Open access • Journal Article • DOI:10.1109/MCG.2014.52

Future Challenges for Ensemble Visualization — Source link

Harald Obermaier, Kenneth I. Joy

Institutions: University of California, Davis

Published on: 17 Jun 2014 - IEEE Computer Graphics and Applications (IEEE Computer Society)

Topics: Visual analytics, Information visualization, Visualization, Data visualization and Scientific visualization

Future Challenges for Ensemble Visualization

H. Obermaier and K. I. Joy
University of California, Davis

Introduction

How often have you executed an algorithm only to find that getting reasonable results means changing parameters and restarting? How much time did you spend on finding the "correct" parameters? Imagine going through the same ordeal with unbelievably complex simulation models used for the prediction of physical phenomena. Scientists have fought this battle for many years and have long been sick of the "sitting, waiting, and restarting" process. Luckily, increased availability of vast computation resources and new computation strategies offer a solution to this dilemma: Scientists can now run several alternative parameter settings or simulation models in parallel, creating an 'ensemble' of possible outcomes for a given event of interest (see Figure 1). In our conversations with simulation scientists and visualization researchers, the visual analysis of ensemble data has repeatedly come up as one of the most important new areas of visualization and we expect it to have a wide impact on our field in the next few years. The challenge is to develop expressive visualizations of properties of this set of solutions, the ensemble, to support scientists in this challenging parameter-space exploration task.

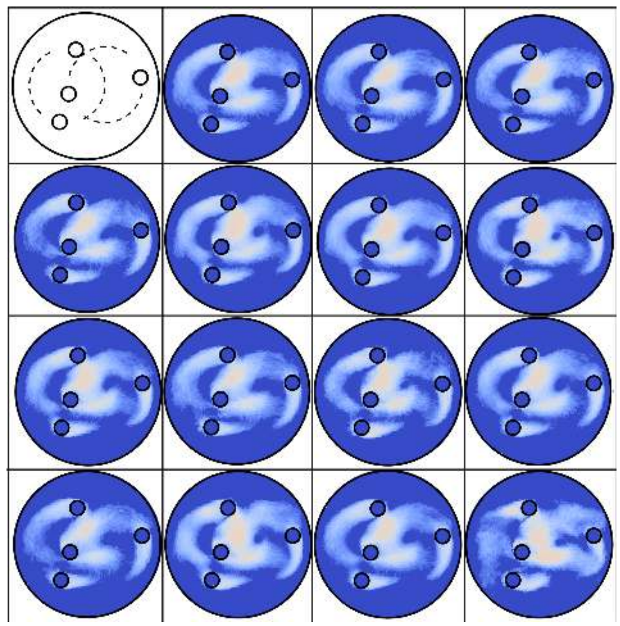

Figure 1: A set of 15 stirring simulations, showing velocity magnitudes for different fluid properties. Visual comparison of ensemble members in a list view is difficult.

The idea to move away from visualizing a single concrete solution to analyzing a family of outcomes is not entirely novel to the field of scientific visualization: Since the mid 1990s, visualization researchers have developed methods to visualize uncertainty and errors in data. However, there is a key difference between uncertain data and ensemble data – uncertain data encodes the probability distributions of values throughout a data set, allowing the identification of a "most likely" or "most common" output, while containing no information about how different outcomes relate. Ensembles, on the other hand, present us with concrete distributions of data, where each outcome can be uniquely associated with a specific run or set of simulation parameters.

This discrete character together with the ability to relate outputs to specific inputs is what makes ensembles so valuable to domain experts. Our challenge is to develop visualization techniques and tools to extract and highlight commonalities, differences, and trends in the set of ensemble members and to allow scientists to discover conceptual drawbacks or the value of simulation models or specific parameter choices.

Visualization of Ensemble Data

Recently, researchers have started to look into the visualization of ensemble data. The few existing approaches can be classified as

- *Feature-Based:* Features are extracted from individual ensemble members and compared across the ensemble.
- *Location-Based:* Ensemble comparison is performed at fixed locations in the data set.

Due to the existence of a variety of prediction models, weather and climate research is one of the central driving forces behind the creation of simulation ensembles. Predictions of climate events rely on a large number of external influences (parameters) and are generally associated with a certain probability of occurrence. This is because there is not only one possible outcome to existing prediction models, but rather a spectrum of possibilities. Sanyal et al. [5] approach the visualization of ensembles of numerical weather simulations by extracting sets of isocontour lines and designing glyphs that illustrate local variances. The so-called *spaghetti plots* created by rendering sets of isocontour lines allow feature-based comparison between members of the ensemble and give a basic impression of how ensembles agree for a given scalar value. A sample image of their technique is shown in Figure 2. In three dimensions, slicing can help create an impression of differences in isosurfaces, as demonstrated by Alabi et al. [1] and shown in Figure 3. As you can see, even simple tasks such as rendering spaghetti plots becomes a visualization challenge in three dimensions.

Location-based methods have largely been used to compute statistical properties of the ensemble throughout the domain. Common statistical measures then provide insights into outliers, agreement or disagreement between members of the ensemble. Means and variances of scalar quantities in climate simulations have been employed by Potter et al. [4], see Figure 4. Notions of variances can also be employed to detect agreement and disagreement in ensembles for arbitrary flow simulations [3]. Figure 5 shows an example of CFD variance-based coloring and trends querying. This provides a more concise representation of a set of flow simulations, than provided by list-based visualizations as shown in Figure 1. Ensemble visualizations can be made especially expressive if (some) ground truth data is available. This allows for an estimation of predictive uncertainty of the ensemble and can aid in identifying outliers in the set of ensemble members [2], see Figure 6. Gosink et al. also propose a visualization of parameter sensitivity, which is a key component of efficient parameter space analysis.

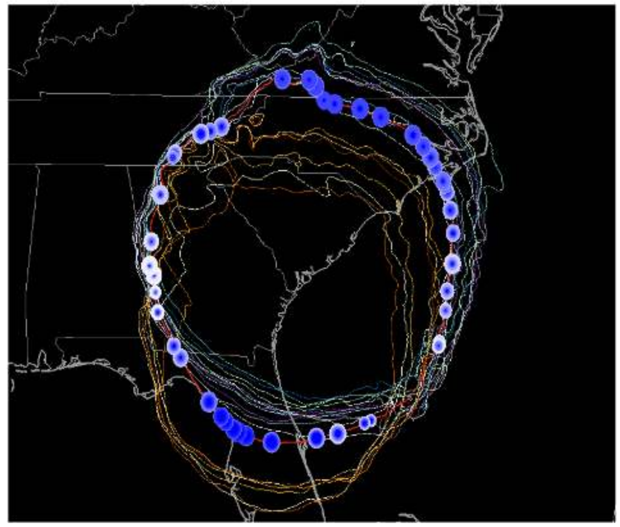

Figure 2: A spaghetti plot of perturbation pressure isocontours along with uncertainty glyphs showing deviation. Ensembles are particularly relevant in weather and climate prediction. Image courtesy of Sanyal et al. [5].

What the Future Holds

We see the need for visualization research to aid simulation scientists in the parameter space exploration task and to support accurate and robust decision

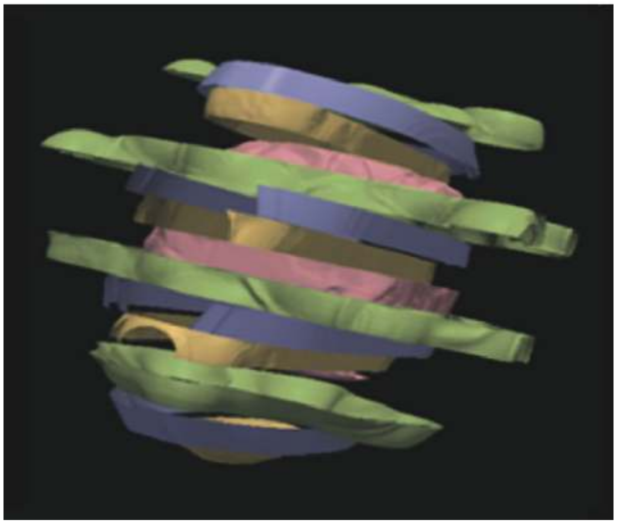

Figure 3: Rendering of sliced isosurfaces from four ensemble members. Feature-based comparison in 3D is challenging and can suffer from occlusion and visual complexity. Image courtesy of Alabi et al. [1].

making in a complex simulation environment. The need for effective visual analysis tools in this area has the potential to open and extend a large variety of research directions and application scenarios. In our correspondence with domain scientists we have identified several key requirements of effective ensemble visualization tools along with conceptual and technical issues. These challenges include

- *Conceptual - Finding the Answer:* Can we help domain experts in finding a "most likely/best prediction" made by the ensemble?
- *Conceptual - Parameter Space:* How can we relate insights gained by ensemble visualization with locations in parameter space?
- *Conceptual - Perception:* How can we visualize such a multitude of data in a precise and easy-to-understand way?
- *Mathematical - Features:* What are statistical or geometric feature definitions that are relevant in the context of ensembles?
- *Mathematical - Metrics:* How can we compare members of the ensemble or their features?

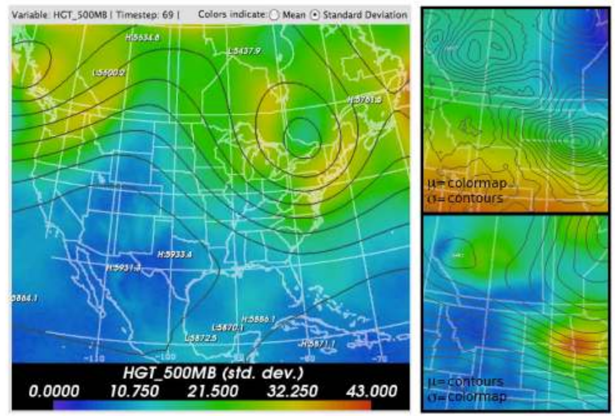

Figure 4: Mean and standard deviation shown as a combination of isocontouring and color-mapping. Correlations of the two variables are conveyed implicitly. Image courtesy of Potter et al. [4].

- *Algorithmic - Data Complexity:* How do we handle the immense increases in memory requirement and data complexity?
- *Algorithmic - Exploration:* How do we enable goal-driven exploration and analysis of parameter space and parameter sensitivity?

One direction we regard as especially challenging is the visualization and exploration of multi-dimensional parameter spaces. In this direction we are investigating how techniques from high-dimensional data visualization can help in making the connection between ensemble and parameter-space analysis. Specifically, the question whether and how recent work in computational steering, parameter-space exploration (e.g., Waser et al. [6]) and multivariate analysis may be applied to complex ensemble visualization problems remains to be answered.

Conclusions

Providing domain scientists with visualization solutions for ensemble data will be a key factor in improving analysis performance in complex simulation environments. Solving the inherent visualization challenges will increase the speed with which scientists

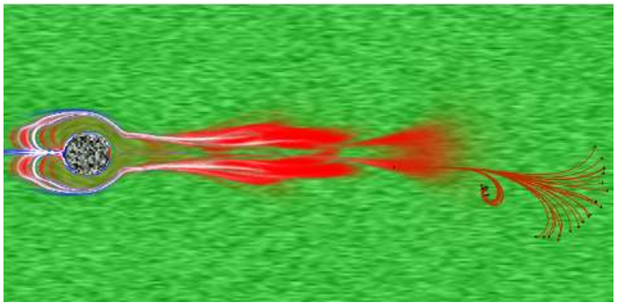

Figure 5: Simulation agreement and disagreement in an ensemble of CFD simulations is visualized through color mapped transport variance. Trends are identified as pathline clusters. A large number of ensemble members may lead to cluttered pathline renderings. Image courtesy of Hummel et al. [3].

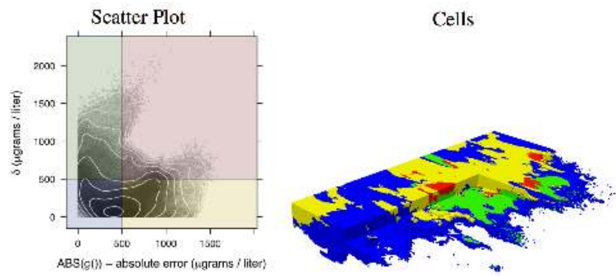

Figure 6: Evaluating predictive uncertainty of a single ensemble member through accuracy classification with respect to a probabilistic ground truth model. Ground truth data provides a basis for more complex analysis and classification of ensemble data. Visualizing complex classifications in 3D proves challenging. Image courtesy of Gosink et al. [2].

can explore, adapt, and validate simulation models. We expect the visualization community to engage in solving this challenging task, and thereby improve the robustness and reliability of simulation-based prediction and decision-making.

Acknowledgements

This work was supported in part by the NSF (IIS 0916289 and IIS 1018097), and the Office of ASCR, Office of Science, through DOE SciDAC contract DE-FC02-06ER25780, and contract DE-FC02-12ER26072 (SDAV).

References

[1] O. S. Alabi, X. Wu, J. M. Harter, M. Phadke, L. Pinto, H. Petersen, S. Bass, M. Keifer, S. Zhong, C. Healey, and R. M. Taylor II. Comparative visualization of ensembles using ensemble surface slicing. In *Proc. SPIE 8294, Visualization and Data Analysis 2012*, pages 82940U–82940U–12, 2012.

[2] L. Gosink, K. Bensema, T. Pulsipher, H. Obermaier, M. Henry, H. Childs, and K. I. Joy. Characterizing and visualizing predictive uncertainty in numerical ensembles through Bayesian model averaging. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2013.

[3] M. Hummel, H. Obermaier, C. Garth, and K. I. Joy. Comparative visual analysis of Lagrangian transport in CFD ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2013.

[4] 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 *IEEE Workshop on Knowledge Discovery from Climate Data: Prediction, Extremes.*, pages 233–240, 2009.

[5] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1421–1430, 2010.

[6] J. Waser, H. Ribičić, R. Fuchs, C. Hirsch, B. Schindler, G. Blöschl, and M. Gröller. Nodes on ropes: A comprehensive data and control flow for steering ensemble simulations. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):1872–1881, 2011.