. Those cells that can be reduced are called as invalid cells.

Two judgements are proposed to identify the valid cells. Firstly, the value range R_{ensemble} between the maximum value and minimum value of all ensemble members in a cell is used to test whether the cell is certainly reserved. For the cell whose value range contains the isovalue, it is directly reserved as a valid cell. However, the ensemble members cannot represent all the possible cases of ensemble data; hence, the valid cells filtered only by R_{ensemble} cannot cover all the cells that the isosurface might cross. For the cell whose value range does not contain the isovalue, if the isovalue locates in a range that is wider than R_{ensemble}, it should also be seen as a valid cell.

Fig. 2 Comparison between Gaussian and GMM for fitting a multimodal distribution. **a** Approximation using Gaussian. **b** Approximation using GMM with 5 Gaussian components

▲ Maximum and minimum of cell's value range

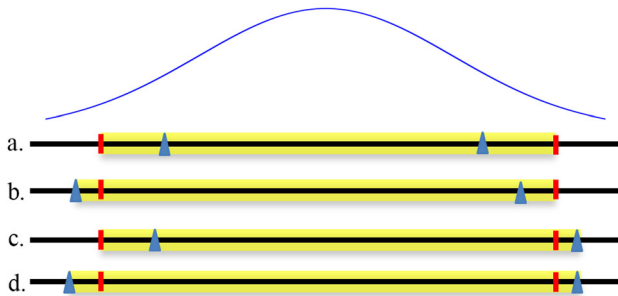❚ Boundaries of range $[\mu - 3\sigma, \mu + 3\sigma]$

Fig. 3 Four possible cases that a cell will be reserved

To increase the accuracy of data reduction, the three sigma rule is introduced to further judge the validity of the indeterminate cells. The three sigma rule is that for a variable that accords with normal distribution, approximately 99.7% of the values locate in the three standard deviations of the mean. The values outside the range of three sigma can be regarded as anomalies. It is a common way to identify the abnormal values for a normal distribution in statistics. If the isovalue locates in the range $[\mu - 3\sigma, \mu + 3\sigma]$ of a cell, the cell will be reserved. Otherwise, the cell will be cut. Through the three sigma rule, most of the cells ignored by the value range R_{ensemble} will be reserved. Moreover, the cells with very low crossing probabilities caused by the abnormal values can also be removed.

All in all, a cell will be reserved only if the isovalue is between the maximum value and the minimum value or locates in the range of three sigma. As shown in Fig. 3, the cell is reserved if the isovalue v_{iso} locates in the ranges represented by the yellow bars, and the value range of ensemble members can be extended in the cases 'a', 'b' and 'c' by introducing the three sigma rule.

Usually, a large amount of cells can be reduced, because the valid cells of uncertainty isosurface usually occupy a small part of the original data. According to the experiments, for most cases, over 70% of all the cells can be reduced. Compared with the method of processing all the original cells, the efficiency of our method is highly increased.

The accuracy of the data reduction method is tested on multiple data sets, through comparing the filtered valid cells with the cells that have nonzero crossing probabilities computed by original data. More than 85% of the cells with nonzero crossing probabilities can be reserved. Moreover, all those lost cells have very low crossing probabilities. If more accurate results are required, the range $[\mu - 5\sigma, \mu + 5\sigma]$ can be used to get a more accurate result with a relatively lower simplification rate. The accuracy rates and the simplification rates of the experiments are listed in Sect. 8.2.

The per-cell GMM modelling and the extraction of uncertainty isosurface are performed only for the valid cells. The crossing probabilities of those invalid cells are set to 0 in the probabilistic crossing field.

5.2 Uncertainty isosurface

After the data reduction, the BIC value is computed for each valid cell using the number of Gaussian components varying from 1 to 4. Then, the number of Gaussian components accorded with the minimum BIC value is recorded for each cell. The number of ensemble members is m, thus each corner point in a cell can be regarded as a random variable that has m values. Multidimensional GMM is modelled for each valid cell based on the joint distribution of all ensemble members on its 8 corner points. This means that the 8 corner points are considered as 8 variables, and therefore, the weights, the mean vectors and the covariance matrices for each cell are obtained through GMM fitting. In this way, the correlation of the 8 adjacent voxels in the cell is contained in the covariance matrix of each Gaussian component in GMM.

Since the potential uncertainty information is very difficult to be represented by the limited m ensemble members, Monte Carlo sampling is carried out for each cell based on the multidimensional GMM. Similar to PMC method [10], each sample of a cell will be determined whether the isosurface cross the cell using symmetry-reduced marching cubes method. As illustrated in Fig. 4, only if v_{iso} is bigger than the maximum value or less than the minimum value, the cell
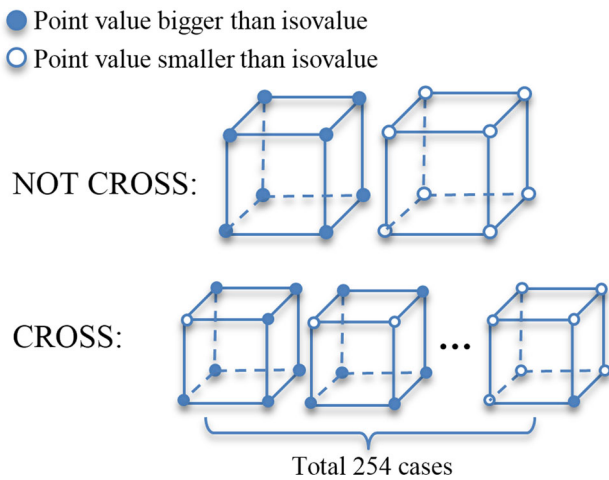
Fig. 4 Illustration of crossing judgement

Algorithm 1 Computing Crossing Probability for Each Cell by Crossing Judgement and Monte Carlo Integration

Input: Sample number N, GMM composed of K components and isovalue v_{iso}
Output: Crossing probability of the cell
1: For each Gaussian component i from 1 to K of GMM, do:
      1.1 Initialize the crossing number m_i of component i into 0 and current step count S into 0.
      1.2 Compute the value range R of the 8 values in the current sample.
      1.3 If the isovalue falls in the range R, then add 1 to the crossing number m_i.
      1.4 Update step count S with $S + 1$.
      1.5 If current step count S is little than N, then to step 1.2.
2: Compute the crossing probability of the cell according to Equation 2.
3: Algorithm end.

is consider as not crossing with the isosurface. So long as the isovalue locates in the value range of the sample values for the cell, the cell is regarded as intersecting with the isosurface. This judgement is simple, but it is accurate enough for the extraction of uncertainty isosurface in data with high resolution.

Different from the PMC method that serially performs the Monte Carlo sampling based on Gaussian, our method uses GMM and implement the process in parallel to obtain higher accuracy and higher efficiency. Since GMM in each cell has different Gaussian components, it is difficult to handle each cell with the same operation in the parallel implementation. To address this issue, the Monte Carlo sampling is performed for each Gaussian component in GMM with the same sampling number and obtains the final crossing probability by the weighted sum of the crossing probabilities for each Gaussian component. Given N samples for each Gaussian component, the crossing probability p for a cell is computed as:

$$p = \frac{\sum_{i=1}^{K} \pi_i m_i}{N} \tag{2}$$

in which K is the component number of GMM, π_i and m_i, respectively, denote the weight and the crossing number of component i. Algorithm 1 explains the process of computing the crossing probability for each cell. After all the cells are processed, the probabilistic crossing field is obtained to quantize the uncertainty. The whole process of our method is illustrated in Fig. 5. Volume rendering can be utilized to show the crossing probability field.

5.3 Comparison with PMC

A synthetic ensemble data set is exploited to demonstrate the advantage of the GMM-based method over the ensemble mean and the Gaussian-based method (PMC). The synthetic ensemble data are built based on the formula $v(x, y, z) = (\cos(7x) + \cos(7y) + \cos(7z))e^{-4.5d}$, $where \ d = \sqrt{x^2 + y^2 + z^2}$ [10]. Four Gaussian noises are added onto the formula to generate ensemble members. The means of Gaussians are shifted symmetrically around 0, and the variances are constant. Sixteen samples are generated for each Gaussian noise, and 64 ensemble members with resolution $128 \times 128 \times 128$ are obtained.

In the experiment of this synthetic ensemble data, the isovalue is set as 0.01215. Figure 6 illustrates the results of the ensemble mean, the Gaussian-based method and our GMM-based method for conveying the positional uncertainty of isosurface. The isosurface of the original synthetic data (computed by the formula) is presented in Fig 6a as the ground truth. Through comparing Fig. 6b with Fig. 6a, it is observed that the ensemble mean expresses little uncertainty and may lose important features, such as the connectivity of isosurface (displayed by the partial detail view on right). As shown in Fig 6c, d, distribution-based method can convey wider positional uncertainty of the isosurface.

Although the Gaussian-based method can present where the isosurface might exist, the corresponding probability values over the uncertainty isosurface are inaccurate. Gaussian distribution always has the highest probability density in the mean value but relatively lower probability densities for the values far from the mean. Hence, for the PMC method, the crossing probabilities are always the highest in the regions near the location of mean isosurface while lower as the positions are far away from the mean isosurface. As shown in the detailed view in Fig. 6c, the fuzzy outer regions of uncertainty isosurface indicate smaller crossing probabilities.

However, according to the added noises, the distribution for ensemble members of the cell in this synthetic data has relatively high probability densities for the values far from the mean. Hence, the true probability values in the outer regions of uncertainty isosurface should be higher than those computed by the PMC method. As for our approach, GMM

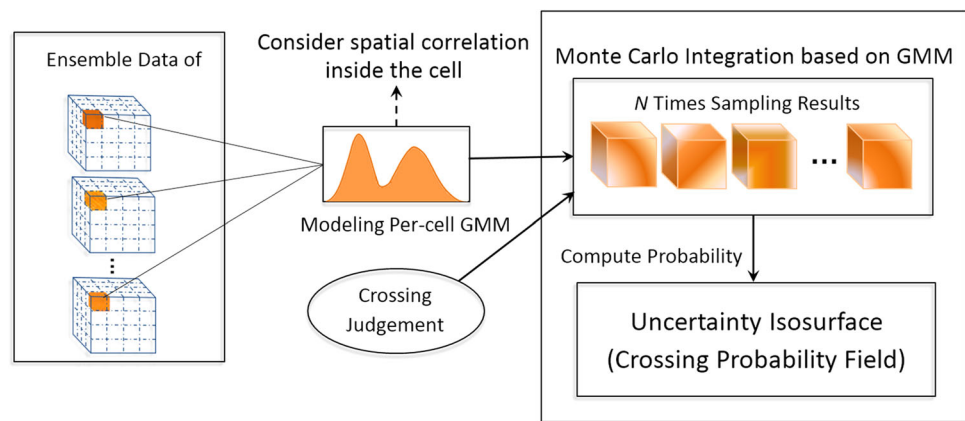**Fig. 5** Process of extracting uncertainty isosurface

Fig. 6 Comparison of uncertainty isosurfaces for synthetic data with isovalue 0.01215 extracted by the ensemble mean, Gaussian-based method and GMM-based method. **a** Ground truth isosurface of original synthetic data. **b** Isosurface of ensemble mean. **c** Uncertainty isosurface extracted by PMC method. **d** Uncertainty isosurface extracted by GMM-based method
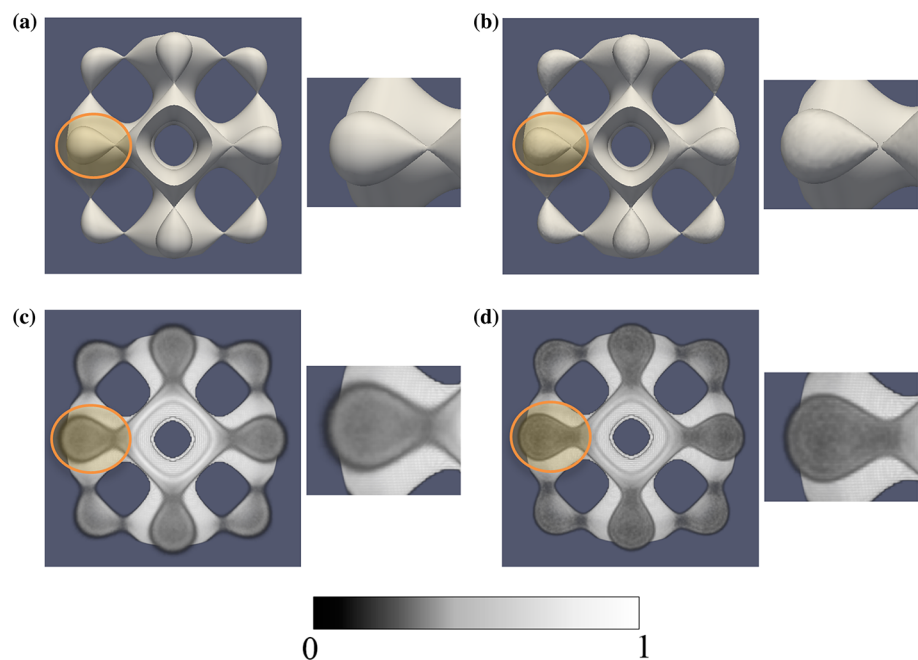
can precisely describe this kind of distributions and provide more reliable results. Compared with Fig. 6c, the uncertainty isosurface in Fig. 6d is clearer in the regions far from the isosurface of the ensemble mean and describes the crossing probabilities more accurately. Therefore, compared with the Gaussian-based PMC method, the GMM-based method can compute the probability field more precisely.

6 Visualization of variable associations

Generally, variable associations can be revealed by mapping the values of the associated variable to the isosurface of the reference variable. However, data uncertainty causes great difficulties in the visualization for ensemble data, because the variable associations of ensemble data are uncertain on space and also have different patterns among the ensemble members. In this section, we describe the visualization of

variable associations between the reference variable and the associated variable, which takes uncertainty into account. To reveal the variable associations, we present a syncretic rendering method that maps the uncertain values of the associated variable onto the uncertainty isosurface of the reference variable. The screen space Monte Carlo integrating strategy [9] is introduced to obtain an overview of the variable associations. Moreover, standard deviation of the associated variable is used to quantify the credibility of variable associations. Through using the syncretic rendering method, we also provide a switchable view and animation to explore the variable associations and the contained uncertainty.

6.1 Syncretic rendering

To handle the uncertainty of the associated variable, per-voxel GMMs are used to model distributions for the ensemble
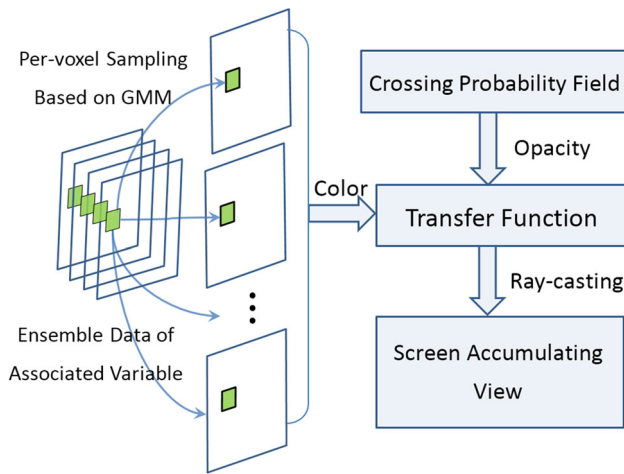
Fig. 7 Process of variable associations visualization

members of the associated variable's voxels that are corresponding to the cells with nonzero crossing probabilities. Then, more samples are generated based on the per-voxel distributions.

Considering the spatial uncertainty, the isosurface-crossing probabilities of the reference variable are used as the opacity and the values of the associated variable are used as the colour in transfer function to realize syncretic rendering. In this way, regions with higher opacities are emphasized to present more certain variable associations, whereas the positions with lower opacities show variable associations with less confidence and are visually weakened.

Considering the value uncertainty, the screen space Monte Carlo integrating strategy [9] is introduced to accumulate the syncretic rendering results of all the samples of the associated variable in screen space. This screen accumulating view is an uncertainty-aware image and can give users the overview of the uncertain variable associations. The volume rendering result of samples' mean will lead to strong colour contrast that may give users an incorrect visual classification of the uncertainty positions, whereas the uncertainty-aware result of the screen accumulating view can give users a more authentic recognition of the uncertain variable associations. The rendering process of the screen accumulating view is illustrated in Fig. 7.

Through this screen accumulating view, the association between the isovalue of the reference variable and the values of the associated variable can be generally revealed. If the association is complex, the colours of the visualization result are chaotic. If the colours of the visualization result are relatively consistent, an explicit association can be deduced. The credibility of the association patterns in uncertain field can be measured by the standard deviation values of the samples.

6.2 Uncertainty exploration

We provide users the switchable view and animation to support further explorations of the uncertain variable associations. The switchable view is made up by the mean view and the standard deviation view. The mean view maps the samples' mean of the associated variable onto the uncertainty isosurface and can support the interactive observation of the general variable associations. Since the result of screen accumulating view does not directly present the strength of uncertainty for the associated variable, standard deviation is utilized to measure the uncertainty of the associated variable. The standard deviation values for all samples of the associated variable are mapped onto the uncertainty isosurface of the reference variable. Then, this standard deviation view combines the positional uncertainty of isosurface for the reference variable with the value uncertainty of the associated variable. Through the standard deviation view, users can recognize the credibility of the associations in different regions. For the reference variable, the regions with high opacity can be seen as credible, because these regions have more possibilities to contain the specific isovalue. For the associated variable, the regions with low standard deviation values can be considered as credible, because the associations change very little on these regions among the samples of the ensemble members.

Generally, the credible associations should have low uncertainty for both variables. Users can draw reliable conclusions from these credible associations. For those regions with low credibility, animation is exploited to reveal the details of data variation among ensemble members of the associated variable. Through browsing the rendering result of each sample and focusing on the regions with not credible associations, users can detailedly understand the inner associations between different variables.

7 Results

In this section, we show the results obtained by our method in visualizing the uncertain associations between different variables, through three data sets.

7.1 Synthetic data set

In this case study, synthetic ensemble data are used to illustrate the effect of visualizing different association patterns and their corresponding credibility. The reference variable is generated as mentioned in Sect. 5.3. Since the uniform pattern and the confusion pattern are two typical cases focused by scientists, the other two artificial variables are generated to show the different association patterns. The associated variables have the same data size as the reference variable and

64 ensemble members. To present the confused association pattern, the volume spaces of the associated variables are divided into blocks with uniform resolution 4^3, and the values of data points in the same block are randomly generated from the same value range, whereas the data values in adjacent blocks are sampled from different value ranges. When creating the first associated variable $Var1$, data values are generated from range [0, 50] and [500, 550] for adjacent blocks alternately, and thus data of ensemble members in adjacent blocks have different means and similar variances. Whereas, for the second associated variable $Var2$, random numbers are produced, respectively, from range [−25, 25] and [−400, 400] for adjacent blocks, thus data of ensemble members in adjacent blocks have similar means and diverse variances.

After the data reduction with range $[\mu - 3\sigma, \mu + 3\sigma]$, we compute the uncertainty isosurface of the reference variable as shown in Fig. 6d. Then, through the syncretic rendering, the screen accumulating views and the standard deviation views that display the different association patterns and different credibility patterns are generated as shown in Fig. 8.

Since the values of the ensemble members for $Var1$ have big differences over space, it is observed from Fig. 8a that the screen accumulating view of $Var1$ has confused colours. Hence, it can be inferred that $Var1$ is chaotic over the uncertainty isosurface, and this pattern usually corresponds to some complex phenomena in scientific data sets, such as simulations of combustion and climate. As shown in Fig. 8b, the colours in the standard deviation view are uniform, and the standard deviation values are relatively low. This means that values of $Var1$ are stable among the ensemble members, so the associations presented in Fig. 8a are credible enough for users to draw conclusions. Besides, the regions marked by the yellow circles, respectively, in Fig. 8a, b have low opacities, so the variable associations have relatively low probabilities to exist in these regions.

The screen accumulating view of $Var2$ is shown in Fig. 8c where the colours are consistent. The standard deviation view of $Var2$ is shown in Fig. 8d where the standard deviation values are distributed over the uncertainty isosurface chaotically. This means that for different regions of the uncertainty isosurface, the credibility of associations is different, though the means for the samples of the associated variable are similar. For the regions with high standard deviation values, the complex information existed among ensemble members of the associated variable is not presented in the screen accumulating view as shown in Fig. 8c. This means that the variable associations shown in these regions are relatively not credible and need to be further explored by users. The associations on the regions with low standard deviation values can be considered as credible. Therefore, the proposed method is capable of presenting various kinds of variable associations and their corresponding credibility.

7.2 Climate data set

In this case study, we explore the variable associations in a climate data set which can be constructed as ensemble data. As we know, climate generally changes with the period of one year. Therefore, the climate of the same month in different years can be seen as the ensembles of a climate model. In this paper, the ECMWF ERA-20C data sets (http://www.ecmwf.int/en/research/climate-reanalysis/era-20c) is used. We combine the data of 31 days in May as a volume data. The data of 111 years (from 1900 to 2010) are adopted to analyse the uncertain associations between different variables. Therefore, 111 ensemble members with resolution $360 \times 181 \times 31$ are formed.

We choose the sea surface temperature (SST) as the reference variable, and the mean sea level pressure (MSLP) is chosen as the associated variable. These two variables are both important factors used to analyse the climate changes, such as El Niño phenomenon [33]. In order to explore the associations between these two variables in different geographical positions, we choose −24,000 and 18,000 as the isovalues for SST to get the uncertainty isosurfaces that are, respectively, located on cold polar regions and tropical regions. Figure 9a shows the volume rendering result of SST in year 1982. Since SST values change with the latitude in space, the uncertainty isosurface of SST −24,000 only occupy a little volume space as shown in Fig 9b and the simplification rate of data reduction can be more than 90%. Figure 9c, d, respectively, show the results of valid cells for isovalue −24,000 after the data reduction using range $[\mu - 3\sigma, \mu + 3\sigma]$ and range $[\mu - 5\sigma, \mu + 5\sigma]$. Compared with Fig. 9a, a large amount of data has been reduced. Even so, the reserved data can cover the uncertainty isosurface extracted without data reduction as shown in Fig. 9b. Therefore, our method of data reduction can highly enhance the performance of extracting the uncertainty isosurface and meanwhile guarantee the accuracy of uncertainty isosurface. The accuracy rate and simplification rate are discussed in Sect. 8.

As shown in Fig. 9b, the uncertainty isosurfaces of isovalue −24,000 are situated in the cold Polar Regions. By performing syncretic rendering, the screen accumulating view as shown in Fig. 10a is generated to show the overview patterns of associations between SST and MSLP. Compared with the mean view in Fig. 10b, the screen accumulating view shows more uncertainty information of MSLP. An interesting result can be observed in Fig. 10a that the MSLP values are relatively low in the regions of the South Pole, but are high in the Arctic Ocean region. Furthermore, the MSLP values distribute consistently in the Arctic Ocean region but distribute chaotically over the uncertainty isosurface in the South Pole. It can be inferred that regions of the Arctic Ocean are relatively homogeneous for MSLP and regions of the South

Fig. 8 Results of synthetic data set. **a** Screen accumulating view of $Var1$. **b** Standard deviation view of $Var1$. **c** Screen accumulating view of $Var2$. **d** Standard deviation view of $Var2$
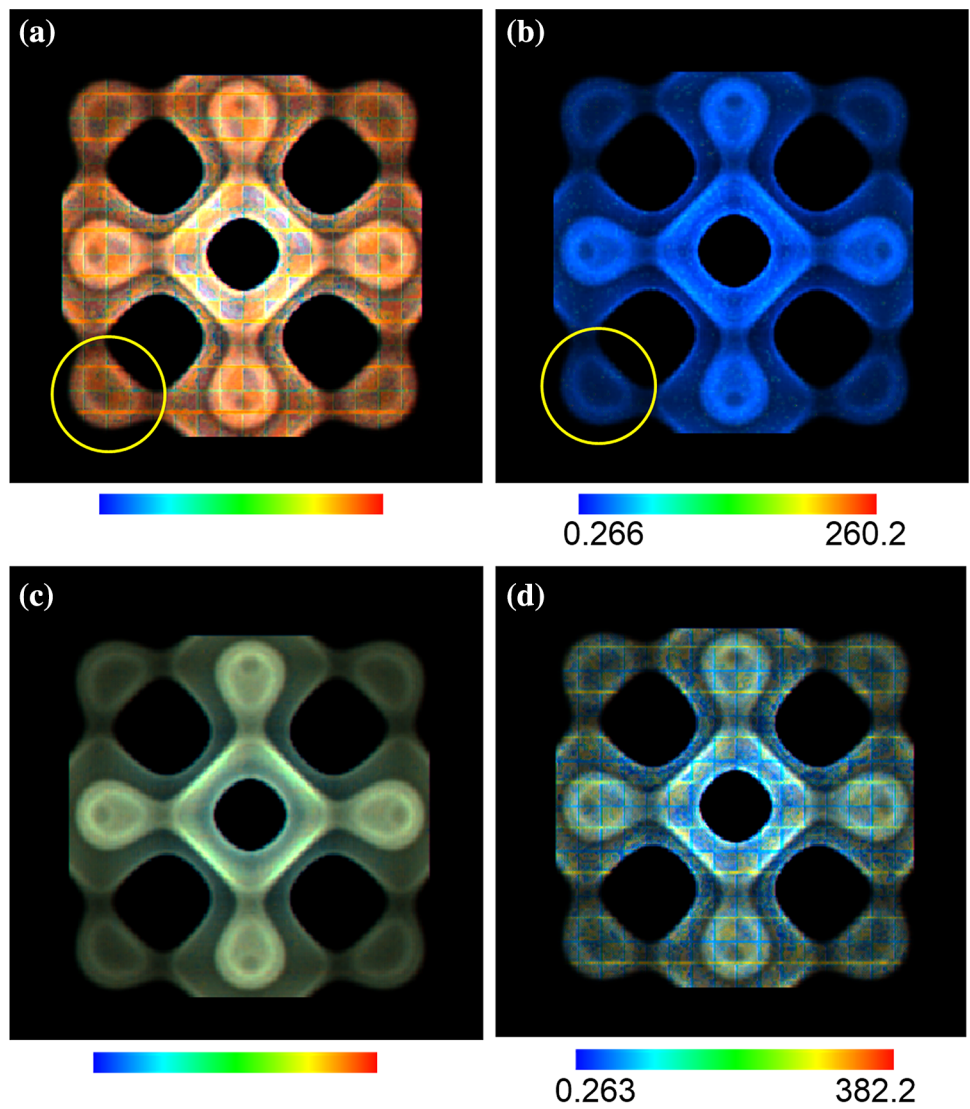

Fig. 9 Effect of data reduction. **a** Volume rendering result for year 1982 of SST. **b** Uncertainty isosurface of SST without data reduction when isovalue is −24,000. **c** Reserved data (112,416 cells) after data reduction with range $[\mu - 3\sigma, \mu + 3\sigma]$. **d** Reserved data (165,192 cells) after data reduction with range $[\mu - 5\sigma, \mu + 5\sigma]$
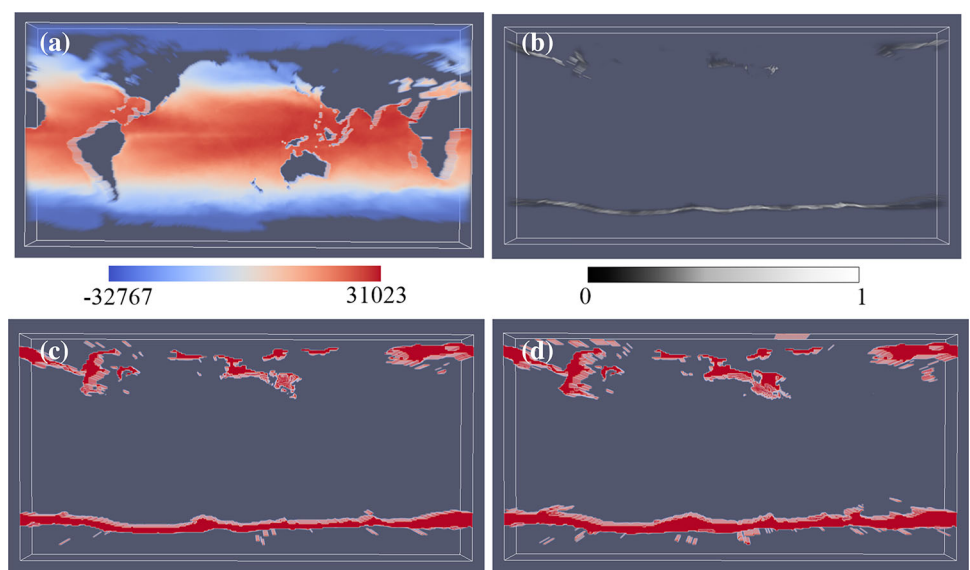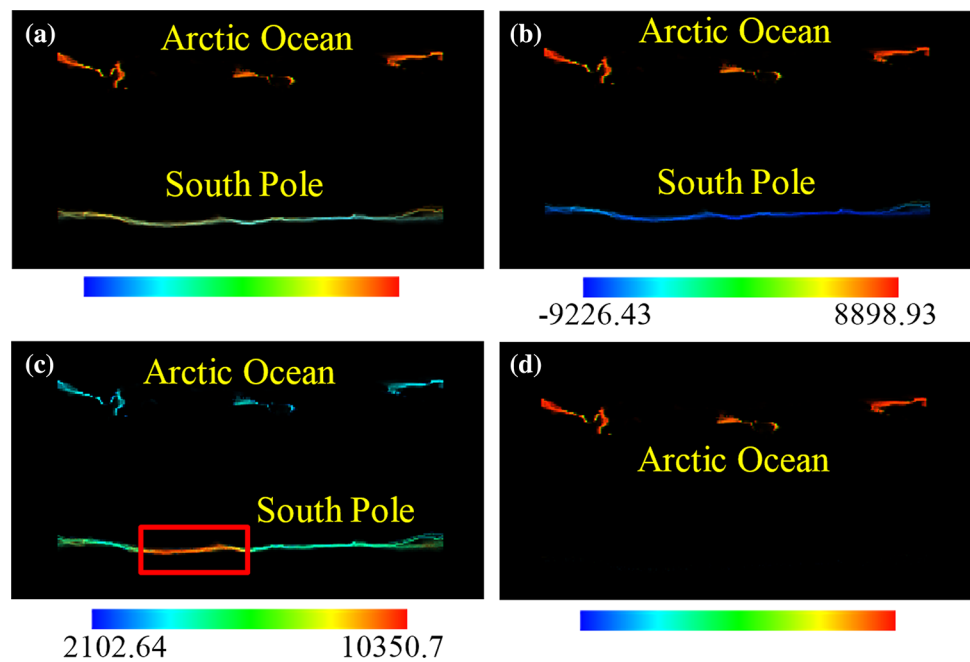
Fig. 10 Results of MSLP and SST with isovalue −24,000. **a** Screen accumulating view. **b** Mean view. **c** Standard deviation view. **d** Query result of the credible associations in mean view

Pole are relatively inhomogeneous for MSLP, when SST is −24,000. However, the inferring may be inaccurate, if the uncertainties of MSLP are high in the regions.

Therefore, we analyse the credibility of the variable associations through the standard deviation view. As shown in Fig. 10c, strong variations of ensemble members in the uncertainty isosurface of the South Pole region can be recognized, especially for the regions marked by the red box, whereas the values of MSLP in the Arctic Ocean region are stable. This means that MSLP values of the Arctic Ocean region are not strongly affected by the climate changes over a long period of time. Through analysing the uncertainty, the conclusion can be drawn that the association pattern presented on the Arctic Ocean is more credible than the association pattern presented on the South Pole.

For those regions with high uncertainty, the detail information in the ensemble members can be further explored through the animation. Fig. 11 presents several frames of the animation for the uncertain regions marked by the red box in Fig. 10c. It is not hard to see that the association patterns change strongly among these frames.

To obtain the regions with relatively credible variable associations, we provide a query function. We perform the query that standard deviations are less than $MinVar + 0.4 * (MaxVar - MinVar)$, where $MinVar$ and $MaxVar$ are, respectively, the minimum value and the maximum value of standard deviations. The query result of the mean view is shown as Fig. 10d, in which the regions of the Arctic Ocean are extracted.

We also explore the associations between SST and MSLP with isovalue 18,000. The simplification rate of data reduction for isovalue 18,000 is also more than 90%. Figure 12a

shows the uncertainty isosurface of isovalue 18,000 that is located on tropical regions. As shown in Fig. 12b, the accumulation values of MSLP are relatively high and are consistently distributed over the uncertainty isosurface. Compared with the mean view in Fig. 12c, the screen accumulating view is an uncertainty-aware result that has more meaningful colours. From Fig. 12d, it can be observed that the standard deviation values are low and consistent, and therefore, the association pattern presented in Fig. 12a is relatively credible. Therefore, the variable associations that MSLP values are high when SST value is 18,000 can be certainly obtained.

Through comparing the results of isovalue −24,000 with isovalue 18,000, we can know that different geographical locations can have different variable associations and different credibility.

7.3 Combustion data set

In this case study, we use a turbulent combustion simulation data that has 120 time steps with the resolution of $240 \times 360 \times 60$. Non-Gaussian noise with mean of 0 is added on the original values to generate 56 ensemble members for each variable in each time step, using the original values as ensemble mean. We apply our method to this turbulent combustion simulation data to analyse the uncertain associations between different variables across space and time.

The mixture fraction (MIX) and the heat release rate (HR) are, respectively, selected as the reference variable and the associated variable, because these two variables are very relevant. The mixture fraction denotes the ratio of fuel and oxidants, which usually indicates the regions of flame. The

Fig. 11 Six frames of the animation for the regions with high uncertainty when the reference variable SST is about −24,000 and the associated variable is MSLP
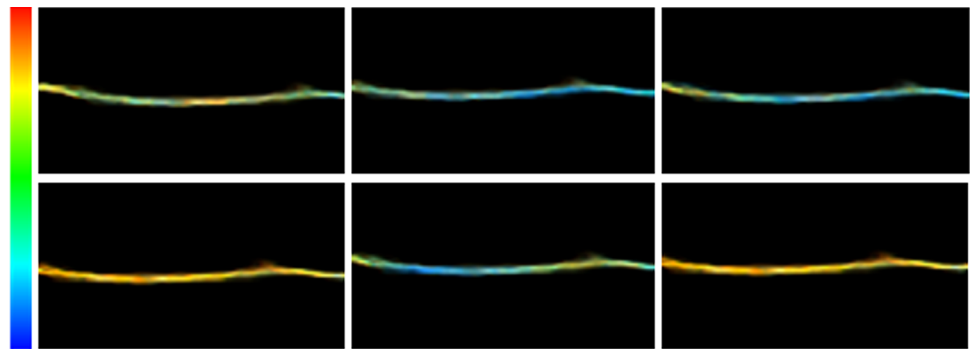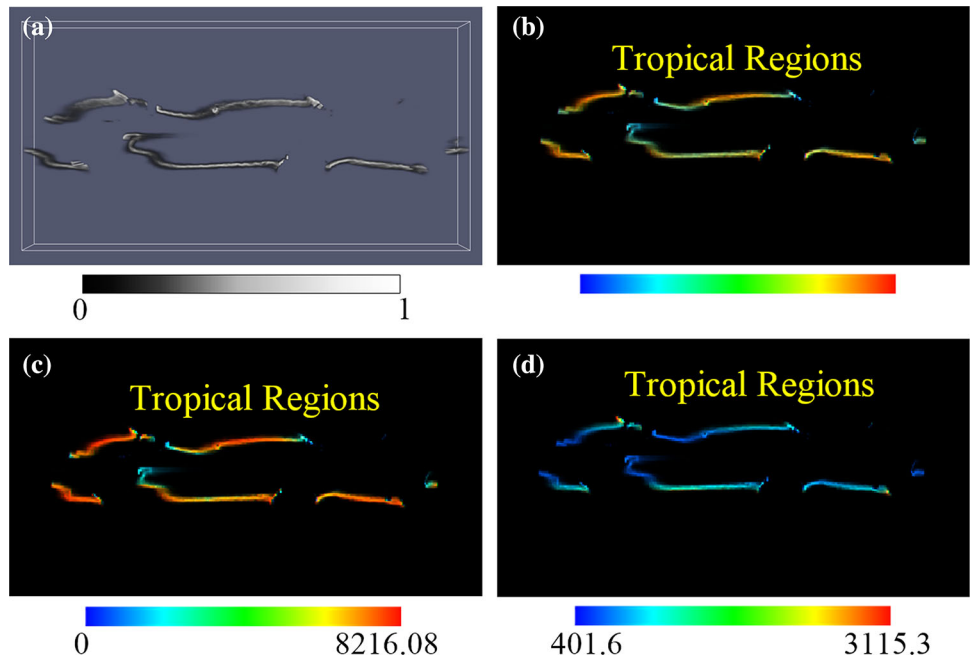

Fig. 12 Results of MSLP and SST with isovalue 18,000. **a** Uncertainty isosurface of SST. **b** Screen accumulating view. **c** Mean view. **d** Standard deviation view

heat release rate can denote the quantity of heat released during the combustion in unit time.

We perform data reduction for isovalue 0.96 using range $[\mu - 5\sigma, \mu + 5\sigma]$, and the simplification rate is 82.65%. By using the case of isovalue 0.96, we compare the uncertainty isosurface extracted by our method with the uncertainty isosurface extracted by PMC method. The isosurface of the original data is shown as Fig. 13a, which is very irregular and complex. This isosurface can be regarded as the ground truth. The distribution of the ensemble members in a voxel is shown as the histogram in Fig. 13b, which is a non-Gaussian distribution. As shown in Fig. 13b, Gaussian approximation has higher probability densities than GMM in the value range (marked by the skyblue boxes) that is near the mean of the voxel but lower probability densities in the value range (marked by the yellow boxes) that is far from the mean of the voxel. The uncertainty isosurfaces extracted by Gaussian and GMM are, respectively, shown as Fig. 13c, d. They are significantly different in the terms of crossing probabilities. The Gaussian-based method gets higher crossing probabilities than the GMM-based method in the regions marked by

the skyblue circles but relatively lower crossing probabilities in the regions marked by the yellow circles.

Since the mean of the noise distribution is zero, the mean of the ensemble members in each voxel is approximately equal to the original data value. Due to the spatial similarity that values are similar in the adjacent space, the means of the skyblue circle regions near the isosurface of ground truth are close to the isovalue 0.96. Therefore, the Monte Carlo sampling based on the distribution approximated by Gaussian will lead to more samples near the isovalue and higher crossing probabilities than GMM, which does not conform to the real distribution of ensemble members. Meanwhile, the sampling based on GMM can obtain more samples in the value ranges marked by the yellow boxes in Fig. 13b. For the yellow circle regions that are far from the isosurface of ground truth, the value ranges are relatively far from the mean value of the voxel and can be near the isovalue. Therefore, our GMM-based method obtains higher crossing probabilities than Gaussian-based method for the regions marked by the yellow circles.

Fig. 13 Comparison of uncertainty isosurfaces of combustion data with isovalue 0.96 extracted by the Gaussian-based PMC method and our GMM-based method. **a** Ground truth isosurface of original data. **b** PDFs of Gaussian fitting and GMM fitting for the ensemble members of a voxel. **c** Uncertainty isosurface extracted by PMC method. **d** Uncertainty isosurface extracted by our method
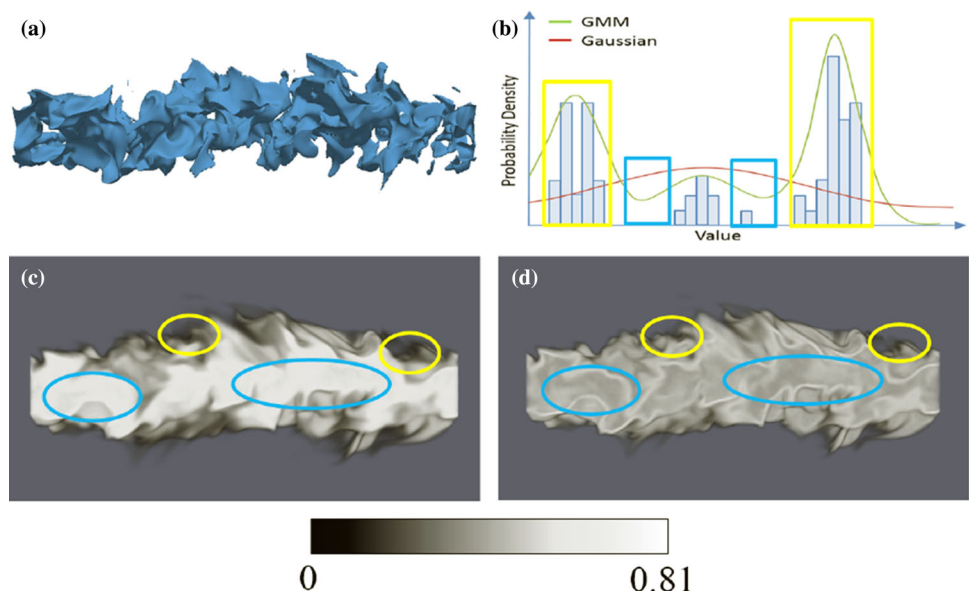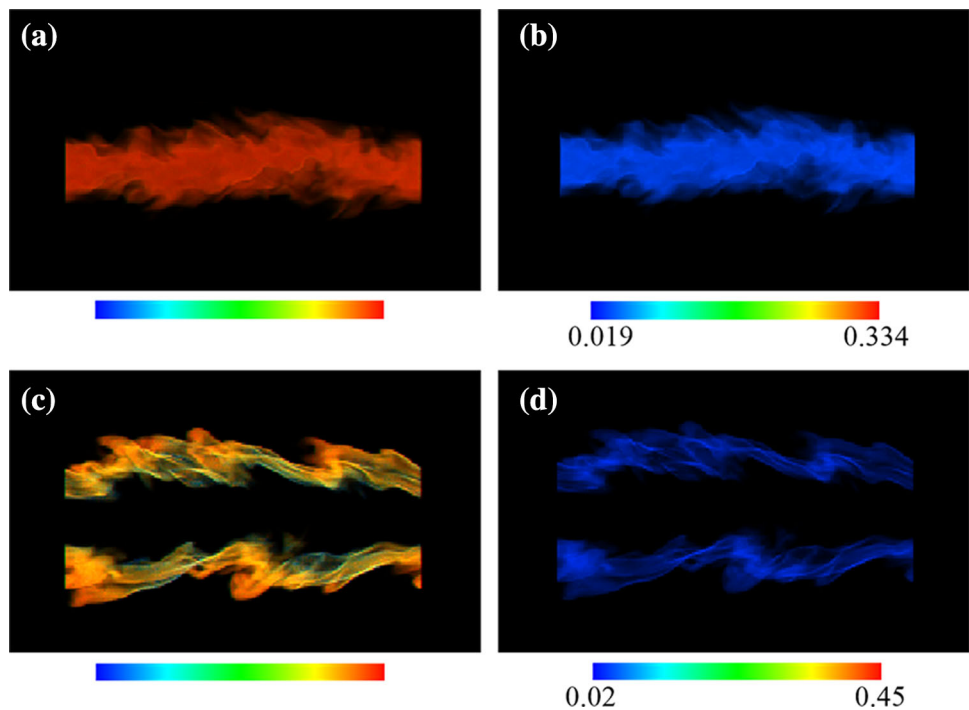

Fig. 14 Results of MIX and HR with isovalue 0.96 and 0.40 in time step 41. **a** Screen accumulating view of isovalue 0.96. **b** Standard deviation view of isovalue 0.96. **c** Screen accumulating view of isovalue 0.40. **d** Standard deviation view of isovalue 0.40


As it is observed in Fig. 13b, GMM better approximates the distribution than Gaussian; therefore, the crossing probability field computed by our method is more accurate than PMC method. If the distribution of ensemble members cannot be fitted accurately, the computed crossing probabilities will have large errors, especially for the serried and complex isosurfaces.

In order to illustrate the different association patterns across space, we explore the cases of isovalues 0.96 and 0.40 in time step 41. Figure 14a presents the screen accumulating view of isovalue 0.96 and relatively high HR values uniformly locate on the regions of the uncertainty isosurface of

MIX, which means MIX value 0.96 is usually corresponding to high HR values and the reactions are weak in these regions. However, in Fig. 14c, relatively diverse colours can be seen over the uncertainty isosurface with isovalue 0.40. This means that complex reactions are occurred in this region. At the same time, these two association patterns have relatively high credibility, because standard deviation values are relatively low as shown in Fig. 14b, d.

To illustrate the variations of associations between different variables in the time series, we explore the variable associations with isovalue 0.40 among time step 41, 61 and 81. As shown in Fig. 14c, green regions appear around the

Fig. 15 Results of MIX and HR with isovalue 0.40 in time step 61 and 81. **a** Screen accumulating view in time step 61. **b** Screen accumulating view in time step 81. **c** Mean view of the query result in time step 61. **d** Mean view of the query result in time step 81
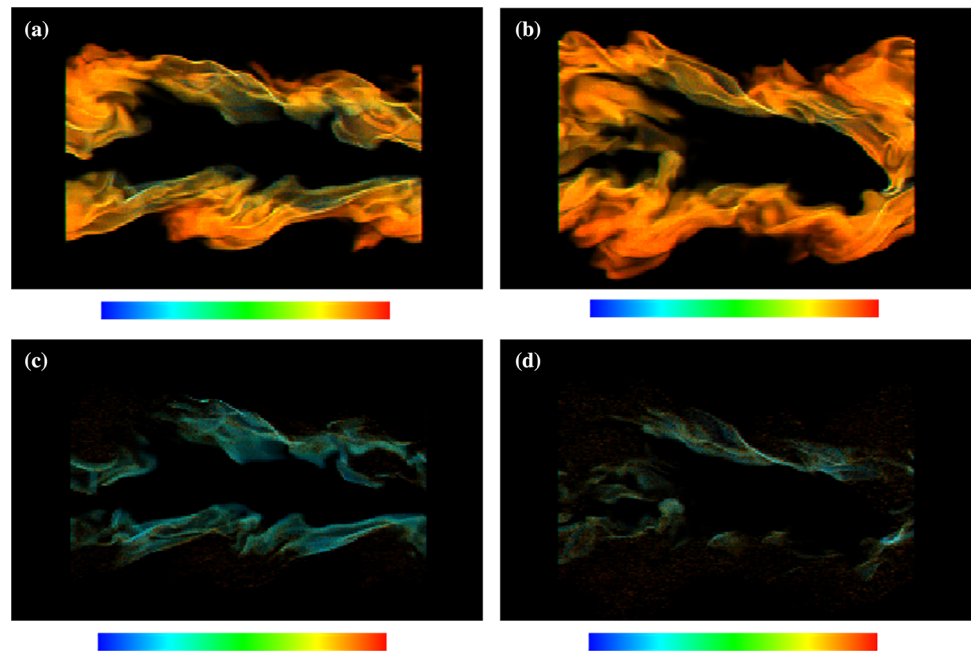
medial edges in time step 41, which is similar to the result of time step 61 in Fig. 15a. However, the association pattern changes in time step 81. As shown in Fig. 15b, the result of the accumulating view for time step 81 has rare green regions. This means that the values of heat release rate have generally increased in time step 81 when MIX is about 0.40, which means more violent reactions are happening. As shown in Fig. 15c, d, through the query of green regions, it is observed that the query result of time step 81 has less green regions and lower opacities. This shows the decrease in green regions more clearly.

8 Discussion

Our method provides an effective way to visualize the uncertain associations between different variables by the extraction of uncertainty isosurface and screen space accumulating method. Compared with the previous methods of visualizing variable associations, our method takes into account the uncertainty and reveals the credibility of variable associations. In this section, we present the selection of isovalues, the choice of parameters, the performance of our method and the evaluation of our method.

8.1 Selection of isovalues

If the users already have some knowledge of the data set, they can select the isovalues according to their knowledge or interest. Besides, users can explore the different isovalues in the field of ensemble mean and select multiple isovalues of interest in advance. We compute the uncertainty isosurfaces

of these isovalues together. For data reduction, it is worth mentioning that computations of the value range, mean and standard deviation for each cell only need to be carried out once for these isovalues. Moreover, GMM modelling of the common valid cells for different isovalues can also be performed only once. Therefore, the pre-processing overhead for multiple isovalues is not much more than the case for processing only one isovalue. Users can interactively vary the isovalues of interest among the group of isovalues in the exploration of uncertainty isosurfaces. The methods for selecting isovalues can be further studied.

8.2 Parameters setting

In this section, we firstly discussed the influence of different sampling number in the process of both extracting uncertainty isosurface and the syncretic rendering.

In order to accurately compute crossing probabilities, the sampling number should be large in the extraction of uncertainty isosurface. We test our method with different sampling number of 500, 1000, 2000, 4000 and 8000. Although little differences can be visually observed when sampling number is larger than 1000, the precisions are higher in the conditions of more samples. For the rendering of the associated variable, at least 300 samples are generated for each voxel. Smoother screen accumulating view can be obtained as the sampling number increases. This can present the contained uncertainty information more accurately.

Secondly, we discuss the selection of the discrimination rule used in the data reduction. According to our experiments for multiple data sets and isovalues, range $[\mu - 3\sigma, \mu + 3\sigma]$ is sufficient for most cases. Nearly, non-difference between

Table 1 Simplification rates and accuracy rates for data reduction of different data sets

Data sets	Isovalue	Range	Simplification rate (%)	Accuracy rate (%)
Synthetic data	0.01215	$\mu \pm 3\sigma$	84.43	100
		$\mu \pm 5\sigma$	75.29	100
Climate data	−24,000	$\mu \pm 3\sigma$	94.2	89.98
		$\mu \pm 5\sigma$	91.48	99.97
Combustion data	0.96	$\mu \pm 3\sigma$	86.38	85.96
		$\mu \pm 5\sigma$	82.65	93.71

the results computed by the original data and the result is obtained after the data reduction using range $[\mu - 3\sigma, \mu + 3\sigma]$. We also evaluate the effectiveness of our data reduction numerically. The simplification rate r_s are computed as:

$$r_s = \frac{C_{\text{invalid}}}{C_{\text{all}}} \quad (3)$$

where C_{invalid} denotes the number of invalid cells after data reduction and C_{all} is the number of all cells in the original data. For the accuracy rate r_a of data reduction, we define it as:

$$r_a = \frac{NC_{\text{reserved}}}{NC_{\text{all}}} \quad (4)$$

where NC_{reserved} denotes the number of the valid cells with nonzero probabilities that are computed using data reduction and NC_{all} is the number of all cells with nonzero probabilities that are computed based on the original data.

Table 1 lists the simplification rate and accuracy rate for different data sets with specific isovalues. The simplification rate is usually high and depends on the selection of isovalues. For synthetic data and climate data, relatively high accuracy rates can be reached. In combustion data, since the uncertainty isosurface with isovalue 0.96 of MIX is very complex, the accuracy rate of range $[\mu - 3\sigma, \mu + 3\sigma]$ is 85.96%. By range $[\mu - 5\sigma, \mu + 5\sigma]$, the accuracy rate can reach 93.71%. Actually, an arbitrary range $[\mu - p\sigma, \mu + p\sigma]$ can be used to perform the data reduction according to their requirements. Generally, the accuracy rate is higher and the simplification rate is lower when p is higher. Users can select the value of p according to their requirements.

8.3 Performance

We use 3 data sets to test the performance of our work, including the extraction of uncertainty isosurface and the visualization of variable associations. The tests are performed on a desktop computer, with a 3.5GHz Intel Core i7 CPU, 16GB memory and a NVIDIA GTX 780Ti GPU.

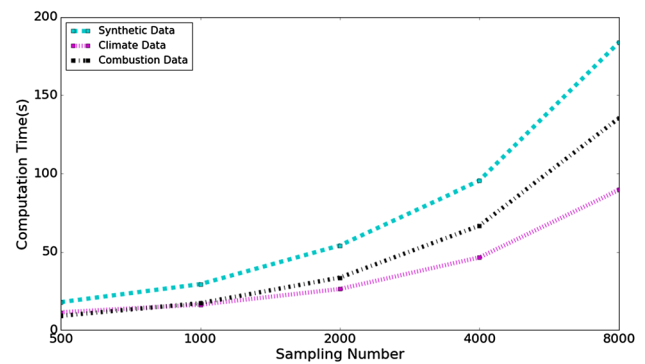In the implementation of extracting uncertainty isosurface, due to the huge data scale and the high computational

Fig. 16 Performance of extracting uncertainty isosurface with different sampling numbers. *Cyan line* denotes the performance of synthetic data set with isovalue 0.01215. *Pink line* denotes the performance of climate data set with isovalue −24,000. *Black line* denotes the performance of combustion data set with isovalue 0.96 in time step 41

complexity of per-cell GMM modelling (include the computation of BIC values), a pre-processing needs to be performed to reduce data and model GMM only for the valid cells. The computation time is greatly saved for a specific isovalue.

After the pre-processing, we can obtain the GMM parameters of weights, means and lower triangular covariance matrices. To accelerate the computation, CUDA is utilized to parallelize the sampling and the computation of crossing probabilities based on each independent cell. Every time, we maintain two layers of data along the z-axis (forming one layer of cells) storing in memory and transfer these cells' parameters of GMMs to GPU. The result of each sample is computed in a thread. We record the time for computing the crossing probability field using Monte Carlo integrating on GPU. Its performance is determined by the Gaussian component numbers of each layer along the z-axis and the threads number. Figure 16 displays the performance in the conditions of different sampling numbers and the same thread number. In Fig. 16, the numbers of valid cells for synthetic data, climate data and combustion data are, respectively, 248,078 (992,312 Gaussian components), 113,828 (444,356 Gaussian components) and 878,146 (2,634,438 Gaussian components). As the sampling number increases, the running time increases slowly unless the cases of 8000 samples for all three data sets.

Table 2 Performance of GMM modelling and sampling in the visualization of variable associations

Data sets	Isovalue	Variable	N_c	$T(s)$
Synthetic data	0.01215	$Var1$	248,078	9.35
	0.01215	$Var2$	248,078	9.49
Climate data	−24,000	MSLP	113,828	7.22
	5000	MSLP	112,271	7.52
	18,000	MSLP	143,939	8.68
Combustion data	0.40	$HR(TIME41)$	2,008,995	29.865
	0.40	$HR(TIME61)$	3,138,213	44.651
	0.40	$HR(TIME81)$	4,386,690	60.507

For the visualization of variable associations, the computation of GMM modelling for the ensemble members in a voxel is relatively fast. We perform GMM modelling and the sampling for the associated variable on-line and accelerate the computation with OpenMP. The sampling number and the size of uncertainty isosurface can affect its performance. The computation times of GMM modelling and sampling for different data sets are demonstrated in Table 2, in which N_c is the number of cells contained in the uncertainty isosurface and T is the computation time of GMM modelling and sampling. The number of samples is 500 for all three data sets. In animation, the per-frame volume rendering is accomplished in real time.

8.4 Evaluation

To evaluate the effectiveness of our visualization method for analysing the variable associations, we have performed a task-oriented user study with 11 graduate students, including 6 females and 5 males. The climate data set has been used in the user study.

The tasks are as follows:

Task1. Identifying the variable associations between the uniform pattern and the confuse pattern over the whole space;

Task2. Identifying the general association pattern and credibility for a given region;

Task3. Searching for the regions with specific general association patterns;

Task4. Searching for the regions with relatively high or low uncertainty;

Task5. Obtaining the detail information among ensemble members for a region with not credible associations.

After a brief explanation to our method and the tasks, the subjects completed the tasks and evaluated the efficiency and usability of our method. For *Task1* and *Task2*, all the subjects can provide the right answers in real time through the screen accumulating view. As for *Task3* and *Task4*, all the subjects

perform the queries for at least 3 times to get the required regions. With regard to *Task5*, some subjects pointed out that the information presented by animation cannot be received very well. They hope that other visualization techniques can be added, such as glyphs.

After summarizing all the questionnaires, it can be found that most of the subjects consider that our method is relatively efficient and is useful to analyse the variable associations for multivariate ensemble data. However, some subjects think that the rendering of association patterns can be further improved for a better visual perception.

9 Conclusion and future work

In this paper, we focus on the analysis of multivariate ensemble data and have presented an effective approach to visualize the uncertain associations between different variables in ensemble data. To accurately and compactly represent the uncertainty, GMM is exploited to model the distribution of ensemble members. Based on the GMM representation, we first extract the uncertainty isosurface of a reference variable to get a salient uncertainty feature. The probabilistic crossing field is obtained to quantify the uncertainty of the feature through Monte Carlo sampling. To effectively reveal the uncertain associations between different variables, the syncretic rendering technique is applied to combine the associated variable with the probabilistic crossing field of the reference variable. Through utilizing the screen space Monte Carlo integrating strategy, we can obtain the general pattern of variable associations in the multivariate ensemble data. Besides, we provide the switchable view and animation to present the credibility of variable associations and further convey the uncertainty contained in the variable associations. In the switchable view, the standard deviations of per-voxel samples for the associated variable is computed to show the credibility of the associations, by combining the spatial uncertainty of the reference variable and the value uncertainty of the associated variable. Query operation is supported to search for the regions with specific association patterns or

credibility. Animation can support the further observation of those regions with high uncertainty and the recognition of unlikely or abnormal cases.

One limitation of our methods may be that only the association between two variables can be visualized at the same time. Actually, our method can be extended by mapping the correlation values of two other variables onto the uncertainty isosurface, which needs a new correlation metric suitable for ensemble data. We leave it to the future work.

Since scientists may not have an entire knowledge of the data sets, another limitation of our work is that the effective selection criterions of significant variables and isovalues are absent. In the future, we plan to give users a better recommendation for the selection of significative variables and isovalues.

About the visualization of variable associations, more deep analyses can be further performed. Due to the analyses of unlikely cases are important for discovering some interesting association patterns, new visualization tools can be further developed to detect and analyse the unlikely cases in the future. Besides, new rendering methods will also be studied to reveal the uncertainty in deep level.

References

1. Wong, P.C., Bergeron, R.D.: 30 years of multidimensional multivariate visualization. Sci. Vis. Overv. Methodol. Tech. 3–33 (1996)
2. Sauber, N., Theisel, H., Seidel, H.P.: Multifield-graphs: an approach to visualizing correlations in multifield scalar data. IEEE Trans. Vis. Comput. Gr. **12**(5), 917–924 (2006)
3. Potter, K., Wilson, A., Bremer, P.T., Williams, D., Doutriaux, C., Pascucci, V., et al.: Ensemble-vis: a framework for the statistical visualization of ensemble data. In: IEEE International Conference on Data Mining Workshops, IEEE, pp. 233–240 (2009)
4. Mihai, M., Westermann, R.: Visualizing the stability of critical points in uncertain scalar fields. Comput. Gr. **41**(6), 13–25 (2014)
5. Schlegel, S., Korn, N., Scheuermann, G.: On the interpolation of data with normally distributed uncertainty for visualization. IEEE Trans. Vis. Comput. Gr. **18**(12), 2305–2314 (2012)
6. Thompson, D., Levine, J.A., Bennett, J.C., Bremer, P.T.: Analysis of large-scale scalar data using hixels. In: Large Data Analysis and Visualization, IEEE, pp. 23–30 (2011)
7. Pöthkow, K., Hege, H.C.: Nonparametric models for uncertainty visualization. Comput. Gr. Forum **32**(3pt2), 131–140 (2013)
8. Athawale, T., Sakhaee, E., Entezari, A.: Isosurface visualization of data with nonparametric models for uncertainty. IEEE Trans. Vis. Comput. Gr. **22**(1), 777–786 (2016)

9. Liu, S., Levine, J.A., Bremer, P., Pascucci, V.: Gaussian mixture model based volume visualization. In: IEEE Symposium on Large-Scale Data Analysis and Visualization, pp. 73–77 (2012)
10. Pöthkow, K., Weber, B., Hege, H.C.: Probabilistic marching cubes. Comput. Gr. Forum **30**(3), 931–940 (2011)
11. Potter, K., Rosen, P., Johnson, C.R.: From quantification to visualization: a taxonomy of uncertainty visualization approaches. In: Dienstfrey, A.M., Boisvert, R.F. (eds.) Uncertainty Quantification in Scientific Computing, p. 226C49. Springer, Berlin (2012)
12. Pang, A., Wittenbrink, C., Lodha, S.: Approaches to uncertainty visualization. Vis. Comput. **13**(8), 370–390 (1997)
13. Bonneau, G.P., Hege, H. C., Johnson, C.R., et al.: Overview and state-of-the-art of uncertainty visualization. In: Hansen, C.D., Chen, M., Johnson, C.R., Kaufman, A.E., Hagen, H. (eds.) Scientific Visualization. Springer, Heidelberg, pp. 3–27 (2014)
14. Brodlie, K., Osorio, R.A., Lopes, A.: A review of uncertainty in data visualization. In: Dill, J., Earnshaw, R., Kasik, D., Vince, J., Wong, P.C. (eds.) Expanding the Frontiers of Visual Analytics and Visualization. Springer, Heidelberg, pp. 81–109 (2012)
15. Ehlschlaeger, C.R., Shortridge, A.M., Goodchild, M.F.: Visualizing spatial data uncertainty using animation. Comput. Geosci. **23**(4), 387–395 (1997)
16. Lundström, C., Ljung, P., Persson, A., Ynnerman, A.: Uncertainty visualization in medical volume rendering using probabilistic animation. IEEE Trans. Vis. Comput. Gr. **13**(6), 1648–1655 (2007)
17. Huber, D.E., Healey, C.G.: Visualizing data with motion. In: IEEE Visualization, IEEE, pp. 67 (2005)
18. Hengl, T.: Visualisation of uncertainty using the HSI colour model: computations with colours. In: Proceedings of the 7th International Conference on GeoComputation, pp. 8–17 (2003)
19. Dinesha, V., Adabala, N., Natarajan, V.: Uncertainty visualization using HDR volume rendering. Vis. Comput. **28**(3), 265–278 (2012)
20. Pöthkow, K., Hege, H.C.: Positional uncertainty of isocontours: condition analysis and probabilistic measures. IEEE Trans. Vis. Comput. Gr. **17**(10), 1393–1406 (2010)
21. Potter, K., Kniss, J., Riesenfeld, R., Johnson, C.R.: Visualizing summary statistics and uncertainty. Comput. Gr. Forum **29**(3), 823C832 (2010)
22. Sanyal, J., Zhang, S., Bhattacharya, G., Amburn, P., Moorhead, R.: A user study to compare four uncertainty visualization methods for 1D and 2D datasets. IEEE Trans. Vis. Comput. Gr. **15**(6), 1209–1218 (2009)
23. Hao, L., Healey, C.G., Bass, S.A.: Effective visualization of temporal ensembles. IEEE Trans. Vis. Comput. Gr. **22**(1), 787–796 (2016)
24. Sanyal, J., Zhang, S., Dyer, J., Mercer, A., Amburn, P., Moorhead, R.J.: Noodles: a tool for visualization of numerical weather model ensemble uncertainty. IEEE Trans. Vis. Comput. Gr. **16**(6), 1421–1430 (2010)
25. Pfaffelmoser, T., Reitinger, M., Westermann, R.: Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. Comput. Gr. Forum **30**(3), 951–960 (2011)
26. Athawale, T., Entezari, A.: Uncertainty quantification in linear interpolation for isosurface extraction. IEEE Trans. Vis. Comput. Gr. **19**(12), 2723–2732 (2013)
27. Gosink, L., Anderson, J., Bethel, W., Joy, K.: Variable interactions in query-driven visualization. IEEE Trans. Vis. Comput. Gr. **13**(6), 1400–1407 (2007)
28. Guo, H., Xiao, H., Yuan, X.: Scalable multivariate volume visualization and analysis based on dimension projection and parallel coordinates. IEEE Trans. Vis. Comput. Gr. **18**(9), 1397–1410 (2012)
29. Biswas, A., Dutta, S., Shen, H.W., Woodring, J.: An information-aware framework for exploring multivariate data sets. IEEE Trans. Vis. Comput. Gr. **19**(12), 2683–2692 (2013)

30. Liu, X., Shen, H.W.: Association analysis for visual exploration of multivariate scientific data sets. IEEE Trans. Vis. Comput. Gr. **22**(1), 955–964 (2016)
31. Zhang, H., Hou, Y., Qu, D., Liu, Q.: Correlation visualization of time-varying patterns for multi-variable data. IEEE Access **4**, 4669–4677 (2016)
32. Pfaffelmoser, T., Westermann, R.: Visualization of global correlation structures in uncertain 2D scalar fields. Comput. Gr. Forum **31**(3pt2), 1025–1034 (2012)
33. Jänicke, H., Böttinger, M., Mikolajewicz, U., Scheuermann, G.: Visual exploration of climate variability changes using wavelet analysis. IEEE Trans. Vis. Comput. Gr. **15**(6), 1375–1382 (2009)

Qi Shang received his B.Sc. degree in 2015 from Northeast Normal University. He is a M.Sc. student in the School of Computer Science and Information Technology, Northeast Normal University. His current research interests include computer graphics and scientific visualization.

Huijie Zhang received her Ph.D. degree in 2009 year from Jilin University. Now she is an associate professor in the School of Computer Science and Information Technology, Northeast Normal University. Her main research interests include scientific visualization, information visualization, visual analysis, computer graphics, 3D model simplification, multi-resolution modeling for terrain and 3DGIS and optimization algorithm. Member of China Computer Federation and IEEE.
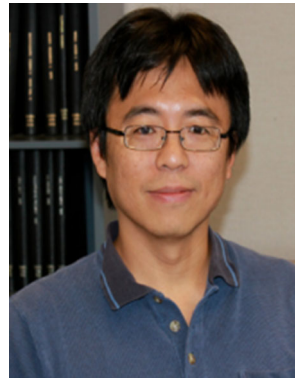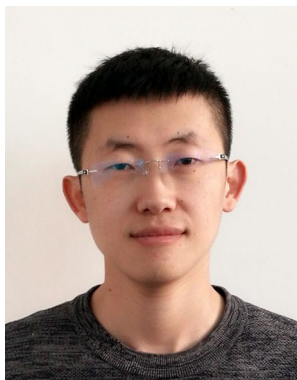
Yafang Hou received her B.Sc. degree in 2014 year from Northeast Normal University. Now she is a M.Sc. student in the School of Computer Science and Information Technology, Northeast Normal University. Her current research interests include computer graphics and scientific visualization. Student Member of IEEE.

Dezhan Qu received his B.Sc. degree in 2014 from Northeast Normal University. Now he is a M.Sc. student in the School of Computer Science and Information Technology, Northeast Normal University. His current research interests include computer graphics and scientific visualization.

Han-Wei Shen is a full Professor of Computer Science and Engineering at The Ohio State University. He is the head of the Graphics and Visualization Study (GRAVITY) research group. He received his Ph.D. degree from Scientific Computing and Imaging Group at University of Utah in 1998. His research interests include data analysis and visualization, computer graphics, and high performance computing.

Quanle Liu received his B.Sc. degree in 2015 from Changchun Institute of Technology. Now he is a M.Sc. student in the School of Computer Science and Information Technology, Northeast Normal University. His current research interests include computer graphics and scientific visualization.