Visualizing Variance and Predictive Uncertainty in Flow Ensembles

1) University of California, Davis 2) Pacific Northwest National Laboratory 3) University of Kaiserslautern 4) University of Oregon

Flow Ensembles

A numerical ensemble is a collection of simulation runs obtained by varying simulation parameters or model characteristics. The data contained in flow ensembles allows us to develop tools that improve our understanding of parameter interactions and model implications in complex flow simulations.

In flow ensembles, research is concerned with investigating the effects of parameter or model changes on fluid transport properties, such as flow divergence, mixing, or the presence of turbulence.

Figure: Three simulation runs of a stirring simulation [1]. All runs produce slightly different simulation outcomes.

Analysis Goals and Challenges

We identify two related classes of ensemble analysis problems:

1) Model selection: Given an ensemble of data models – often using experts to filter from an extremely large base: How do we tell which is the “correct” underlying model?

2) Parameter selection: Assuming a trusted model, we can create an ensemble of runs by sampling parameter-space: What parameters create “correct” results?

Intermediate analysis goals include being able to:

• Extract and convey trends, outliers [2,3]
• Identify commonalities, differences, and variances [2,3]
• Highlight predictive uncertainties [2]

The nature of ensemble data poses several interesting challenges to the analysis and visualization process [1]:

• Data size and complexity: How can we process and visualize large amounts of complex data?
• Feature definition and extraction: What types of features are relevant in the ensemble context?
• Visual complexity: How can we effectively represent outputs of multiple simulations?
• Interaction complexity: How do we relate inputs and outputs to allow for intuitive interaction?

Variance and Predictive Uncertainty

How can ensemble analysis aid model and parameter selection?

In the presented application scenarios, two properties of time-varying flow fields are relevant: Transport variance and predictive uncertainty of chemical concentrations.

Transport variance can be used to identify differences and commonalities of time-varying flow fields in an ensemble. Further analysis of particle distributions allows the identification of outliers and trends. This allows us to highlight and study the effects of parameter changes and provide means for target-driven parameter selection.

We propose a different approach for model selection:

Given a set of real-world measurements and an ensemble of models, we construct a probabilistic estimate of the ground truth, using Bayesian Model Averaging. Based on this estimate, we identify different classes of predictive uncertainty within an ensemble and draw conclusions about model properties.

Figure: Principal Components Analysis of advected tracer positions reveals positional variance in fluid transport of a time-varying flow ensemble.

Results and Work in Progress

• Scalable ensemble analysis and visualization
  • Variance and uncertainty classification
  • Model and parameter selection
  • Interactive ensemble exploration
  • Implemented through VisIt

• Ongoing work
  • Full extension to time-varying data
  • Investigate modality of distributions
  • Stronger integration of parameter space

References