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 finding the correct parameters? Imagine going through the same ordeal with unbelievably complex simulation models used for predicting physical phenomena. Scientists have fought this battle for many years and have long been sick of sitting, waiting, and restarting.

Luckily, the increased availability of vast computation resources and new computation strategies offer a solution to this dilemma. Scientists can run 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 visualization’s most important new areas, and we expect it to have a wide impact on our field in the next few years. The goal is to develop expressive visualizations of an ensemble’s properties to support scientists in this demanding parameter-space exploration.

The move from visualizing a single concrete solution to analyzing a family of outcomes isn’t entirely new to scientific visualization. Since the mid-1990s, visualization researchers have developed methods to visualize uncertainty and errors in data. However, there’s a key difference between uncertain data and ensemble data. Uncertain data encodes the probability distributions of values throughout a dataset, allowing identification of the most likely or most common output, while containing no information about how different outcomes relate. Ensembles, on the other hand, present concrete distributions of data, in which 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. We have two main challenges. The first is to develop visualization techniques and tools to extract and highlight commonalities, differences, and trends in ensemble members. The second is to enable scientists to discover conceptual drawbacks or the value of simulation models or specific parameter choices.

Visualizing Ensemble Data
The few existing ensemble visualization approaches fall into two categories:

- Feature-based visualization extracts features from individual ensemble members and compares them across the ensemble.
- Location-based visualization compares ensemble properties at fixed locations in the dataset.

Owing to the variety of prediction models, weather and climate research is a central driving force behind the creation of simulation ensembles. Predictions of climate events rely on many external influences (parameters) and are generally associated with a certain probability of occurrence. This is because prediction models produce not just one possible outcome but a spectrum of possibilities.

Feature-Based Visualization
Jibonananda Sanyal and his colleagues visualized ensembles of numerical weather simulations by extracting sets of isocontour lines and designing glyphs that illustrate local variances. By rendering the sets, they created spaghetti plots that allow feature-based comparison between ensemble members and give a basic impression of how ensembles agree for a given scalar value. Figure 2 shows a sample image.

In 3D, slicing can help emphasize the differences
in isosurfaces, as Oluwafemi Alabi and her colleagues demonstrated (see Figure 3). As you can see, even simple visualization tasks such as rendering spaghetti plots become a challenge in 3D.

**Location-Based Visualization**

Researchers have used location-based methods largely to compute an ensemble’s statistical properties throughout the domain. Common statistical measures then provide insights into outliers and agreement or disagreement between ensemble members. Kristin Potter and her colleagues’ climate simulations employed means and variances of scalar quantities (see Figure 4). Researchers have also employed notions of variances to detect agreement and disagreement in ensembles for arbitrary flow simulations. Figure 5 shows computational-fluid-dynamics variance-based coloring and trend querying. This represents a set of flow simulations more concisely than list-based visualizations such as Figure 1 do.

Ensemble visualizations can be especially expressive if (some) ground truth data is available. As Luke Gosink and his colleagues showed, this allows estimation of the ensemble’s predictive uncertainty and can help identify ensemble members that are outliers (see Figure 6). Gosink and his colleagues also proposed visualizing parameter sensitivity, a key component of efficient parameter-space analysis.

**What the Future Holds**

The need for effective visual-analysis tools for ensembles could open and extend a variety of research directions and application scenarios. Through our correspondence with domain scientists, we’ve 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 find the most likely or best prediction made by the ensemble?
- **Conceptual**—parameter space. How can we relate insights gained by ensemble visualization to locations in parameter space?
- **Conceptual**—perception. How can we visualize such a multitude of data in a precise, easy-to-understand way?
- **Mathematical**—features. What statistical or geometric feature definitions are relevant for ensembles?
- **Mathematical**—metrics. How can we compare ensemble members or their features?
- **Algorithmic**—data complexity. How do we handle the immense increases in memory requirements and data complexity?
- **Algorithmic**—exploration. How do we enable goal-driven exploration and analysis of the parameter space and parameter sensitivity?
One especially challenging direction is the visualization and exploration of multidimensional parameter spaces. We’re investigating how high-dimensional data visualization techniques can help connect ensemble and parameter-space analysis. Specifically, a major question still needs answering: whether and how recent research in computational steering, parameter-space exploration (for example, Jürgen Waser and his colleagues’ research), and multivariate analysis is applicable to complex ensemble visualization problems.

Providing domain scientists with ensemble visualization solutions 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 can explore, adapt, and validate simulation models. We expect the visualization community to engage in this pursuit, thereby improving the robustness and reliability of simulation-based prediction and decision making.

Acknowledgments
This research was supported partly by the US National Science Foundation (contracts IIS 0916289 and IIS 1018097) and the Office of Advanced Scientific Computing Research, Office of Science, through US Department of Energy SciDAC (Scientific Discovery through Advanced Computing) contracts DE-FC02-06ER25780 and DE-FC02-12ER26072 (SDAV; Scalable Data Management, Analysis, and Visualization).

References


Harald Obermaier is a postdoctoral researcher in the Computer Science Department at the University of California, Davis. Contact him at hobermaier@ucdavis.edu.

Kenneth I. Joy is a professor in the Computer Science Department at the University of California, Davis. Contact him at kijoy@ucdavis.edu.

Contact department editor Theresa-Marie Rhyne at theresamarierhyne@gmail.com.