https://doi.org/10.2352/ISSN.2470-1173.2017.1.VDA-384

An Interactive Tool for Analyzing the Correlation, Uncertainety, and Clustering (ACUC) over members within Ensembles in Climate Dataset

***Najmeh Abedzadeh*[1]**
Mississippi State University; Mississippi State, MS

Abstract

Weather scientists are looking to better understand the atmospheric conditions. We propose a new tool to detect the most significant association between variables in the multidimensional multivariate time-varying climate datasets. In this case, we represent the correlation between variables, the uncertainty between different members within ensembles, and several clustering methods. The climate dataset is collected in different time steps and locations. One of the most important research questions for weather scientists is the relationship between various variables in different time steps, or dissimilar spatial locations. In this paper, we present a set of techniques to evaluate the correlation and association between different variables within a time step and spatial location. In another way, we perform static analysis on a single point in space-time, then extending that analysis either in the temporal or spatial dimension(s), followed by an aggregation of the individual results to get an "overall" correlation. We created a tool that not only can be used to visualize the correlation and uncertainty between two time series of all ensembles, but also spatial locations. Mini-batch- K-Means clustering is applied to these datasets to identify the most substantial patterns within them. We study the Pearson correlation and integrate glyphs and color mapping into our design to demonstrate the trend of changing the correlation values of a single, pair, or triple of variables. Statistical calculations are applied to derive an accurate interpretation of the time-varying correlations between members within all of the ensembles as well as the uncertainty of the correlation values. The uncertainty visualizations provide insight toward the effects of parameter perturbation, sensitivity to initial conditions, and inconsistencies in model outputs. To evaluate the tool, we apply this technique to a climatology dataset.

Introduction

Representing the correlation among variables in a dataset is one of the most significant part of many types of research. Correlation analysis applies in different subjects, including climatology, business, education, biology, and many other fields. For example, in climatology, correlation analysis is beneficial in detecting the hidden relationship between variables in the dataset and discovering the new patterns to forecast accurately. However, performing such investigations rapidly in the climate datasets, which is a multidimensional, multivariate dataset is difficult.

Visualizing the uncertainty of the data is beneficial in understanding the trends of climate change or the reliability of weather predictions. Uncertainty can represent the accuracy of the data and errors are mostly coming from the simulation of faulty measurements of the initial conditions, or sensitivity of input parameters. For this reason, researchers combine multiple simulations using different parameters, and conditions and create various ensembles, which in turn provide uncertainty. Visualizing the data uncertainty, scientists have a better understanding of the data and consistency of weather predictions.

For this purpose, we propose to concentrate on the correlations and uncertainty between climate variables across different members within ensembles. In this study, we address three different challenges in weather datasets: correlation between different variables over ensembles, comparing the correlation values, and illustrating the uncertainty of correlation values over ensembles. The correlations between the variables of each ensemble are described through data clustering and a correlation metric in a spatial plot. We can show the clustering for every variable in the climate dataset, similar to pressure, temperature, height, humidity, wind longitudinal and latitudinal velocity. The uncertainties between the members within ensembles are also shown from the differences between the ensemble values for each of the variables.

RELATED WORKS

Correlation analysis in climate dataset have been studied by many different research groups. Several important topics related to our work, including multifaceted data, time-varying datasets, Visualization techniques, and statistical analysis are discussed in this section. Correlation analysis has been used in different research domains to detect the hidden patterns in the datasets. Many current works in the correlation analysis are including scatterplots, parallel coordinates, color mapping, and glyphs. Correlation analysis is a part of the multivariate relationship, by which it can discover the significant association between multiple variables in a dataset.

One of the biggest challenges for weather scientists is working with multifaceted data. The multifaceted data are categorized into spatiotemporal (spatial structure), multivariate (different attributes), multimodal of data sources, multirun (different ensembles), and multimodel of simulation (physically interacting phenomena such as the atmosphere) [14]. Many works are done on multivariate visualization, including dimension reduction, pattern mapping, and correlation analysis. The dimension reduction and pattern mapping are two of the multivariate visualization techniques. Barlowe et al. [2] used differentiation to enhance high dimensional visualization and represent the correlation between data. The dimension reduction and knowledge discovery are the results of their multivariate visualization. The sampling-based approaches are beneficial in simplifying these patterns[6].

The climate dataset are multidimensional multivariate time-varying datasets. Many researchers use Query-driven methods to handle time-varying multivariate data [10]. The developing of a textual pattern matching has shown to be beneficial in identifying the temporal patterns in time-varying datasets, but they lack detecting the events that cross many timesteps [9][8]. Hierarchical clustering is used in another study to detect the similarity in the correlation relation in time-varying multivariate climate datasets [11].

Scatterplots and parallel coordinates are two of the most important visualization techniques in representing the correlation between variables. Scatterplots are famous in representing the correlation between two variables. They are powerful in visualizing the multidimensional data. Due to the simplicity and readability of scatterplots, they are still famous in various areas and researchers have tried to augment them into a new design with a higher visual perception of the relationship between two data variables [5]. Parallel coordinates are one of the most famous visualization methods in representing the correlation between different variables in a dataset. Different studies have enhanced parallel coordinates to fit in their dataset. Blass et al. [4] developed parallel coordinates to show the usability of parallel coordinates in the large datasets, similar to the climate dataset, which is a multivariate, multidimensional time-varying dataset [12]. In another research study, the authors enhanced most of the related improvements into the new interactive design of visualizing parallel coordinates that represent the significant association between variables in a climate dataset [23].

The combination of statistical analysis and visualization techniques have shown to be effective in representing the features in the large time-varying datasets. [3]. Integrating statistical methods and feature extraction into Principal Component Analysis (PCA) and correlation analysis has shown to be effective in detecting the internal relationship between the data attributes in the time-varying datasets [20]. The point-wise correlation coefficients and canonical correlation analysis are applied to the time-varying multivariate climate data set to reveal the patterns and the connections between data attributes [24]. The query-driven visualization and correlation exploration are implemented to investigate the relationship between variables in the time-varying multivariate volumetric and particle data sets [25].

The purpose of this research is to investigate the new patterns in the climate dataset, which is a multidimensional, multivariate, time-varying dataset in a specific time and location between members within different ensembles and visualize the correlation between data attributes and the uncertainty among various members within ensembles in a dataset. Abedzadeh [1] implemented a new method for representing the correlation between varibales in multidimensional multivariate time varying datasets. Nguyen, et al. [18] applied the object and field-based visualization into climate dataset to represent the variation of a data attribute and their correlation at a specific time and location. Qu, et al. [21] embedded parallel coordinates, circular pixel bar charts, and weighted complete graphs into a polar system to detect patterns in the air pollution dataset in Hong Kong. Sanyal, et al. [22] visualized the uncertainty between weather ensembles by developing the current techniques and tools. Collins [7] represented an interactive method for visualizing the relationship and connection between visualizations by using interaction techniques such as filtering. Zhang, et al. [26] investigated the uncertainty information on the feature-level uncertainty and data-level uncertainty in 2D and 3D datasets.

Potter, et al. [13] proposed an ensemble visualization framework to visualize weather ensembles, their uncertainty and the statistical computations behind the data. Phadke, et al. [17] visualized ensemble members and compared them to simplify the interpretation. Alabi, et al. [19] sliced the ensemble members into equal-width strips and then combined them to highlight the surface shape differences between the strips members. Hao, et al. [15] analyzed and visualized the ensemble members and extracted the important shape transition between ensemble members over time. Recently, Hao, et al. [16] clustered, compared, and visualized different ensemble members to represent the inter-member shape and data similarity in the datasets. In this research, we used glyphs to detect the new patterns in the climate dataset, which is a multidimensional, multivariate, time-varying dataset. To represent the results, we used correlation analysis, clustering algorithms, and uncertainty visualization. These patterns are unable to show in any other way.

ANALYZING THE CORRELATION, UNCERTAINTY, AND CLUSTERING (ACUC).

ACUC (Analyzing the Correlation, Uncertainty, and Clustering) is a tool that helps weather forecasters discover new patterns in their datasets. Analyzing the correlation, uncertainty, and clustering is beneficial in investigating the hidden association between variables in a dataset. Climate dataset is a multidimensional multivariate time-varying dataset and weather scientists are looking to predict the weather more accurate by evaluating these datasets. ACUC tool is divided into three section, including correlation, uncertainty, and clustering that we will discuss in this section.

correlation

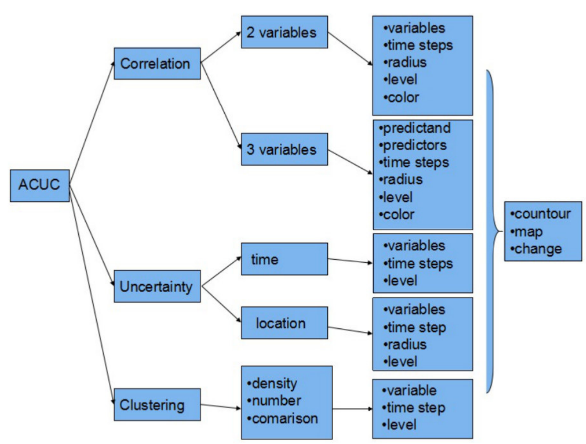

Figure 1. *The stucture of ACUC is represented*

Correlation study is beneficial in exploring the unknown relationship between different variables in a dataset. ACUC evaluates the correlation in two different ways, correlation between two variables and correlation between three variables. For representing the correlation between two variables, they have different options. First they choose the variables that they want to be correlated. Then, they choose the time step, the radius of the circle around of any spatial location in the dataset, the level of pressure(The data are collected from different value of the pressure, which gives us various), and finally they can choose the color of the correlation lines. Having the ability to choose different colors helps them to compare the correlation lines by their color. For representing the correlation between three variables, they have three variables, two predictors and one predictand. Other options are the same as correlation between two variables. In this case, instead of illustrating the correlation between two variables, they predict one variable based on the correlation between three of them. Equation 1(Pearson correlation) is used for calculating the correlation between two variables and Equation 1 is used for calculating the correlation between three variables.

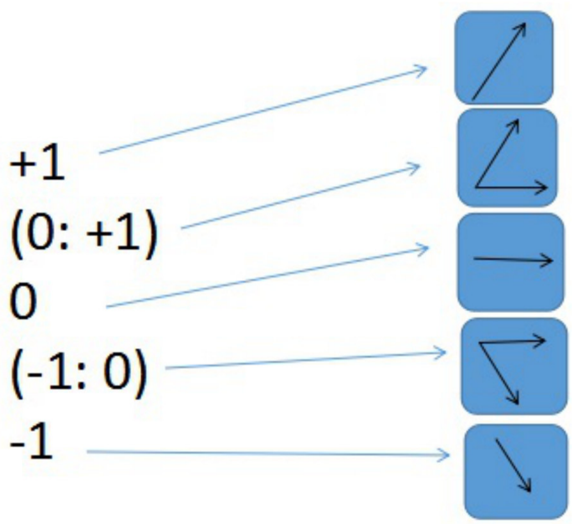

Figure 2. *The meaning of the correlation lines. The value* $"-1"$ *has a line with a slope of* $"-75"$ *degree. The value between* $"-1"$ *and* 0 *has a line with a slope between* $"-75"$ *and* 0 *degrees. The value* 0 *means there is no correlation between them, and it is a horizontal line. The value between* 0 *and* $"+1"$ *has a line with a slope between* 0 *and* $"+75"$ *degrees. Finally, The value* $"+1"$ *has a line with a slope of* $"+75"$ *degree*

Let x and $y \in R^N$ are the related points in an n- dimensional space, the correlation between x and y is defined as Equation 1. However, the result, r is a value between -1 and $+1$. Thus, we projected the correlation values to vector lines. Consequently, the correlation represented as a vector line with slope between -75 to $+75$ degrees. We chose this degree because it is wide enough to visually compare the correlation between variables. As depicted in Fig. 2, the value -1 has a line with a slope of -75 degree. The value between -1 and 0 has a line with a slope between -75 and 0 degrees. The value 0 means there is no correlation between them, and it is a horizontal line. The value between 0 and $+1$ has a line with a slope between 0 and $+75$ degrees. Finally, The value $+1$ has a line with a slope of $+75$ degre.

$$r = \frac{\sum_{i=0}^{n}(y_i - y_0)(x_i - x_0)}{\sqrt{\sum_{i=0}^{n}(y_i - y_0)^2}\sqrt{\sum_{i=0}^{n}(x_i - x_0)^2}} \tag{1}$$

In this experiment, we propose to show the correlation between variables for every single point separately. Therefore, we calculated the correlation between two variables with n observations, and the result is a number with a value between -1 and +1. At this point, we have one value for all of the data points. Our goal is having different values for every single point to show the variation between correlation values, otherwise, every point will have the same slope for its vector line.

$$r = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2r_{xz}r_{yz}r_{xy}}{1 - r_{xy}^2}} \tag{2}$$

In Equation. 2, we are calculating the correlation between three variables, rxz, ryz, and rxy. rxz is the correlation between x and z, ryz is the correlation between y and z, and rxy is the correlation between x and y, which are computed using Equation. 1. Using Equation. 2, we can detect some new patterns in the dataset, which are not detectable using any other way.

Uncertainty

The second application of ACUC is visualizing the uncertainty. To calculate the uncertainty between members within different ensembles, we represented this uncertainty using two different methods, including time uncertainty, and location uncertainty. To represent the uncertainty of correlation between two variables with different ensembles and show their confidence level, we calculated their correlation over time and location. The user can choose the variables, the time steps, and the level of pressure to calculate them over time. In this case, we used historical data. For example, to calculate the correlation between pressure and temperature with different ensembles between time step 42 and 48, we calculated the correlation between pressure and temperature between timestep 42 and 43, 42 and 44, 42 and 45, 42 and 46, 42 and 47, and 42 and 48. We kept the min and max values for every single ensemble and continued calculating them for all 32 ensembles. Then, we represented the range of correlation values for all ensembles by arcs. Finally, to show the distribution of correlation values between different ensembles, we highlighted the correlation between pressure and temperature between timestep 42 and 48 with arrows for every ensemble separately. This structure is beneficial in representing the values that correlation values are more concentrated in uncertainty (Figure 6.a).

For representing the uncertainty over location, user can choose the variables, the time step, the radius of the circle around of every single spatial location, and the level of the pressure. Having different values for different ensembles, we could show the range of correlation values with arcs around of them. To show the distribution of these correlation values, we visualized the correlation for every single ensemble separately, which is shown by lines inside of the arcs, which has the same benefits as time uncertainty. The difference is that location uncertainty is the uncertainty in correlation values between different variables within a circle around of every point in a specific time, while time uncertainty is the uncertainty in correlation values between different variables between two different time steps in one spatial location (Figure 6.b).

Clustering

The last benefit of ACUC is clustering. They have the ability of choosing the variables that they want to cluster, their time step, and the level of pressure. Clustering the data points provides a framework for determining the points that belong to the same class and investigates them more accurately. There are so many variables in the multivariate datasets that clustering them is beneficial in detecting a global trend between all of the variables in the dataset. We used Mini-Batch K-Means clustering algorithm and clustered all the variables in the dataset. This clustering algorithm computes the cluster centers by breaking them into pieces of mini-batches. We allocated a class to a subdivision of points based on their distance to the nearest mean. It works similar to K-Means, but it chooses the initial cluster centers in a smart way, which speeds it up. Therefore, it predicts the closest cluster that each sample in the dataset fits.

There are several different clustering algorithms. In this research, we have a large dataset, and for the sake of reducing the number of distance computations per iteration, we applied Mini-Batch K-Means clustering. The new way of clustering the variables is represented in three different ways, including ***number of clusters***, ***clustering density***, and ***clustering all variables***. The ***number of clusters*** algorithm, clusters all ensembles for every variable separately and then counts the number of clusters in every single spatial location. For example, clustering 5 different variables represented that all of the variables belong to the same cluster, if the number of clusters is one for a point. Also, if the number of clusters is 5 for a point, it shows that all of the variables belong to different clusters, which results to 5 different clusters in that point (Fig. 4(a)).

To visualize the cluster density algorithm, after clustering every single variable with different ensembles, we added the cluster numbers together and then divided the result over five (number of the variables) to get the density value for every point. If the density value is one for a point, it shows that all of the variables belong to the same cluster. Also, if the density value is 5 for a point, it shows that all of the variables belong to the different clusters, which means that we have 5 different clusters in that point (Fig. 4(b)).

For ***clustering all variables***, we put every single variable in one row and all ensembles as the columns and clustered them. The clustering results were shown with five different colors, and it separated the critical regions similar to storm regions from other areas. In conclusion, we can say that every clustering algorithm could show some new results and investigate the new patterns in the dataset (Fig. 5).

Finally, adding ***contour lines***, ***map***, and ***variables change***, they can get a better conclusion from their dataset. The contour lines give the users the ability of showing the boundaries of a variable in a specific level and time step. This is mostly beneficial in visualizing the specific phenomena similar to storm regions. The ***variables change*** is an option for representing the locations that one variable is increasing. This option helps them to decide on other variables whether they are increasing or decreasing in an specific time period and level.

RESULTS

We applied ACUC on the storm dataset to detect the patterns inside of the dataset. In this section, we show the related results for correlation, clustering, and uncertainty. We visualize the related images, including correlation between two variables, the correlation between three variables, a ***number of clusters***, ***Cluster Density***, ***clustering all variables***, uncertainty over time, and uncertainty over the location.

Superstorm dataset

Superstorm dataset is called the storm of the century happened between March 12 and 14 1993 from Central America to Canada. This storm had over 250 mortalities and happened because of the low pressure and temperature alongside with winds and snowfall. They used National Weather service science operations in Weather Research and Forecasting (WRF).

Correlation between ensembles

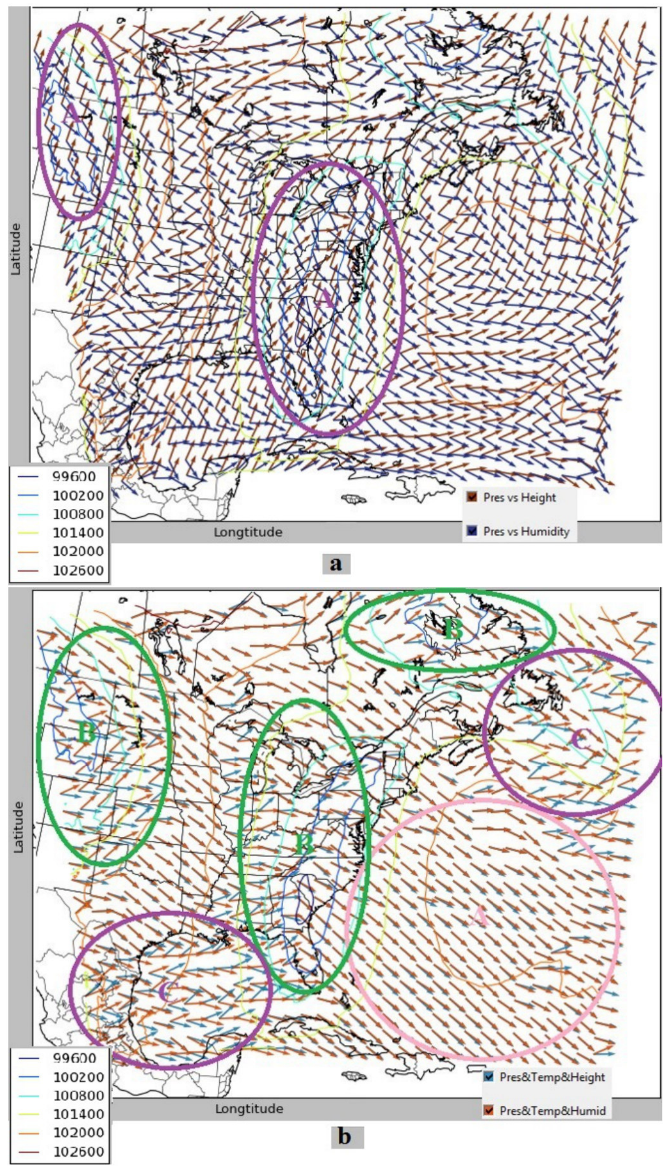

Figure 3. *Comparing the correlation between two variables versus three varibales*

Fig. 3(a) and Fig. 3(b) visualize the correlation between variables. The important regions in these images are highlighted by author to represent the difference between various patterns in the pictures. In Fig. 3(a), we represent the difference between the correlation values for pressure versus height with orange color and pressure versus humidity with blue color. The larger difference between these two correlation values is coming from the storm areas in regions A. In these areas, pressure versus height is highly positive while pressure versus humidity is highly negative. This pattern has happened in other areas too, but the angle between these arrows has the highest value in these regions.

Fig. 3(b) compares the difference between the correlation values among pressure, temperature, and height versus pressure, temperature, and humidity. The contour plot is the representation of pressure values, which can show the storm areas by the blue contour. In this image, we are considering pressure as a dependent variable, which its value is defined by the values of temperature, height, and humidity. The blue arrows represent the correlation between pressure versus temperature and height while the orange arrows show the correlation between pressure versus temperature and humidity. Section A shows the same pattern for these two variables while section B and C represent a lot of tribulations between arrows. The difference between region B and C is that region B is where the storm is happening, while there is no storm in region C. It can be concluded that the difference between the angle of the correlation lines demonstrate the storm areas (region B). In another way, it can be concluded that the difference between height and humidity is affecting the storm event.

The difference between Fig. 3(a) and Fig. 3(b) represents the role of temperature in correlation values. In both figures, pressure is the independant variable, which its correlation with other variables is visualized. The three variables including: pressure, height, and humidity are incorporated in both images, but the difference between them comes from the temperature. Comparing Fig. 3(a) and Fig. 3(b), we can see the difference between the correlation values in different regions.

Clustering ensembles

number of clusters

One of the techniques for visualizing the clustering of several different variables is ***number of clusters***. Fig. 4(a) represents the clustering algorithm, which named ***number of clusters***. In this case, every color is the representation of the number of the clusters in one particular longitude and latitude location. To get to this point, we clustered every single variable across the ensemble members separately and then counted the number of clusters in every spatial location. We can have from one to five clusters. One cluster means that all variables are showing the same behavior in that point and they are related to each other. Also, five clusters means that all variables are from different clusters, and they are not related to each other in that spatial location. As it is represented in Fig. 4(a), most of the regions have two or three clusters. One interesting region is the dark violet oval in almost the center of the Fig. 4(a) (cluster 1), which is near to storm area.It can be concluded that all of the variables are correlated to each other around of the storm area.

Cluster Density

One of the methods for visualizing the clusters is ***Cluster Density***. Fig. 4(b) represents the clustering algorithm, which named ***Cluster Density***. The contour plot shows the pressure level and storm areas. To represent the cluster density, we clustered every single variable separately and then added the cluster number for every single variable up together. Then, we divided them into the number of the variables and calculated their density. For example, cluster 1 shows that all of them are from cluster 1. All dark violet regions are standing on the right part of the storm areas, which shows that they are all from the same cluster. Storm regions are the dark blue contour areas.

clustering all variables

Fig. 5 visualizes the ***clustering all variables*** algorithms. In this algorithm, we created a matrix with 5 rows of 5 variables, including pressure, temperature, height, humidity, and wind magnitude. The columns are the 32 ensembles. Then, we clustered the whole matrix. It is interesting how the storm areas, which are the blue contour regions are put in the same group, which is dark pink. We added the correlation between pressure and temperature

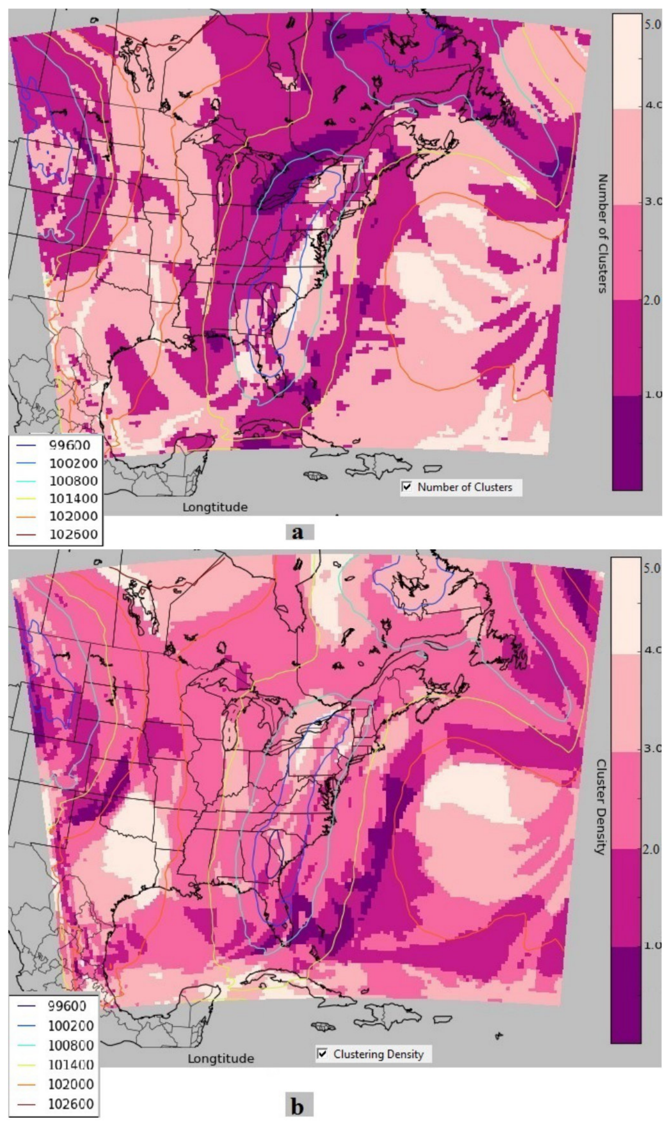

Figure 4. ***clustering all variables*** *in all ensembles using a)* ***number of clusters*** *method (the locations with the same color have the same number of clusters), b)* ***clustering density*** *method(the locations with the same color represent the same behaviour).*

to this image to show more patterns between data. These correlation values are represented in Fig. 5 with green arrows. The orange scatterplots represent the places that pressure is increasing. There are so many different patterns that we can investigate from this image. For example, the regions of cluster 3 with middle pink color are with negative line slopes with no orange scatterplot, which shows that pressure is decreasing and the temperature is increasing. A weather scientist evaluated this picture and recognized some patterns in this image. He claimed that cluster number 4 is the representation of the cold sector of the cyclone. He also mentioned that cluster number 2 in the middle is the area of the cyclone and cluster number 5 is probably the warm sector. These patterns are not easy to represent in any other way.

Uncertainty between ensembles

In this paper, uncertainty is visualized in two different ways, over time and location. This storm dataset, which has been used in this work has 72 time steps. The former experiments on this dataset has shown that storm is stronger between time steps 42 and 48. To show the results over time, we chose to visualize the correlation between time step 42 and 48 to have a higher performance in our results. Fig. 6(a) shows the difference in correlation values for different ensembles between time step 42 and 48. The most uncertainty happened in the center of the storm. The difference in correlation values comes from the correlation values between two time step in a particular region across different ensemble members. For example, the extent of the arcs in Fig. 6(a) comes from different correlation values. These correlation values are calculated between pressure and temperature from time step 42 to 48 for different ensembles. Fig. 6(b) represents the uncertainty between different ensembles for correlation values between pressure and temperature in a circled spatial location with radius 5 around of every point. In this case, the scope of the arcs comes from the difference between correlation values for various ensemble members. As depicted in Fig. 6(b), the most uncertainty is happening in the ocean.

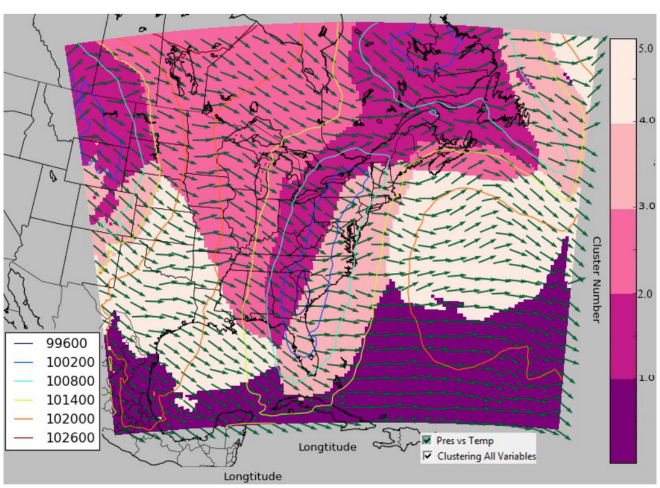

Figure 5. ***clustering all variables*** *in all ensembles using and integrating correlation lines, scatterplots, and contour into design.*

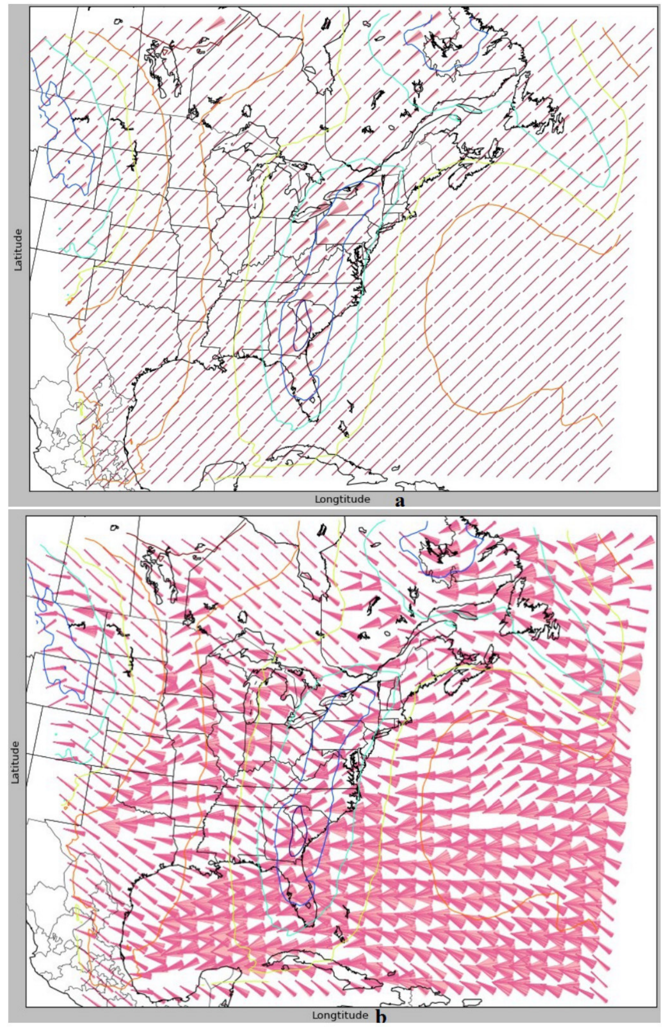

Figure 6. *Representing uncertainty among members within different ensembles a) between different time step 42 and 48, b) between different spatial locations with a circle around of every point with radius 5.*

CONCLUSION AND FUTURE WORKS

In this paper, we proposed a new tool for visualizing the correlation, uncertainty, and clustering for several different variables across different ensembles in the climate datasets. We have presented a novel framework that combines statistical analysis with visualization indicators to analyze large datasets. We used multiple layers of glyphs to indicate the difference between the correlation values. Moreover, Clustering techniques benefitted us in representing the similarity between variables. The uncertainty visualization helped us to find the statistical significant of the correlation values among different ensembles over time and location. Meteorologists can use this method to detect the patterns across members within different ensembles in a novel way.

References

[1] N. Abedzadeh. Correlation analysis in multidimensional multivariate time-varying datasets. *IEEE Scientific Visualization Conference (SciVis)*, pages 139–140, 2015.

[2] S. Barlowe, T. Zhang, Y. Liu, J. Yang, and D. Jacobs. Mul-

tivariate visual explanation for high dimensional datasets. *IEEE Symposium on Visual Analytics Science and Technology*, pages 147–154, 2008.

[3] J. C. Bennett, V. Krishnamoorthy, S. Liu, R. Grout, E. R. Hawkes, J. H. Chen, J. Shepherd, V. Pascucci, and P.-T. Bremer. Feature-based statistical analysis of combustion simulation data. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 17(12):1822–1831, 2011.

[4] J. Blaas, C. P. Botha, and F. H. Post. Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 14(6):1436–1443, 2008.

[5] Y.-H. Chan, C. D. Correa, and K.-L. Ma. The generalized sensitivity scatterplot. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 19(10):1768–1781, 2013.

[6] C. Chen, C. Wang, K. M. M. Tech, and A. T. Wittenberg. Static correlation visualization for large time-varying volume data. *IEEE Pacific Visualisation Symposium*, pages 27–34, 2011.

[7] C. e. a. Collins and S. Carpendale. Revealing relationships amongst visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1192–1199, 2007.

[8] M. Glatter, J. H. S. Ahern, J. Daniel, and A. Lu. Visualizing temporal patterns in large multivariate data using textual pattern matching. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 14(6):1467–1474, 2008.

[9] M. Glatter, J. Huang, S. Ahern, and A. Lu. Visualizing temporal patterns in large multivariate data using textual pattern matching. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 18(12):2392–2401, 2012.

[10] L. J. Gosink, J. C. Anderson, E. W. Bethel, and K. I. Joy. Variable interactions in query-driven visualization. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 13(6):1400–1407, 2007.

[11] Y. Gu and C.Wang. A study of hierarchical correlation clustering for scientific volume data. pages 437–446, 2008.

[12] H. Guo, H. Xiao, and X. Yuan. Scalable multivariate volume visualization and analysis based on dimension projection and parallel coordinates. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 18(9):1397–1410, 2012.

[13] P. B. D. W. C. D. V. P. K. Potter, A. Wilson and C. R. Johnson. Ensemble-vis: A framework for the statistical visualization of ensemble data. *In Proceedings of the 2009 IEEE International Conference on Data Mining Workshops*, pages 233–240, 2009.

[14] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 19(3):495–513, 2013.

[15] C. H. L. Hao and S. Bass. Effective visualization of temporal ensembles. *IEEE TVCG*, 22(1):787796, 2016.

[16] S. A. B. L. Hao, Ch. G. Healey and H. Y. Yu. Visual correlation analysis of numerical and categorical data on the correlation map. *IST International Symposium on Electronic Imaging 2016 Visualization and Data Analysis*, pages 1–10, 2016.

[17] O. A. J. H. R. M. T. I. X. W. H. P. S. A. B. M. N. Phadke, L. Pinto and C. G. Healey. Exploring ensemble visualization. *IEEE Transactions on Visualization and Computer Graphics*, page 82940B82940B, 2012.

[18] H. T. Nguyen, T. V. Tran, P. V. Tran, and H. Dang. Multivariate cube for visualization of weather data. *in Control, Automation and Information Sciences (ICCAIS), 2013 International Conference on*, pages 311–316, 2013.

[19] J. M. H. M. P. L. P. H. P. S. B. M. K. S. Z. C. G. H. O. S. Alabi, X. Wu and R. M. T. II. Comparative visualization of ensembles using ensemble surface slicing. *Visualization and Data Analysis*, page 112, 2012.

[20] S. Oeltze, H. Doleisch, H. Hauser, P. Muigg, and B. Preim. Interactive visual analysis of perfusion data. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 13(6):1392–1399, 2007.

[21] H. Qu, W.-Y. Chan, A. Xu, K.-L. Chung, K.-H. Lau, and P. Guo. Visual analysis of the air pollution problem in hong kong. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 13(6):1408–1415, 2007.

[22] D. J. M. A. A. P. M. R. Sanyal J1, Zhang S. Noodles: a tool for visualization of numerical weather model ensemble uncertainty. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 16(6):1421–1430, 2010.

[23] C. A. Steed, J. E. S. II, T. Jankun-Kelly, and P. J. Fitzpatrick. Guided analysis of hurricane trends using statistical processes integrated with interactive parallel coordinates. *IEEE Symposium on Visual Analytics Science and Technology*, pages 19–26, 2009.

[24] J. Sukharev, C. Wang, K.-L. Ma, and A. T. Wittenberg. Correlation study of time-varying multivariate climate data sets. *In Proc. IEEE Pacific Visualization*, pages 161–168, 2009.

[25] C. Wang, H. Yu, R. W. Grout, K. Ma, and J. H. Chen. Analyzing information transfer in time-varying multivariate data. *IEEE Pacific Visualisation Symposium*, pages 99–106, 2011.

[26] Z. Zhang, K. T. McDonnell, E. Zadok, and K. Mueller. Visual correlation analysis of numerical and categorical data on the correlation map. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 21(2):289–303, 2015.