Visual Scalability of Spatial Ensemble uncertainty

Sujan Anreddy[*]
Mississippi State University

Song Zhang [†]
Mississippi State University.

Andrew Mercer [‡]
Mississippi State University.

Jamie Dyer [§]
Mississippi State University.

J. Edward Swan II[¶]
Mississippi State University.

Abstract

Weather Research and Forecasting (WRF) models simulate weather conditions by generating 2D numerical weather prediction ensemble members either through perturbing initial conditions or by changing different parameterization schemes, e.g., cumulus and microphysics schemes. These simulations are often used by weather analysts to analyze the nature of uncertainty attributed by these simulations to forecast weather conditions and to track storms. The number of simulations used for forecasting is growing with the advent of increase in computing power. Hence, there is a need for providing better visual insights of uncertainty with growing number of ensemble members. We propose a geo visual analytical framework that uses visual analytics approach to resolve visual scalability of ensemble members. Our approach naturally fits with the workflow of an analyst analyzing ensemble spatial uncertainty. Meteorologists evaluated our framework qualitatively and found it to be effective in acquiring insights of spatial uncertainty associated with multiple ensemble runs that are simulated using multiple parameterization schemes.

Index Terms: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 Problem Statement

Weather forecasters simulate weather conditions using numerical weather prediction models. These models are run multiple times in order to reduce error and uncertainty attributed by one single run. Each simulation run is termed an ensemble member that is generated either by perturbing initial conditions or by changing parameterization schemes. Some of the tasks that are performed with ensemble members and parameterization schemes are to find an ensemble run that represents consensus of all ensemble runs and also to find outliers from the ensemble. Therefore, in this paper "uncertainty" refers to the degree to which 2D scalar ensemble members agree or disagree with each other. A meteorologist is interested in understanding uncertainty attributed by parameterization schemes whereas an operational weather forecaster is interested in forecasting weather using these simulations. The climatologists also use these simulations to predict climate.

Ensemble-Vis[7] and Noodles[8] were built to provide insights on 2D numerical weather simulation models. They are efficient in presenting insights of overall uncertainty but are inefficient in handling various distributions of uncertainty as their spatial visualization techniques are not effective in identifying and tracking outliers. These frameworks initially depend on providing insights based on spatial characteristics of uncertainty and use those insights to understand multiple parameterization schemes. So, if the spatial visualization techniques used in these frameworks do not yield better representations of underlying distributions of uncertainty, it would affect a meteorologist′s understanding of parameterization schemes.

The use of visual metaphors such as graduated glyphs, ribbons are good at presenting the overall uncertainty or spread of ensemble members on single large display, but are not helpful in analyzing uncertainty attributed by a single ensemble member or group of similar ensemble members. These visualization techniques do not scale well with growing number of ensemble members as it becomes cumbersome to identify individual outliers and to track uncertainty across multiple spatial locations. The identification of individual ensemble runs or cluster of similar ensemble runs is important for a meteorologist as it helps them to narrow down the parameters that were used for generating these ensemble members. It also helps operational weather forecaster in terms of identifying an ensemble run that reflects the consensus of all the ensemble runs and predict weather based on a particular ensemble representative instead of using descriptive statistics like mean, median and standard deviation of all the ensemble runs. Hence, there is a need for presenting spatial uncertainty attributed by individual ensemble members effectively with precision and control.

2 Methodology

In order to overcome the disadvantages with frameworks discussed in the earlier section, we propose a geo visual analytical framework that presents initial insights on parameterization schemes and use's knowledge from these initial insights to dynamically build explore spatial characteristics of uncertainty. This approach naturally fits with the workflow of simulations, as the uncertainty in these simulations is driven by parameterization schemes.

We primarily divide our framework into two spaces, overview space and spatial exploratory space. These spaces guides the analyst to build their knowledge based on type of uncertainty characteristics acquired at each stage of the framework as shown in Figure 1. We will further describe our work with the help of an ensemble dataset used in evaluating our work. This ensemble dataset is simulated for 73 forecast hours using intial atmospheric conditions from 1999 storm event that occured on East coast. This dataset is simulated using 5 microphysics, 2 cumulus and 3 planetary boundary layer (PBL) schemes.

The purpose of this overview space is to present insights on the relationship between parameterization schemes based on uncertainty characteristics attributed by them. The euclidean distance metric is used to calculate the pairwise distance between ensemble members followed by clustering these ensemble members into selected number of cluster groups based on their calculated pairwise distances. The clustering results of n ensemble members are plotted using a dendrogram that identifies each ensemble member with parameterization schemes that were used in simulating them. As shown in Fig 1, this dendrogram presents insights on ensemble clustering groups in relation to parameterization schemes.

The spatial exploratory space guides analysts to perform a spa-

[*]e-mail: sra116@msstate.edu
[†]e-mail: szhang@msstate.edu
[‡]e-mail: mercer@gri.msstate.edu
[§]e-mail: jdyer@msstate.edu
[¶]e-mail: swan@cse.msstate.edu

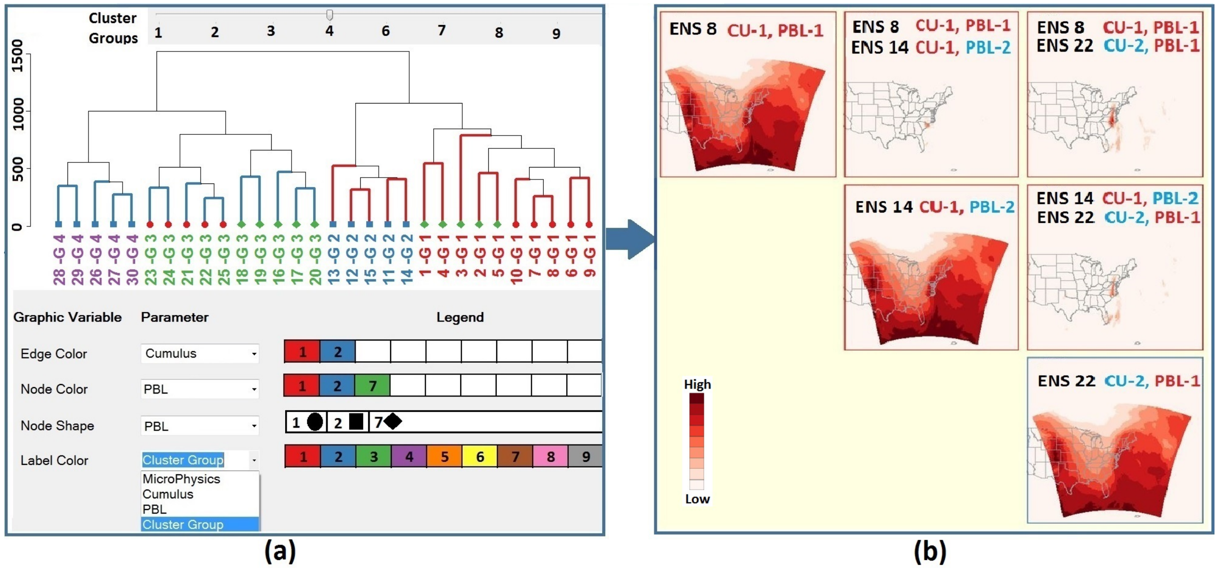

Figure 1: This figure represents 30 ensemble members of 850mb Temerature simulated for 73 hours as described previously. (a) This figure shows an interactive dendrogram where node attributes like node's shape and color are mapped to Cumulus while their corresponding edge colors are mapped to Planetary boundary layer(PBL) parameterization schemes. It provides insight on uncertainty with reference to parameterization schemes. The pattern in this dendrogram draws towards an insight that uncertainty is primarily driven by cumulus followed by PBL. Color Brewer scheme "Set1" is used for encoding categorical qualitative data like parameters belonging to parameterization schemes. (b) This shows the spatial uncertainty between ensemble members 8, 14 and 22 based on the insights gained in the overview space. The pairwise comparisons of ensemble members (8, 22) and (14, 22), exhibit greater uncertainty compared to the pairwise comparison of ensemble members (8, 14). It presents spatial characteristics of uncertainty using knowledge gained from the insight in the analysis phase (a). The color Brewer scheme "Reds" is used to encode sequential quantitative scalar spatial data. The color scale for ensemble members and pairwise comparisons is different even though we used the same color scheme to encode spatial data.

tial analysis of uncertainty attributed either by selected individual ensemble member or by a selected group of ensemble members based on the insight acquired from overview space. The purpose of this space is to provide insights on spatial characteristics like location and magnitude of the uncertainty. The spatial exploratory space consists of small multiples that represent individual ensemble members and their pairwise comparisons arranged in different layouts. These small multiples help the analyst build dynamic spatial patterns using knowledge gained from overview space. The matrix layout of small multiples helps analysts to acquire the spatial consensus of ensemble members, while the tabular layout of small multiples helps to make 1:n number of pairwise comparisons. This spatial exploratory space is tightly integrated with multiple levels of interactive features that help analysts explore spatial characteristics of ensemble members.

3 Evaluation and Conclusion

The novelty of this framework comes from using pairwise comparison approach to visualize ensemble spatial uncertainty and integrating this approach with unsupervised cluster analysis to find similar ensemble members, thereby reducing the total number of ensemble members required to visualize, thus resolving the visual scalability of spatial ensemble uncertainty. The entire framework is evaluated qualitatively by meteorologists in exploring parameter space and analyzing spatial uncertainty. Even though both evaluated the tool independently, they inferred the same insights from the dataset saying that "uncertainty is primarily driven by cumulus parameterization schemes followed by PBL scheme".

Acknowledgements

Funding for this work was partially provided by the Pacific Northwest National Laboratory, under U.S. Department of Energy Contract DE-AC05-76RL01830. This material is based upon work supported by the National Science Foundation under award NSF1117871.

References

[1] K. A. Cook and J. J. Thomas. Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2005.

[2] B. Fisher, T. M. Green, and R. Arias?Hernndez. Visual analytics as a translational cognitive science. *Topics in Cognitive Science*, 3(3):609–625, 2011.

[3] T. M. Green and R. Maciejewski. A role for reasoning in visual analytics. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pages 1495–1504, 2013.

[4] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323, 1999.

[5] A. M. MacEachren and J. H. Ganter. A pattern identification approach to cartographic visualization. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 27(2):64–81, 1990.

[6] O. Maimon and L. Rokach. *Data Mining and Knowledge Discovery Handbook*. Springer-Verlag New York, Inc., 2005.

[7] K. Potter, A. Wilson, P. T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. R. Johnson. Ensemble-vis: A framework for the statistical visualization of ensemble data. In *Data Mining Workshops, 2009. ICDMW '09. IEEE International Conference on*, pages 233–240, 2009.

[8] 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. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1421–1430, 2010.

[9] D. Wilks. Statistical methods in the atmospheric sciences, 2011.