EnsembleGraph: Visualizing Variations for Ensemble Simulations Exploration

Qingya Shu [1] Hanqi Guo [1,2] Limei Che [1] Weicong Lyu [1] Xiaoru Yuan [1,2*]

1) Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University, Beijing, P.R. China
2) Center for Computational Science and Engineering, Peking University, Beijing, P.R. China

Abstract

Ensemble simulation is increasingly important in various scientific domains. One of the most important tasks in ensemble simulation data visualization is to investigate consensus and disagreement between runs. In this work, we present an ensemble visualization tool which is capable of showing an overview of variations in the ensemble data. By providing visual summaries of ensemble variations, we enable users to quickly find regions of interests such as the disagreement regions across the members. We apply our prototype system on a weather simulation result to demonstrate the exploration of high variation regions in the dataset.

1 Introduction

Ensemble simulation, which is increasingly important in various scientific domains, it is used to evaluate model sensitivities by using different parameters or boundary conditions. In ensemble simulation results, the output values on the same location are usually different in different runs. Visualizing and analyzing ensemble data sets are regarded as challenging problems, as the data is usually multivariate, multi-valued and time-varying.

Existing ensemble visualization methods include map and statistical visualization [2], glyphs [3], visual comparisons [1], etc. Ensemble simulation data sets are usually time-variant. Animation and juxtaposing methods are used to visualize the multiple time steps [4].

One of the most important tasks for ensemble visualization is to investigate consensus and abnormalities patterns accross the runs. Scientists routinely make comparisons between the runs, in order to discover regions of interests in their workflow. In our observation, regions that appear to be highly similar or highly different among members are usually important in general.

In this work, we present *EnsembleGrpah*, a graph-based visualization tool for ensemble simulation comparison and exploration. The graph-based visual representation gives a summary of high variation regions in the data. Flexible user interaction is also designed to facilitate the exploration. In addition, we also provide spatial view for further checking data from spatial domain. Combining with linked view showing spatial distribution of variation regions, *EnsembleGraph* allows users to gain overview of ensemble data sets from the facet of variation in both space and time.

2 Our Method

The pipeline of *EnsembleGraph* is shown in Fig. 1. We first construct a graph, which depicts the evolution of variations in the dataset, and then create a graph layout for visualization. Together with spatial views, the prototype system enables user to explore and investigate high variation regions in the ensemble simulation results.

*e-mail:{qingya.shu,hanqi.guo,limei.che,weicong.lyu,xiaoru.yuan}@pku.edu.cn

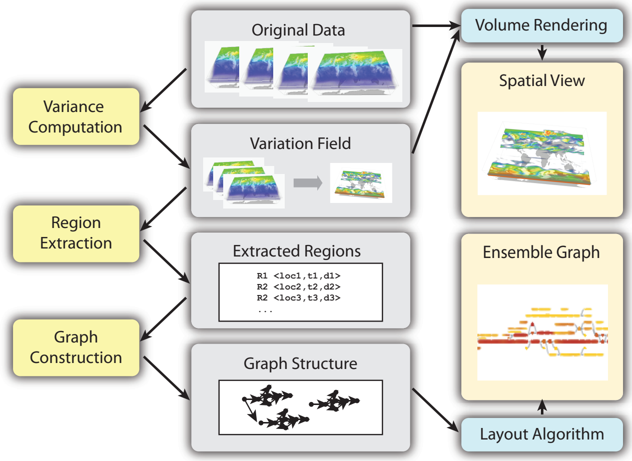

Figure 1: The pipeline of our method. The variation field is computed from the ensemble data, and then a graph is constructed based on the extraction of high variation regions. A graph layout is then computed and visualized.

2.1 Data Preprocessing and Graph Construction

In this work, *variation* means the degree of disagreement across ensemble members. If different output values at each sample point are close to each other, then it is considered as agreement, and vice versa. In our applications, we use standard deviation or variance as the metric of variation. For each location, we compute the variation of the sample values from all runs for each time step. Thus a time-varying scalar field is obtained, so called the *variation field*. Next, we extract all high variation regions according to the user-defined threshold. This process is repeated for each time step, and then a series of interesting variation regions across the temporal domain is obtained. As the variation regions may move in space, or change shape during time, the extracted regions over different time steps may be connected. These temporal coherent regions are extracted by scanning all neighborhood time steps. Then we obtain a graph that depicts the evolution of disagreement in the ensemble data. In this directed graph, nodes represent high variation regions, and edges are their connections. The directions of edges means the transformation as time evolves.

2.2 Graph View

Intuitive graph layout results are necessary for understanding the evolution of high variation regions. We pose the layout as a streaming graph, which aligns the time steps from left to right. There are many series of nodes connected together, due to the fact that variation region may sustain for a duration of time. In order to make the

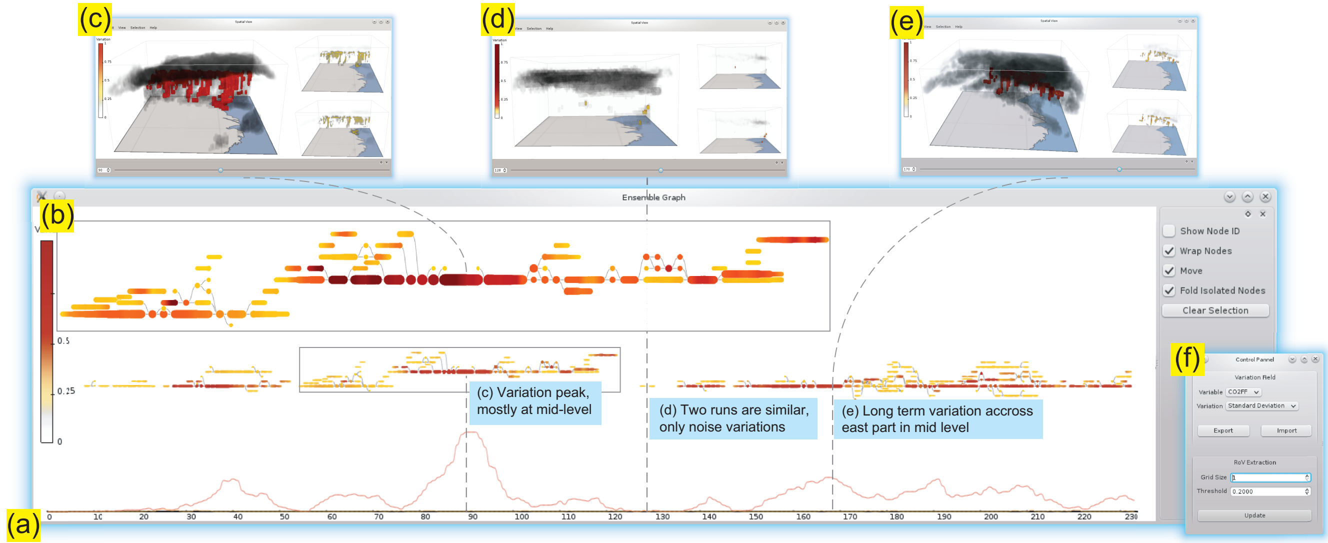

Figure 2: Interface. (a): graph view, (b): zoomed graph, (c), (d), (e): example nodes. (f) is control panel for parameters.

result less occluded for easier reading, we reduce the edge crossing by iteratively scanning time steps from left to right, and straighten the longest paths in the graph at each iteration.

2.3 Spatial View

A spatial view is provided to help users to investigate the spatiotemporal locations of high variation regions corresponding to the graph nodes. The graph view and the spatial view are linked. The regions corresponding to the selected nodes in the graph view are highlighted in the spatial view. Users can also select a high variation region by mouse clicking on the target, and corresponding nodes are then highlighted in graph view. Both the variation field and the raw ensemble data are visualized by GPU-accelerated volume rendering techniques. When there are regions highlighted, all other regions will be shown as context in gray colors.

3 Case Study

We apply *EnsembleGraph* on a ensemble WRF (Weather Research and Forecasting) simulation result. The scientists who conducted this simulation would like to find out how urbanization influences the weather. The dataset contains a set of output variables, and spatially covers East China. Two runs are conducted with different initial conditions. The first run is based on real observations, while second run is conducted by an ideally assumed circumstances that all urban areas are replaces by vegetarian landuse. The two results are all at a spatial resolution of $100 \times 100 \times 27$, and both include 237 time steps of hourly output. Comparisons on the two runs are necessary to help the scientists to validate their hypothesis.

The visualization results with *EnsembleGraph* are shown in Fig. 2. In this case, QCLOUD (Cloud Water Mixing Ratio) is chosen to investigate the variations. After setting the variable and threshold in control panel, the graph layout will be shown. We also visualize the amount of extracted regions using line chart at the bottom of the graph view, so that users know when do most high variations happen. From the graph view and the line chart, we can see that there are about 3 series of variation occurrences, peaks appear at around time step 40, 60, 90 and 170. We select nodes around those peak time steps, the variation of this node will be shown in the spatial view, most of which appeared in middle level. User can drag the time slider bar and check the animation of variation field. It seems that the mid levels are very usual places where output of cloud mixing ratio is more influenced by different initial conditions. From the spatial view we can see that, high levels have continuous variations occur, but their variation values are not very high comparing to regions at mid level. The graph view shows less nodes at around time step 130, indicating that two runs are more similar at that time. We can also spot a little noise variation starting at that time, which occurs in low levels just above the east ocean.

In this case, *EnsembleGraph* provides an overview of high variation regions in the time-varying ensemble datasets, thus enables users to quickly understand the data and find regions of interest.

4 Conclusions

In this work, we present a graph-based visualization method for ensemble simulation data. The graph provides a visual summary of high variation regions and their evolutions. Results show that our method is capable of finding high variation regions in the exploration of an ensemble weather simulation result.

Acknowledgements

This work is supported by NSFC No. 61170204. This work is also partially supported by NSFC Key Project No. 61232012 and the "Strategic Priority Research Program - Climate Change: Carbon Budget and Relevant Issues" of the Chinese Academy of Sciences Grant No. XDA05040205.

References

[1] J. Poco, A. Dasgupta, Y. Wei, W. Hargrove, C. Schwalm, R. Cook, E. Bertini, and C. Silvia. SimilarityExplorer: A visual inter-comparison tool for multifaceted climate data. *Comput. Graph. Forum*, 33(3):341–350, 2014.

[2] K. Potter, A. T. Wilson, P.-T. Bremer, D. N. Williams, C. M. Doutriaux, V. Pascucci, and C. R. Johnson. Ensemble-vis: A framework for the statistical visualization of ensemble data. In *ICDM 2009: Proceedings of IEEE International Conference on Data Mining Workshops*, pages 233–240, 2009.

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

[4] A. T. Wilson and K. C. Potter. Toward visual analysis of ensemble data sets. In *Proceedings of the 2009 Workshop on Ultrascale Visualization*, pages 48–53, 2009.