The Scalable Data Management, Analysis, and Visualization Institute http://sdav-scidac.org

Visualizing Variance and Predictive Uncertainty in Flow Ensembles

H. Obermaier[1] and K. I. Joy[1]. Joint work with: L. Gosink[2], K. Bensema[1], T. Pulsipher[2], M. Henry[2], H. Childs[4], M. Hummel[1,3], C. Garth[3]

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.

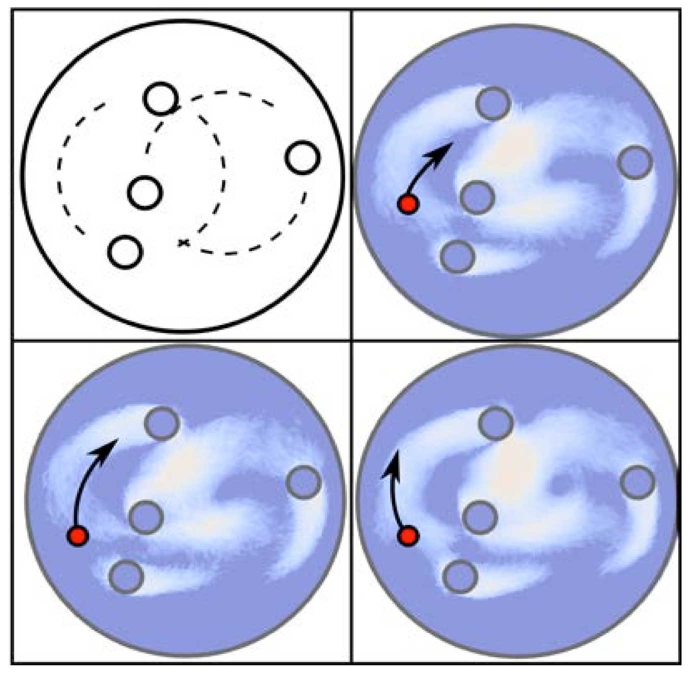


Figure: Three simulation runs of a stirring simulation [1]. All runs produce slightly different simulation outcomes.

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**.

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.

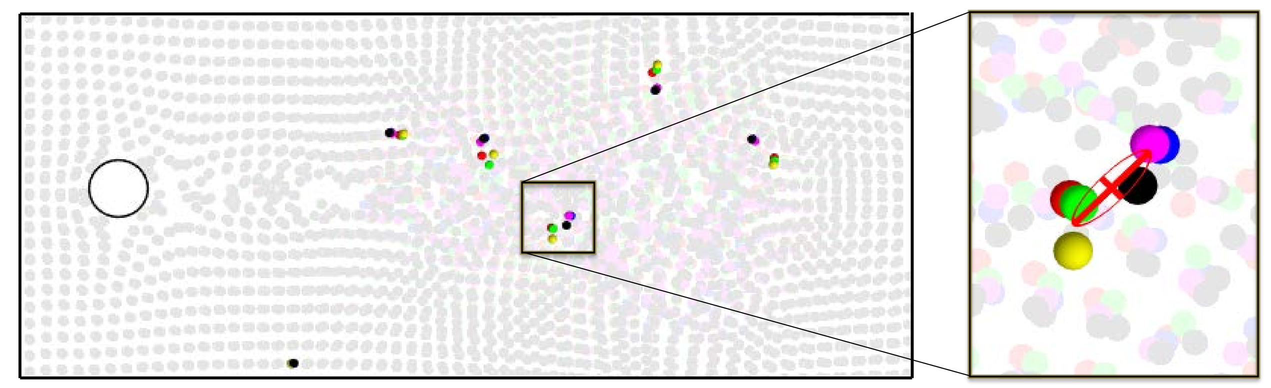


Figure: *Principal Components Analysis* of advected tracer positions reveals positional variance in fluid transport of a time-varying flow ensemble.

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.

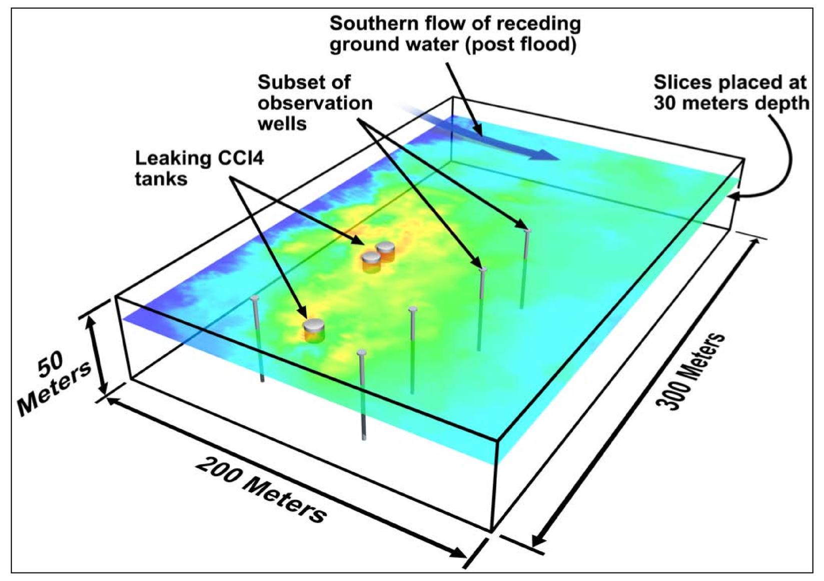
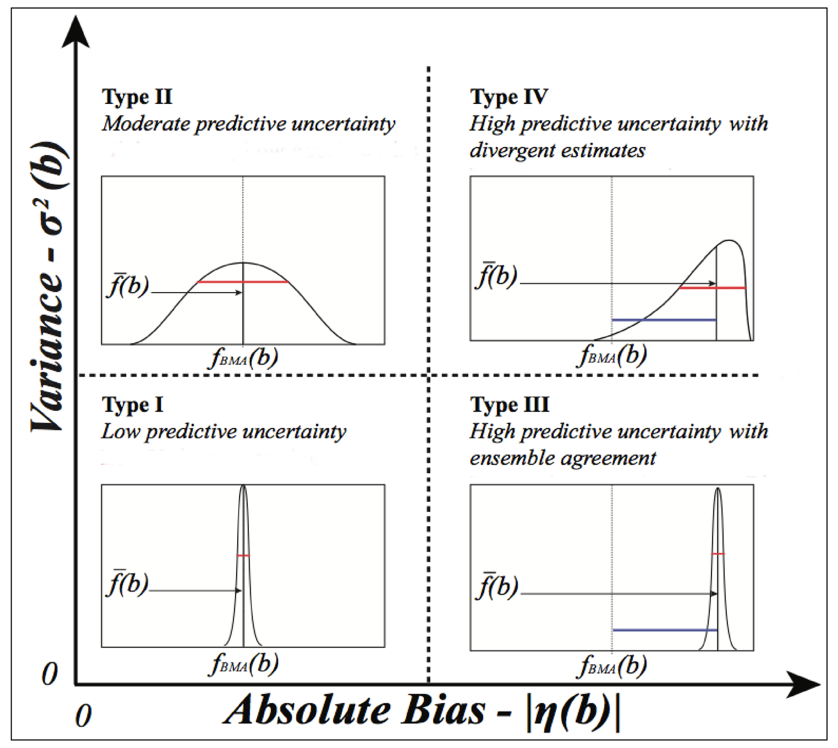


Figure: Aqueous transport of carbon tetrachloride. Subsiding groundwater moves CCl_4 away from the containment tanks.



References

[1] H. Obermaier and K. I. Joy: Future Challenges for Ensemble Visualization, IEEE CG&A, May/June 2014.

[2] L. Gosink, K. Bensema, T. Pulsipher, H. Obermaier, M. Henry, H. Childs, K. I. Joy. Characterizing and Visualizing Predictive Uncertainty in Numerical Ensembles Through Bayesian Model Averaging, IEEE TVCG, vol. 19, no. 12, pp. 2703-2712, 2013.

[3] M. Hummel, H. Obermaier, C. Garth, K. I. Joy. Comparative Visual Analysis of Lagrangian Transport in CFD Ensembles, IEEE TVCG, vol. 19, no. 12, pp. 2743-2752, 2013.
best paper award

Visualization and Analysis

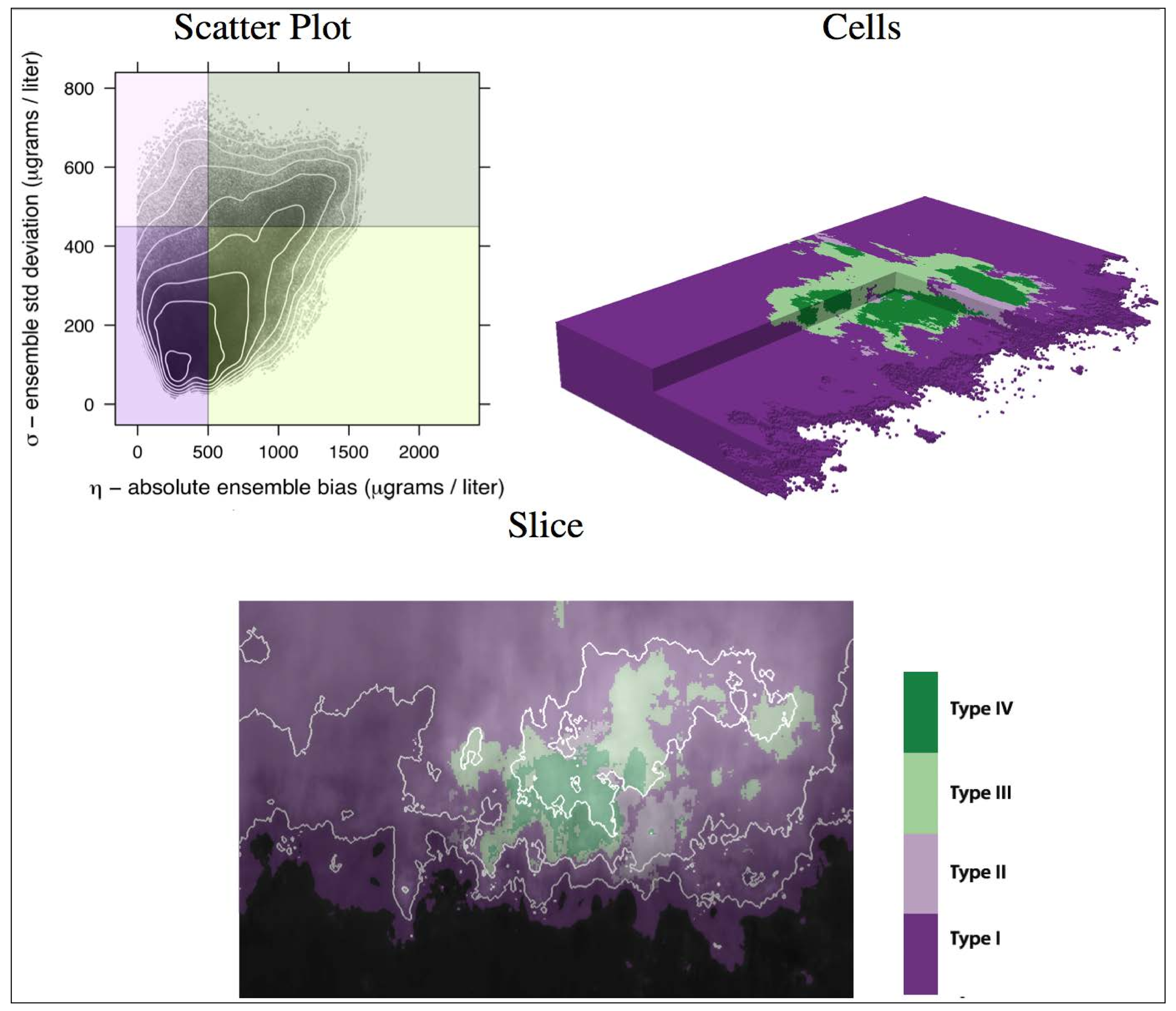


Figure: CCl_4 transport study. Ensemble analysis and visualization performed in VisIt, showing classification of predictive uncertainty [2].

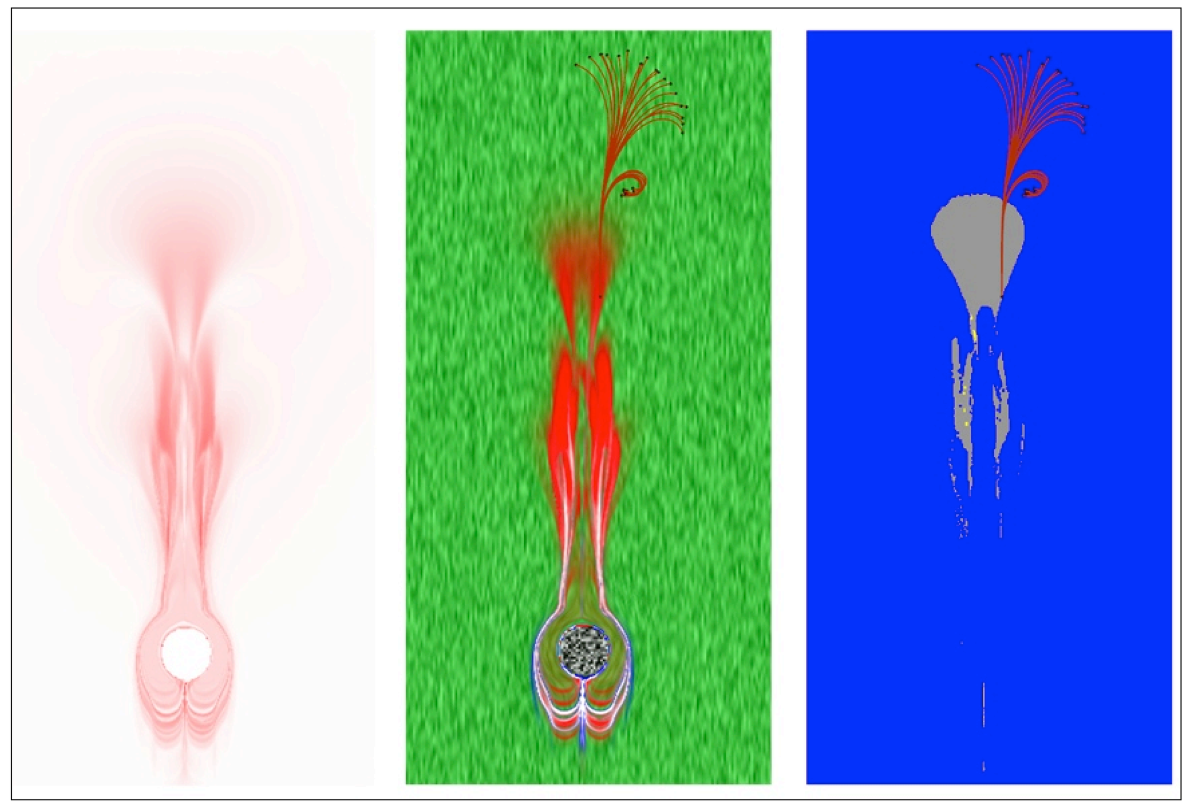


Figure: Transport variance and trends in an ensemble of heat convection simulations [3]. Path lines reveal two prevalent trends of transport behavior.

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