Visualization of Uncertainty

Kristin Potter
Scientific Computing and Imaging Institute
University of Utah

TAMU

May 14, 2010

Advanced Computing and Scientific Data

Jaguar, ORNL

- More bandwidth, storage, & computational power
- Larger data sets:
  - Higher resolutions
  - Longer runs
- More sophisticated models

All this leads to huge amounts of complex data

Uncertainty in Data

- Scientific data sets are incomplete without indications of *uncertainty*
- Express error, accuracy, confidence level
- Sources include acquisition, transformation, sampling, quantization, interpolation, and visualization

SCI INSTITUTE

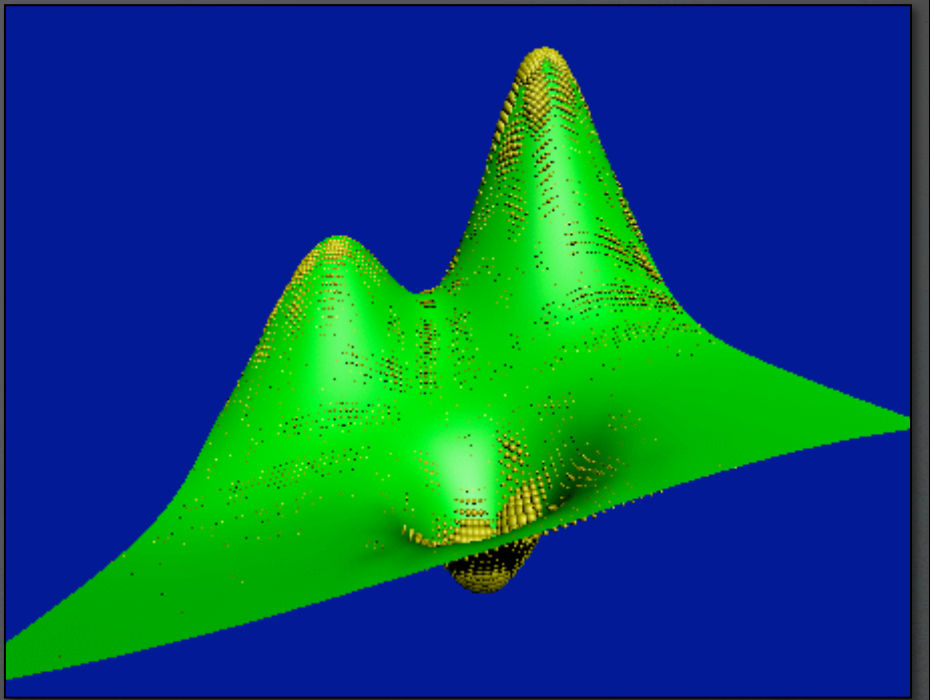
Types of Uncertainty

- Experimental Uncertainty
  - NIST defines uncertainty as standard deviation of a measurand*
- Geometric Uncertainty
- Simulation Uncertainty
- Visualization Uncertainty

* Barry N. Taylor and Chris E. Kuyatt. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. NIST Technical Note 1297, 1994.

Types of Uncertainty

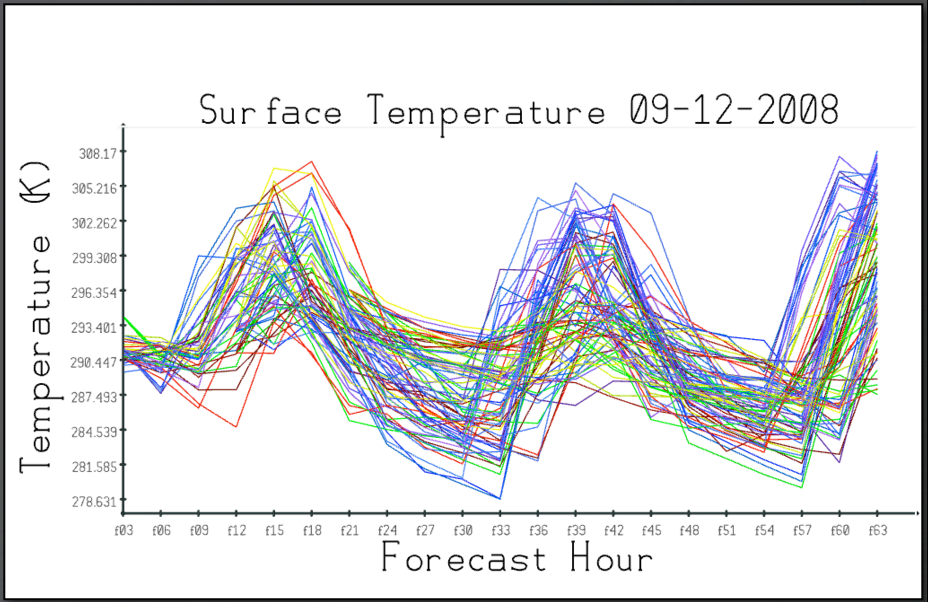
- Experimental Uncertainty
- Geometric Uncertainty
  - Unknowns in spatial positions
- Simulation Uncertainty
- Visualization Uncertainty



* S. Lodha, B. Sheehan, A. Pang and C. Wittenbrink. Visualizing geometric uncertainty of surface interpolants In Proc Graphics Interface '96 pp. 238--245. 1996.

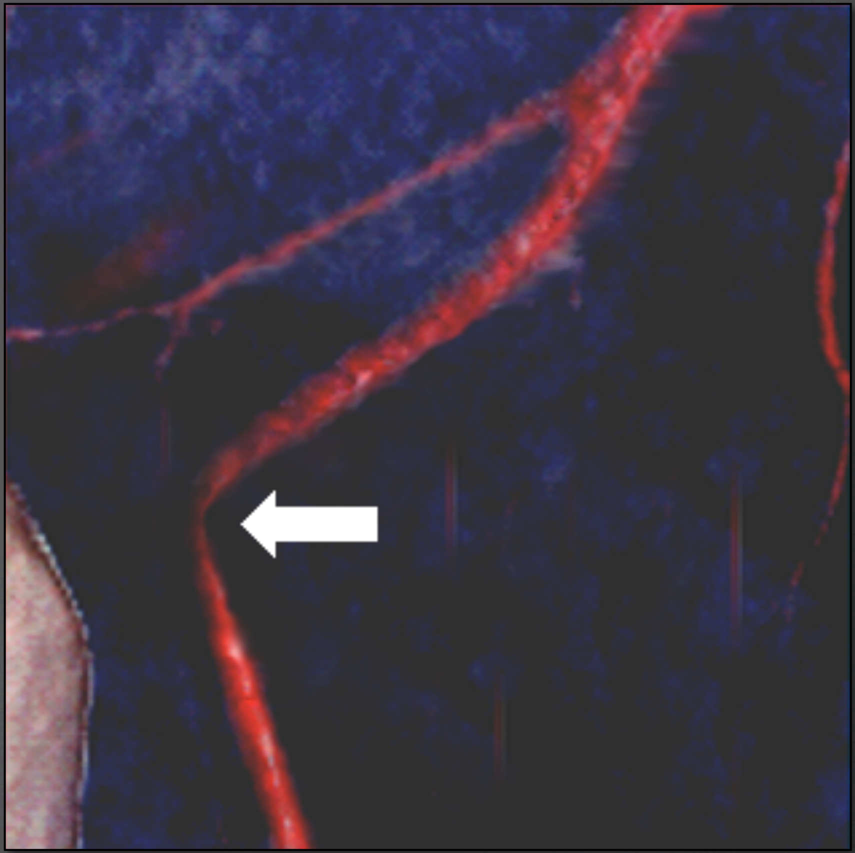
Types of Uncertainty

- Experimental Uncertainty
- Geometric Uncertainty
- Simulation Uncertainty
  - Multimodel, ensembles or non-determinininistic
- Visualization Uncertainty



Types of Uncertainty

- Experimental Uncertainty
- Geometric Uncertainty
- Simulation Uncertainty
- Visualization Uncertainty
  - Parameters of technique lead to differences



* C. Lundström, P. Ljung, A. Persson, and A. Ynnerman, Uncertainty Visualization in Medical Volume Rendering Using Probabilistic Animation, In IEEE TVCG, 13(6,) pp. 1648-1655, 2007,

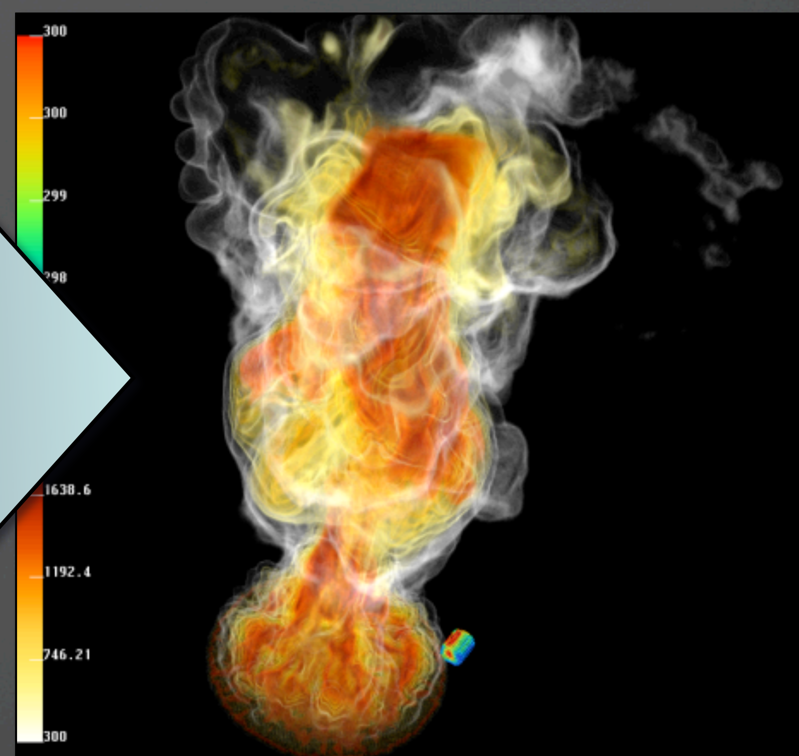
Types of Uncertainty

- Experimental Uncertainty
- Geometric Uncertainty
- Simulation Uncertainty
- Visualization Uncertainty
  - Parameters of technique lead to differences

* C. Lundström, P. Ljung, A. Persson, and A. Ynnerman, Uncertainty Visualization in Medical Volume Rendering Using Probabilistic Animation, In IEEE TVCG, 13(6,) pp. 1648-1655, 2007,

Visualization is Communication

- Translate data into images, “see” the data
- Brings out relationships & features in data
- Lets scientists communicate within their fields and out to others



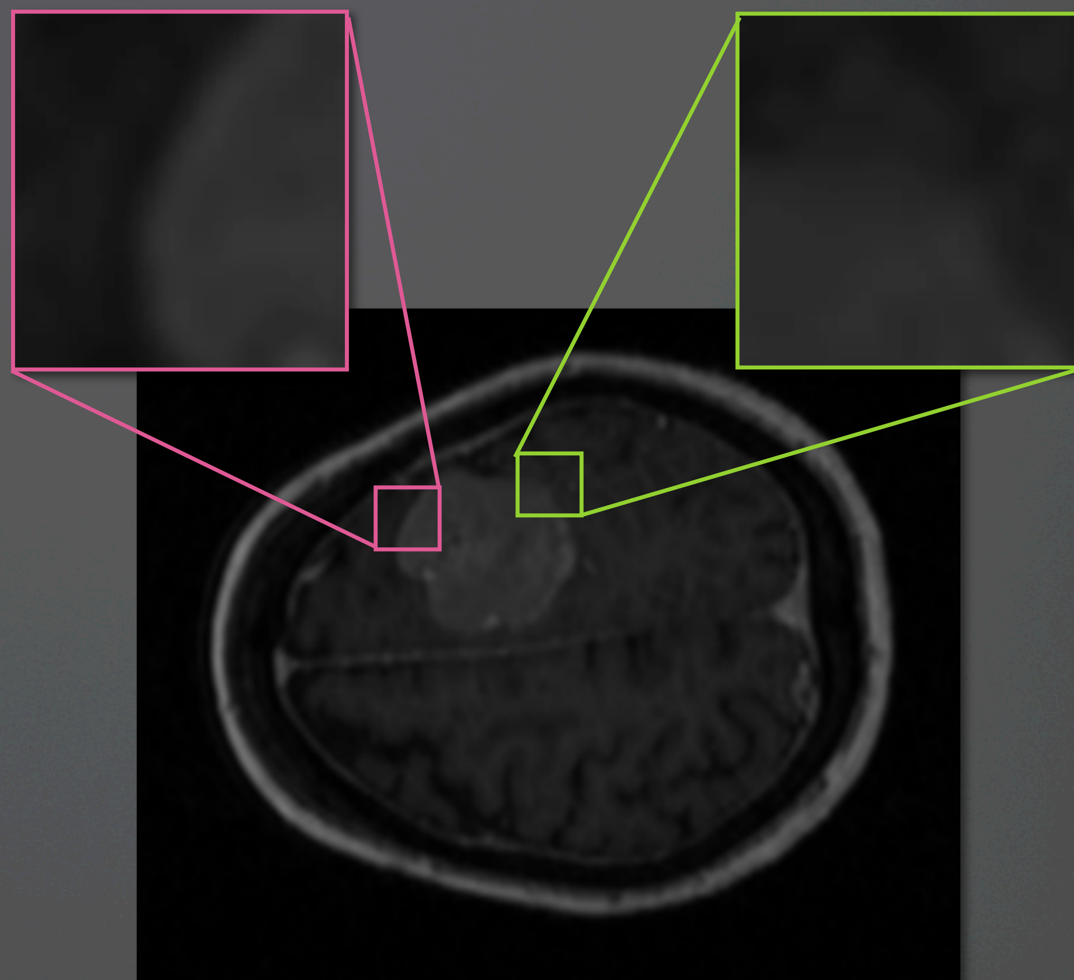
Uncertainty Visualization

- Visually depict uncertainties
- Faithfully present data
- Improve vis as a decision making tool
- Top visualization research problem *

* Chris R. Johnson.
Top Scientific Visualization Research Problems,
In *IEEE CG&A* 24(4) pp. 13--17, 2004.

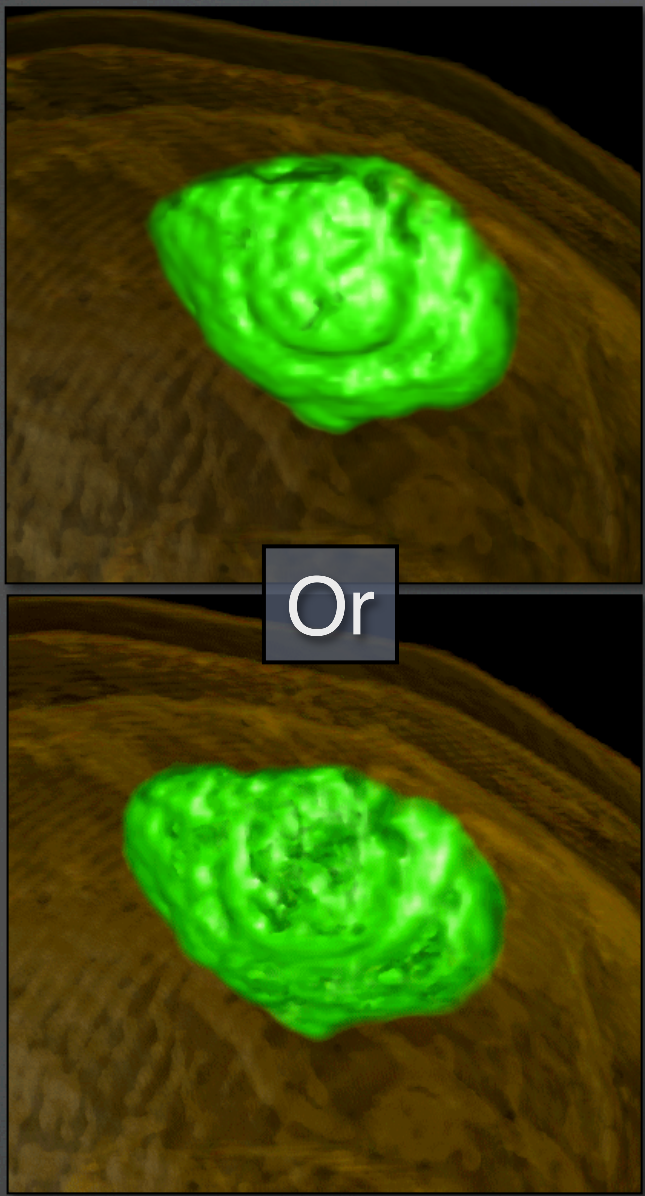
Brain Tumor Example

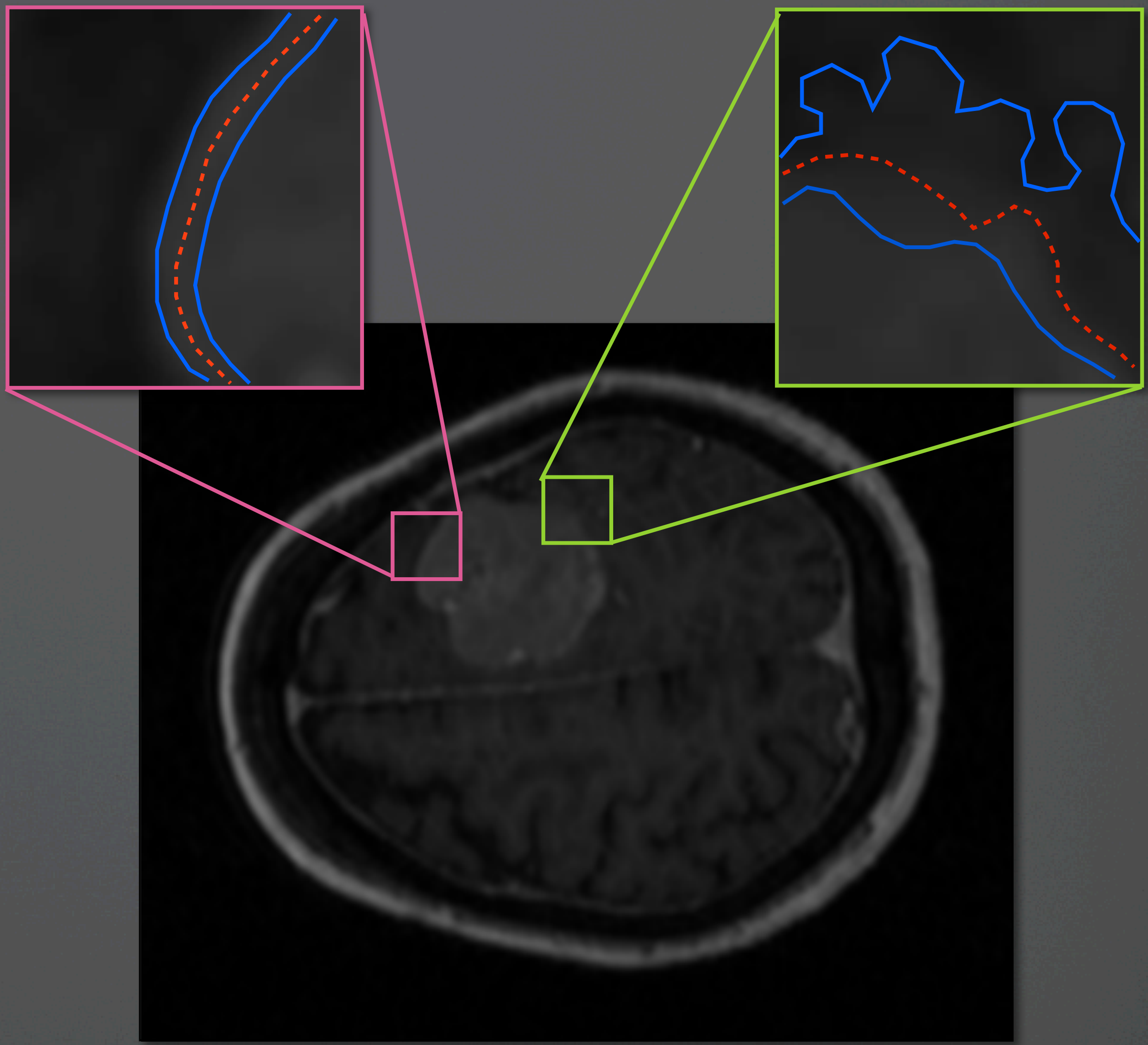
- Fuzzy boundaries exist in the data
- How to distinguish between tumor and gray matter?
- Pre-operative planning: Doctors (and patients!) need to know confidence of the line delinating tissue types



Whole Brain Atlas
http://www.med.harvard.edu/AANLIB/home.html

Brain Tumor Example - cont



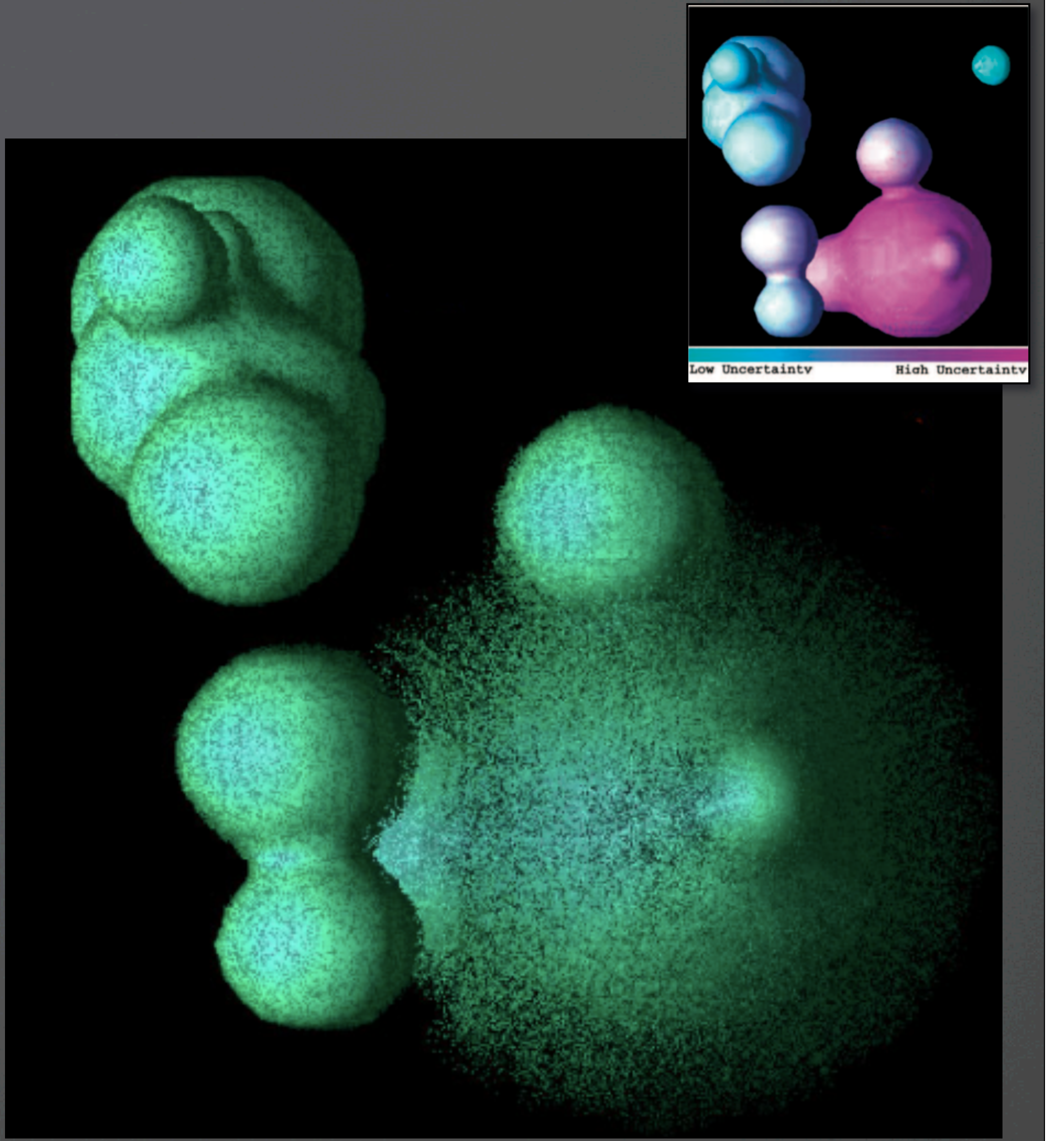


Why is Uncertainty Vis Hard?

- Not clear how to present uncertainty
- Increased visual clutter & complexity
- Data presentation may be obscured
- Increasing visual “uncertainty” can decrease understanding

Example of an "Uncertain" Image

- People can interpret blur and fuzz as uncertainty
- But they cannot **quantify** the amount of uncertainty from blur



Gevorg Grigoryan and Penny Rheingans. Point-Based Probabilistic Surfaces to Show Surface Uncertainty
In *IEEE TVCG*, 10(5), pp. 546--573, 2004.

Roadmap

- Summaryplot for 1D distribution data
  - Combination of graphical data analysis methods
  - Highlight feature characteristics & uncertainty
- Visualization of multidimensional distribution data
- Ensemble-Vis Framework for visual data analysis
- Current research and ongoing problems

Roadmap

- Summaryplot for 1D distribution data
- Visualization of multidimensional distribution data
  - Explore relationship between input parameters and resulting outcomes
  - Global qualitative view, local quantitative
- Ensemble-Vis Framework for visual data analysis
- Current research and ongoing problems

Roadmap

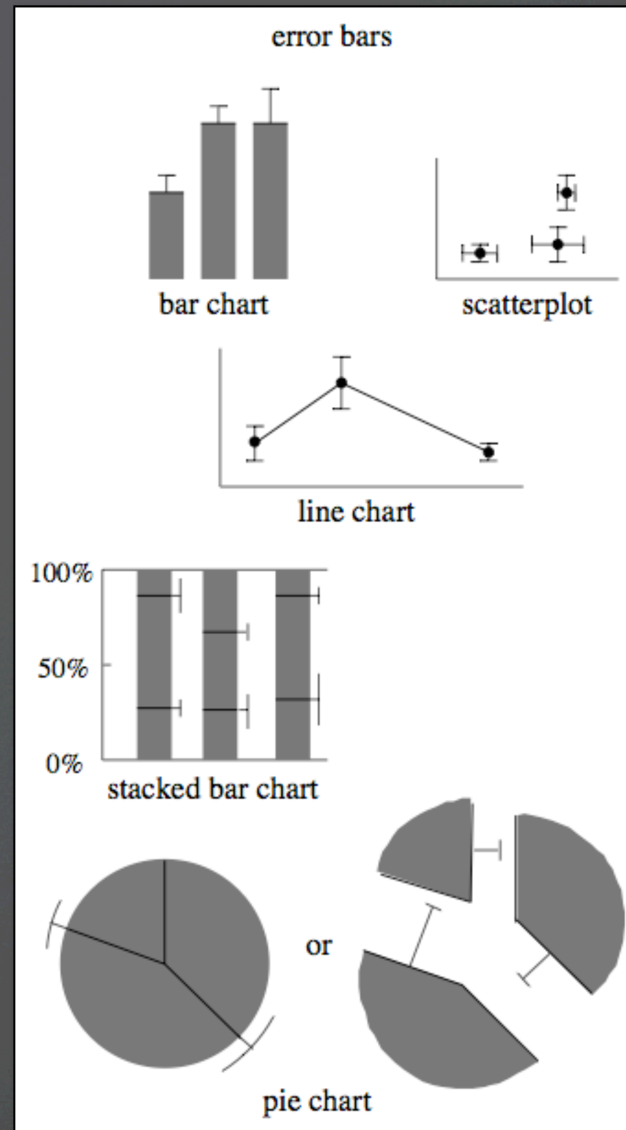
- Summaryplot for 1D distribution data
- Visualization of multidimensional distribution data
- Ensemble-Vis Framework for visual data analysis
  - User driven, component-based
  - Combine information and scientific visualizations
- Current research and ongoing problems

Roadmap

- Summaryplot for 1D distribution data
- Visualization of multidimensional distribution data
- Ensemble-Vis Framework for visual data analysis
- Current research and ongoing problems
  - Segmentation and classification of brainweb data

How do we usually see uncertainty?

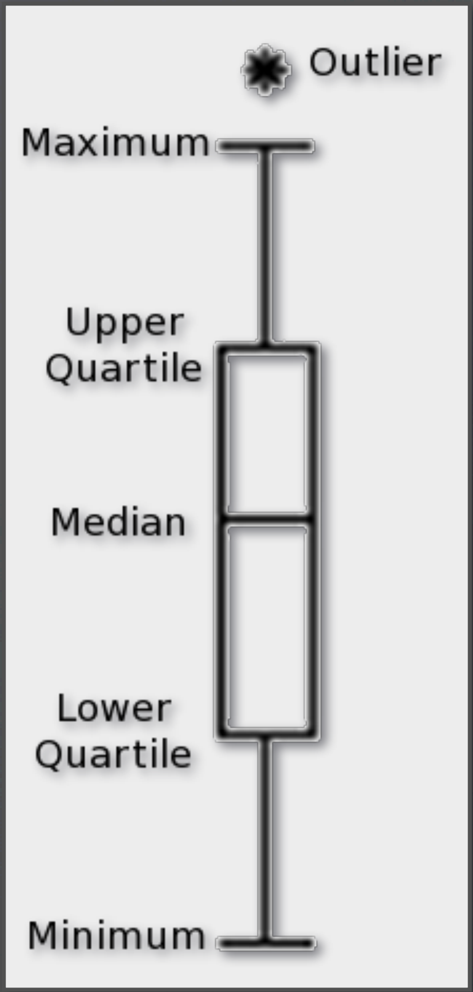
- Error bars
  - variation
  - simple addition to display



Chris Olston and Jock D. Mackinlay. Visualizing Data with Bounded Uncertainty. In *Proc InfoVis'02,* pp. 37-40, 2002.

How do we usually see uncertainty?

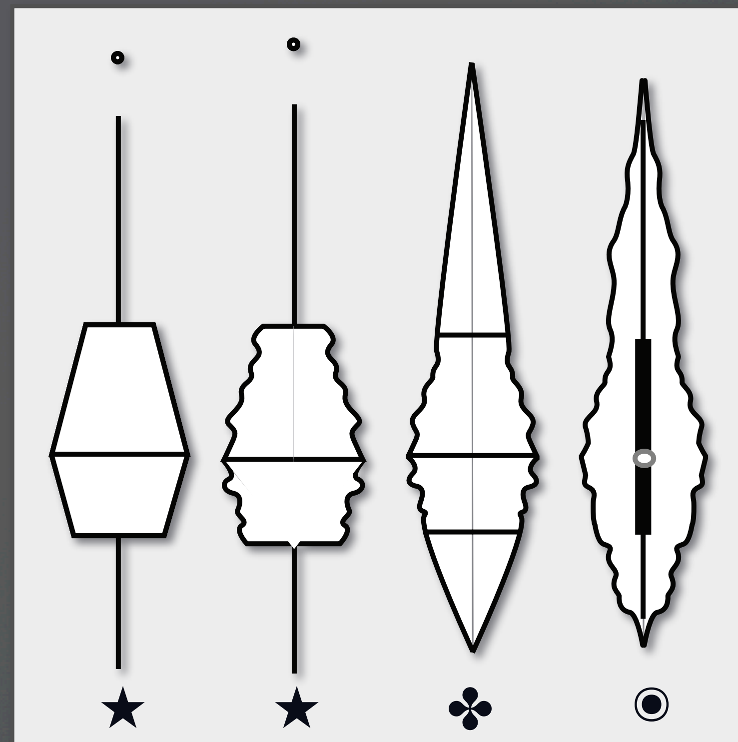
- Boxplots
  - Show range of data
  - Quartile range, including median
  - Outliers



John W. Tukey.
Exploratory Data Analysis.
Addison-Wesley, Reading, MA. 1977.

Boxplot Modifications

- Use the box sides to encode more information
  - Density information
  - Confidence levels
  - Skew & modality

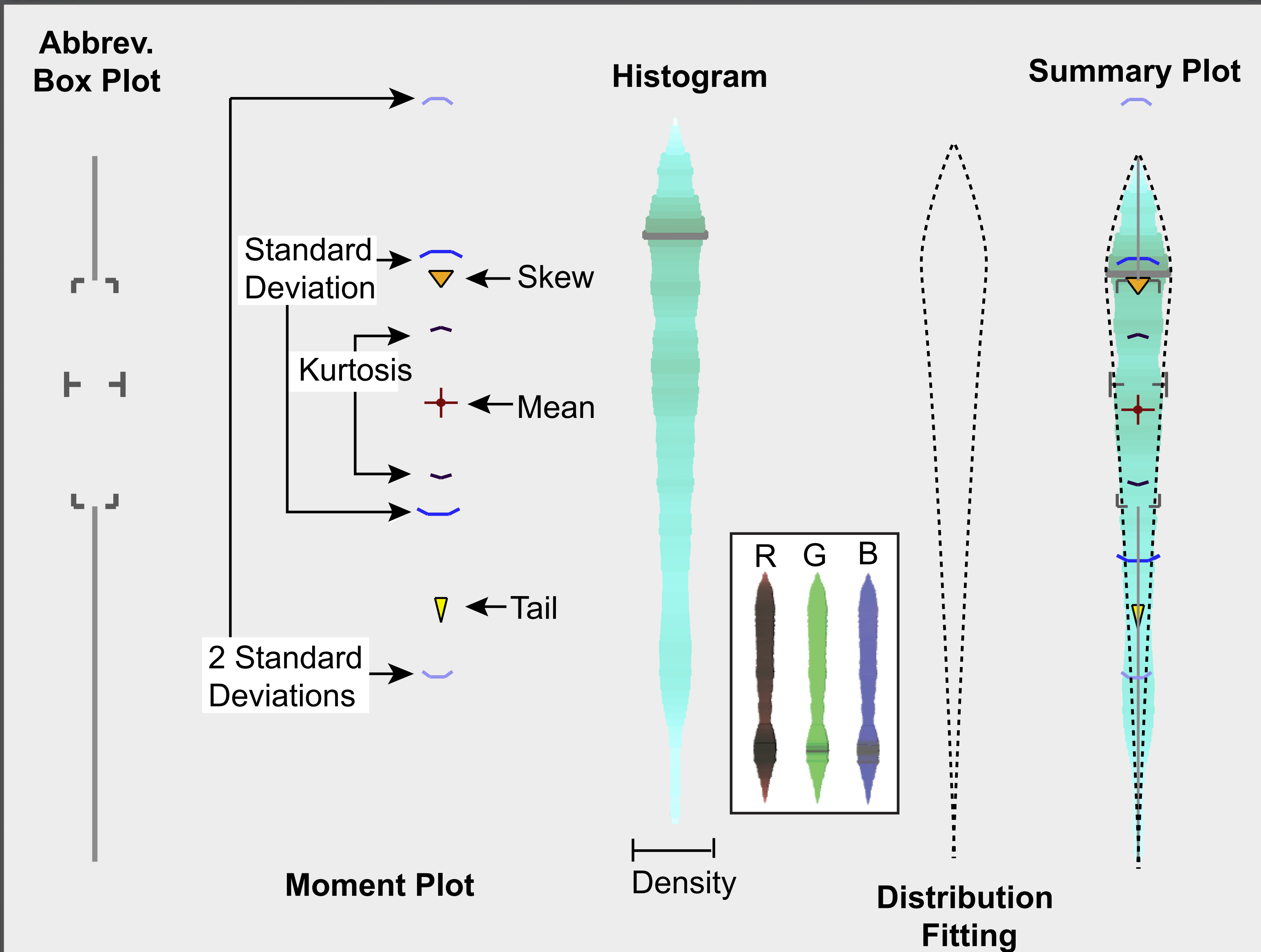


★ Yoav Benjamini.
Opening the box of a boxplot.
TAS 42(4) ,1988.

✤ W. Esty and J. Banfield.
The box-percentile pot.
JSS 8(17) 2003.

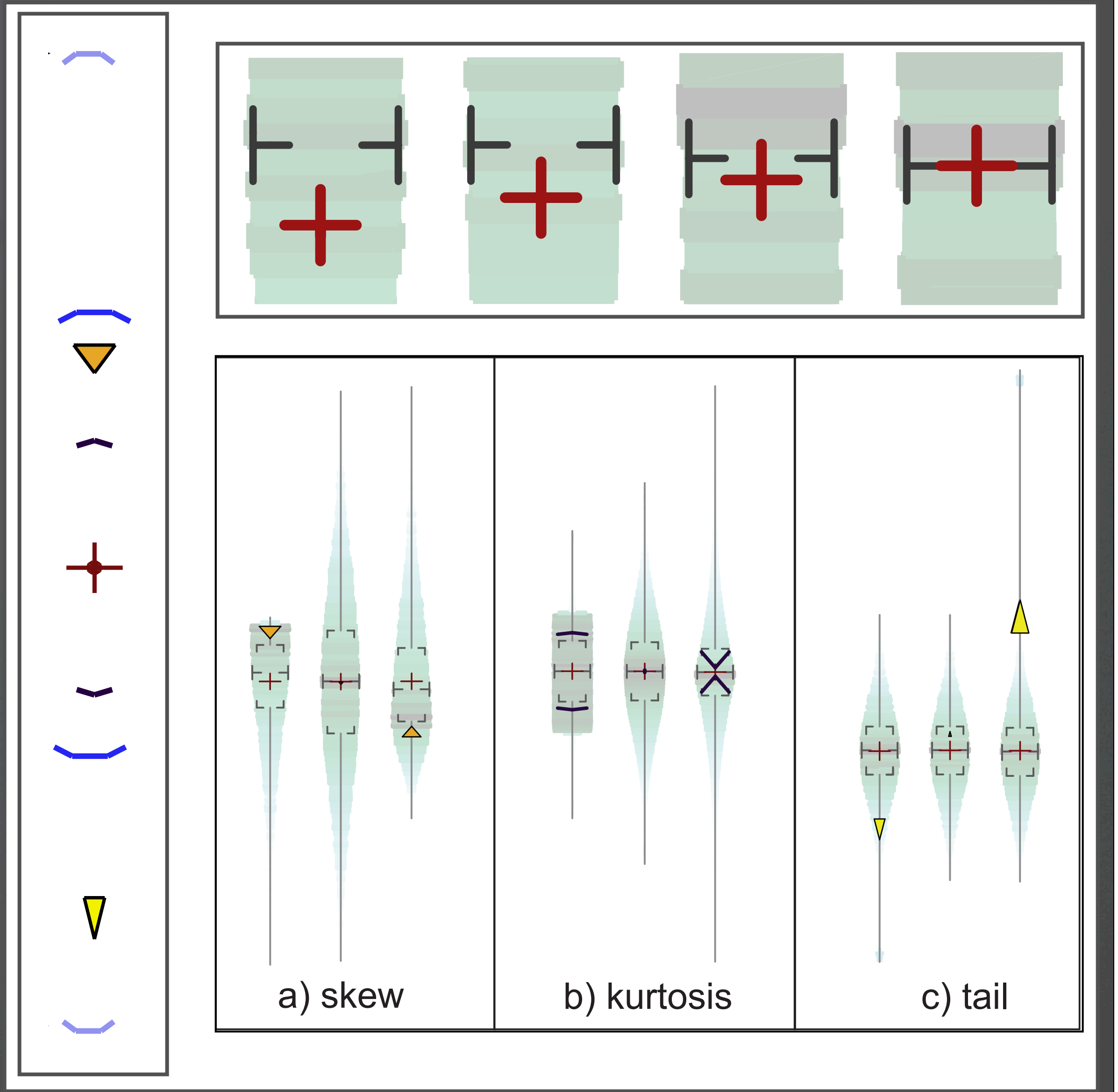
◉ J. Hintze, and R. Nelson.
Violin Plots: A Box Plot-Density Trace Synergism.
TAS 52(2) 1998.

The Summary Plot



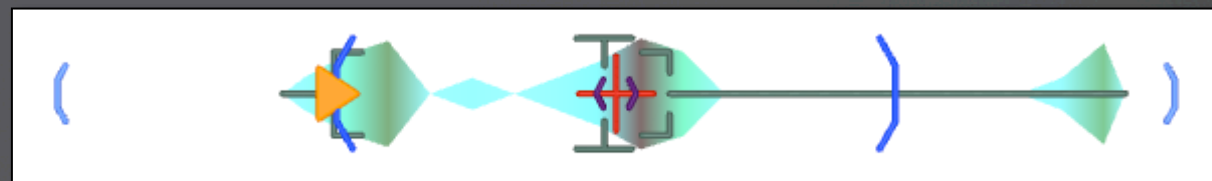
Moment Plot

- Statistical measures of feature characteristics
- Signature similar to boxplot
- Can express features hidden by boxplot (e.g. asymmetry)

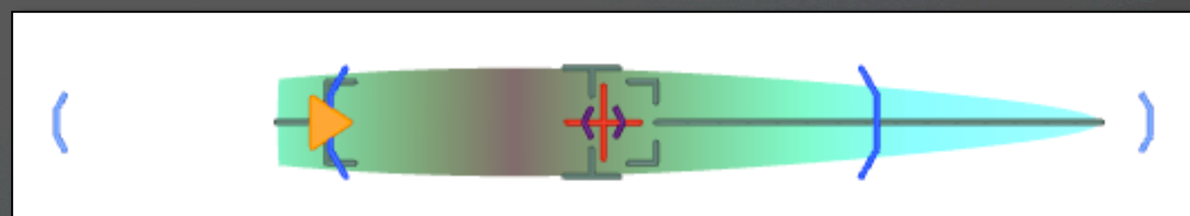


Density Plot

- Redundantly encode density through colormap and width
- Symmetric display on either side of plot
- Type of estimator influences display



Histogram, 20 bins, 84 samples

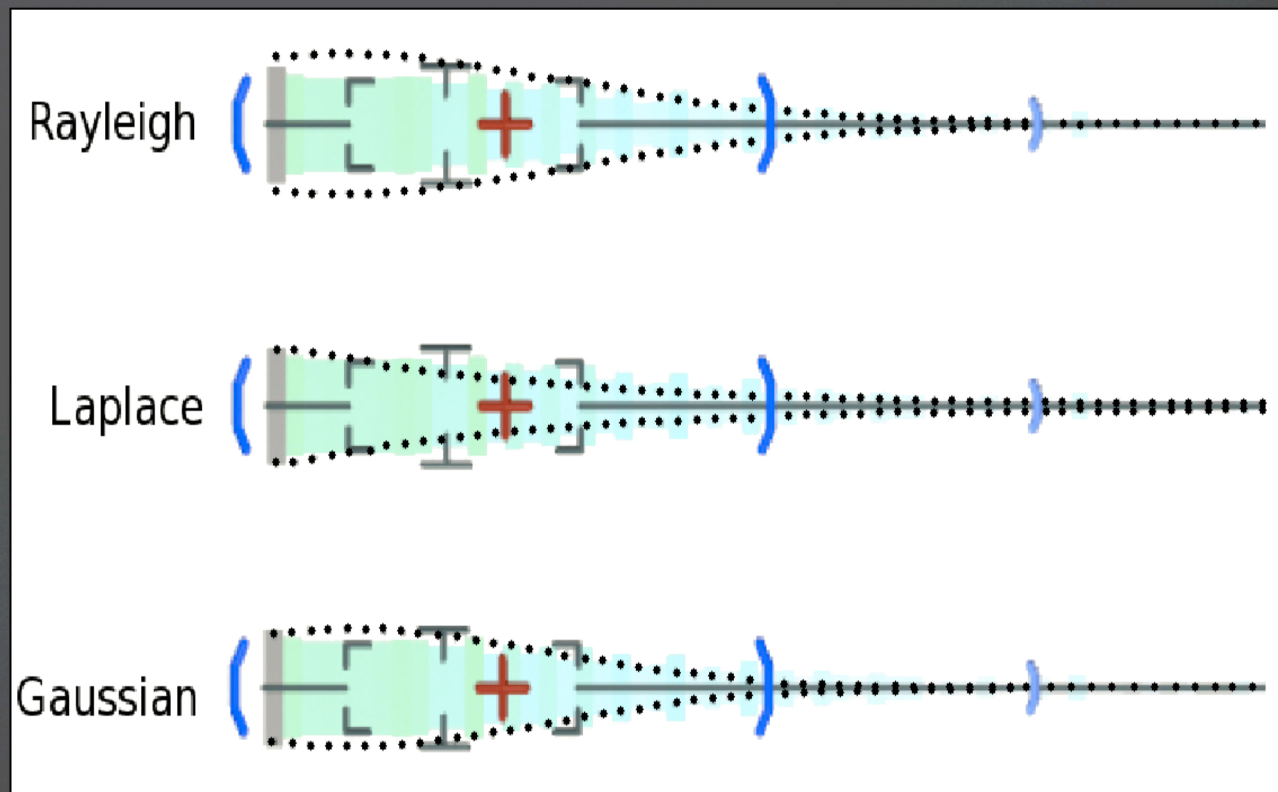


Kernel Density Estimation*

* Emmanuel Parzen.
On estimation of a probability density function and mode. The Annals of Mathematical Statistics, 33, 3, pp. 1065–1076, 1962.

Distribution Fitting

- Fit to canonical distributions from library
- Find a best fit
- Or fit to a chosen distribution



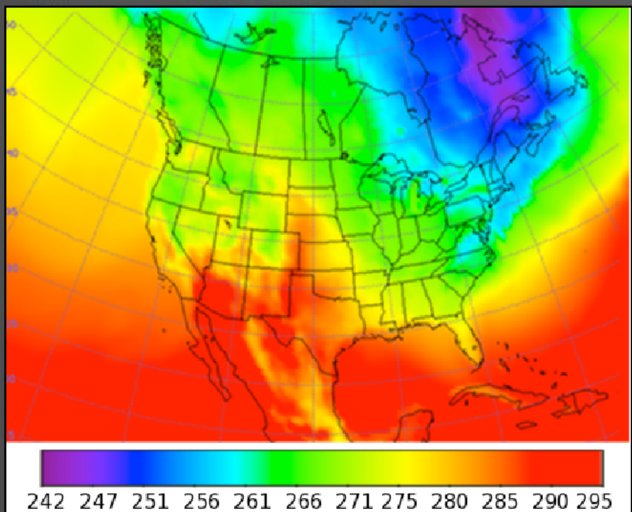
Using the Summary Plots

Short-Range Ensemble Forecasts (SREF)

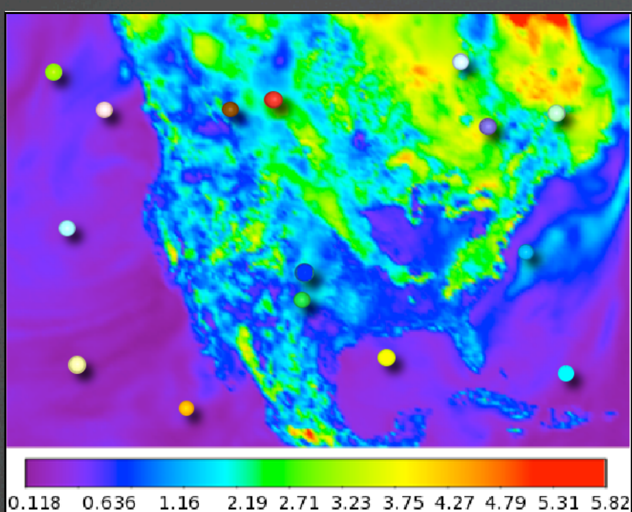
- Domain across North America
- Forecast weather variables out to ~3.5 days
- 4 models using perturbations in initial conditions & parameters

Short-range ensemble forecasting. http://wwwt.emc.ncep.noaa.gov/mmb/SREF/SREF.html.

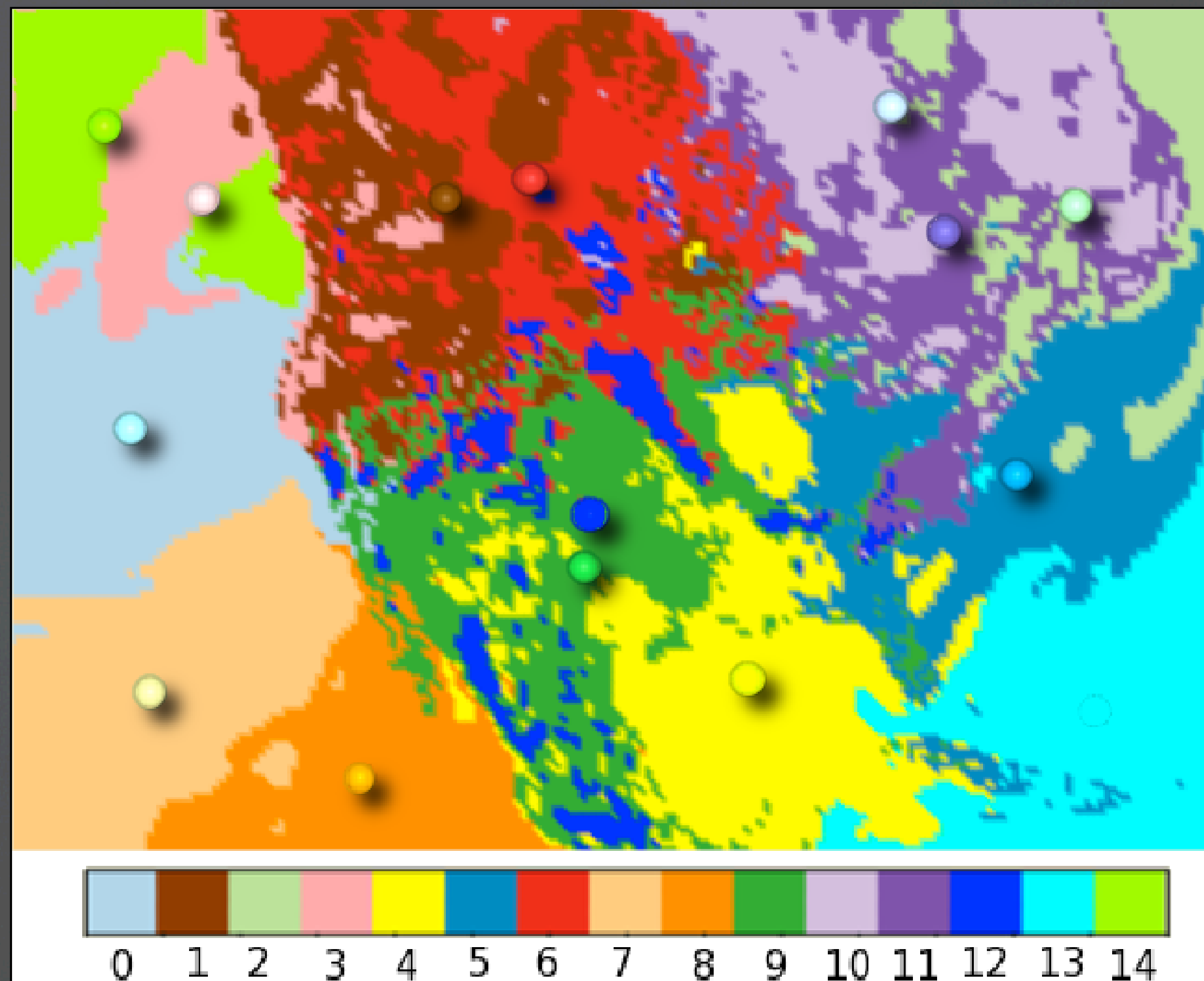
Choose Clusters Based on Variance



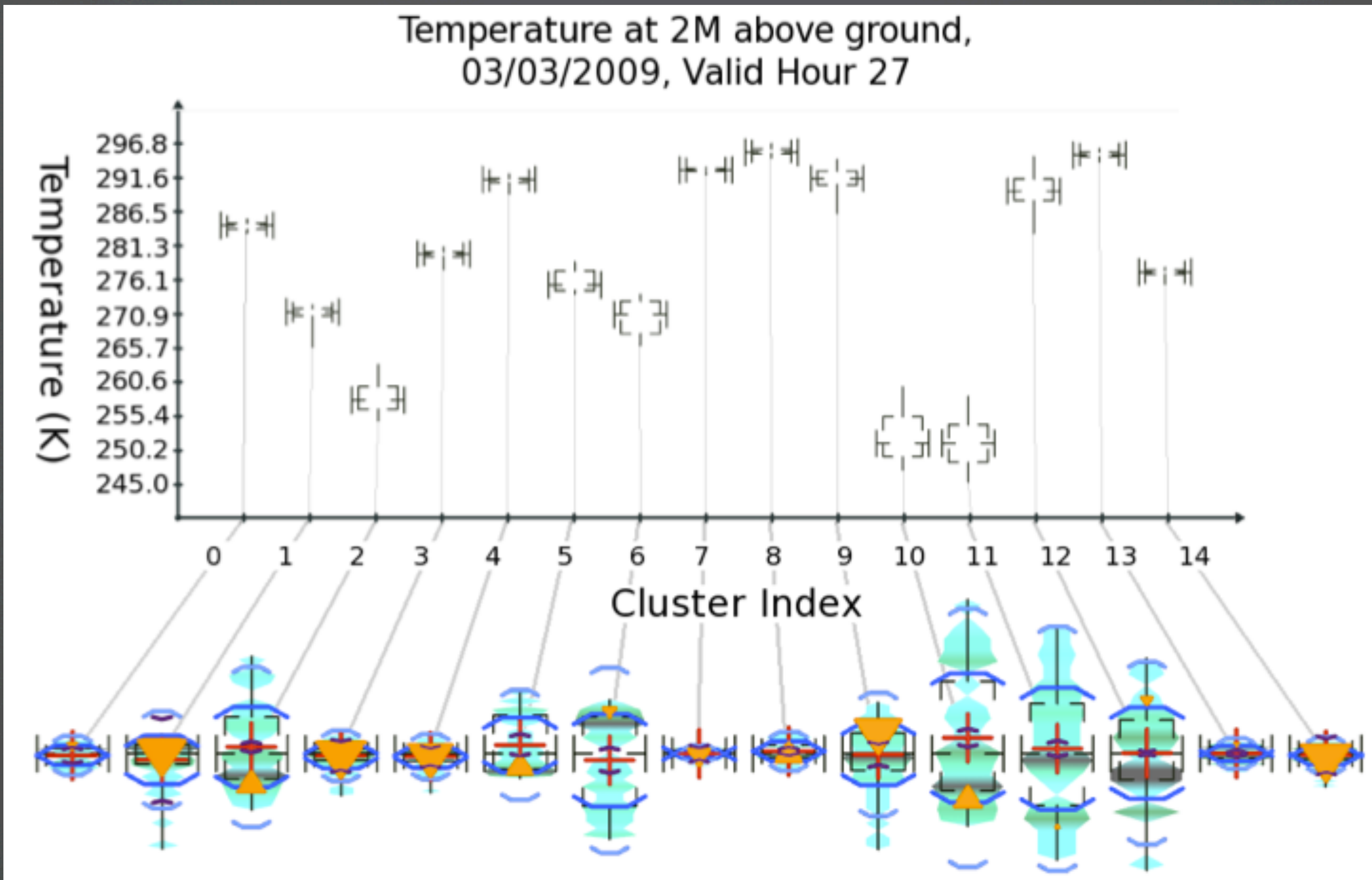
Mean



Standard Deviation



Summary Plots on Clusters



In Summary

- Explore 1D data distributions
- Summarize data
- Highlight salient features
- Redundantly encode information

In Progress: R Package (June 2010)

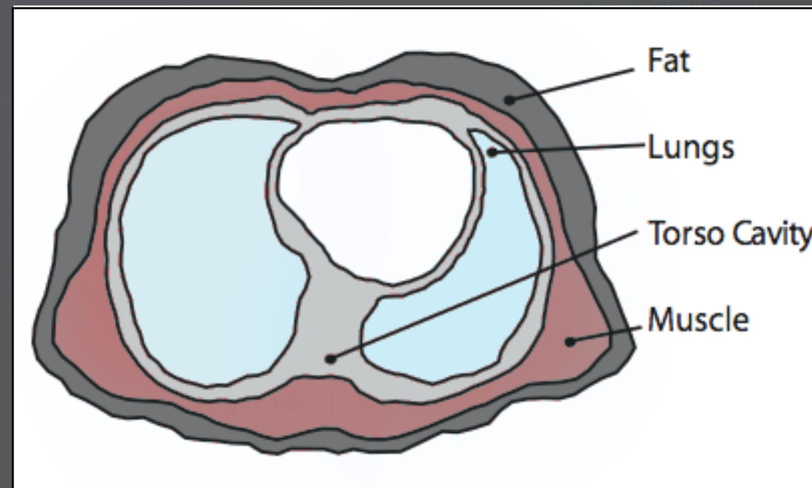
K. Potter, J. Kniss, R. Riesenfeld and C.R. Johnson.
Visualization of Summary Statistics and Uncertainty.
EuroVis 2010 (to appear).

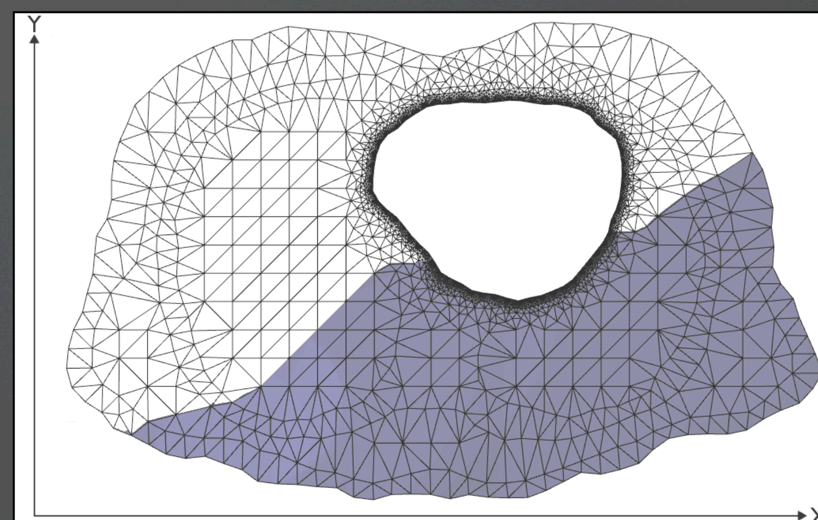
What about 2D data?

- Spatial grid of data
- 1D distribution at each grid pt
- Spatial location important
- Uncertainty across entire spatial domain, not just a single position
- Impact of input parameters on outcome

Electrical Conductivity of the Heart

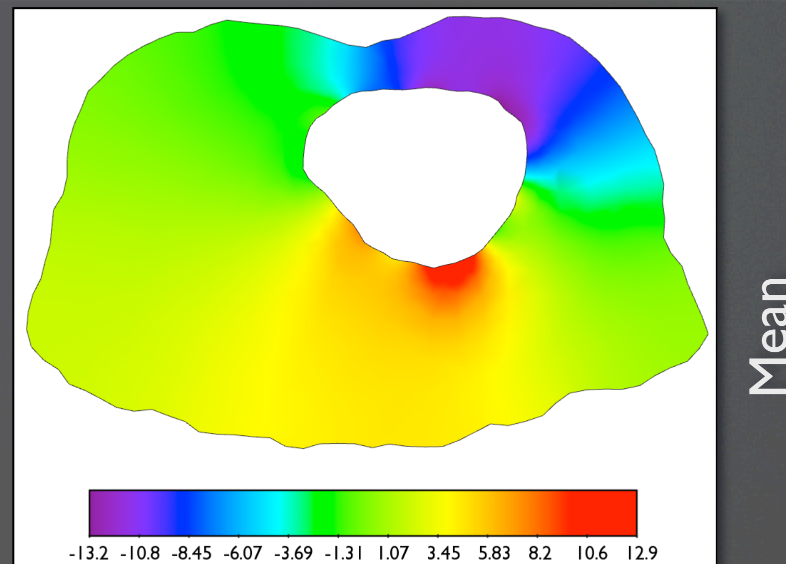
- Electrocardiogram
- Simulate how signals from the heart propagate across the torso
- Vary tissue conductivity rather than have a single value for each tissue

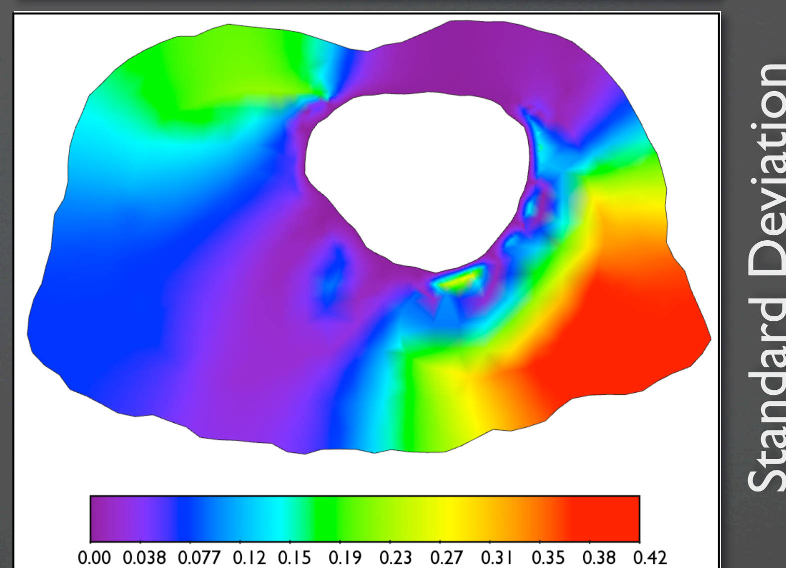




Potentials Data

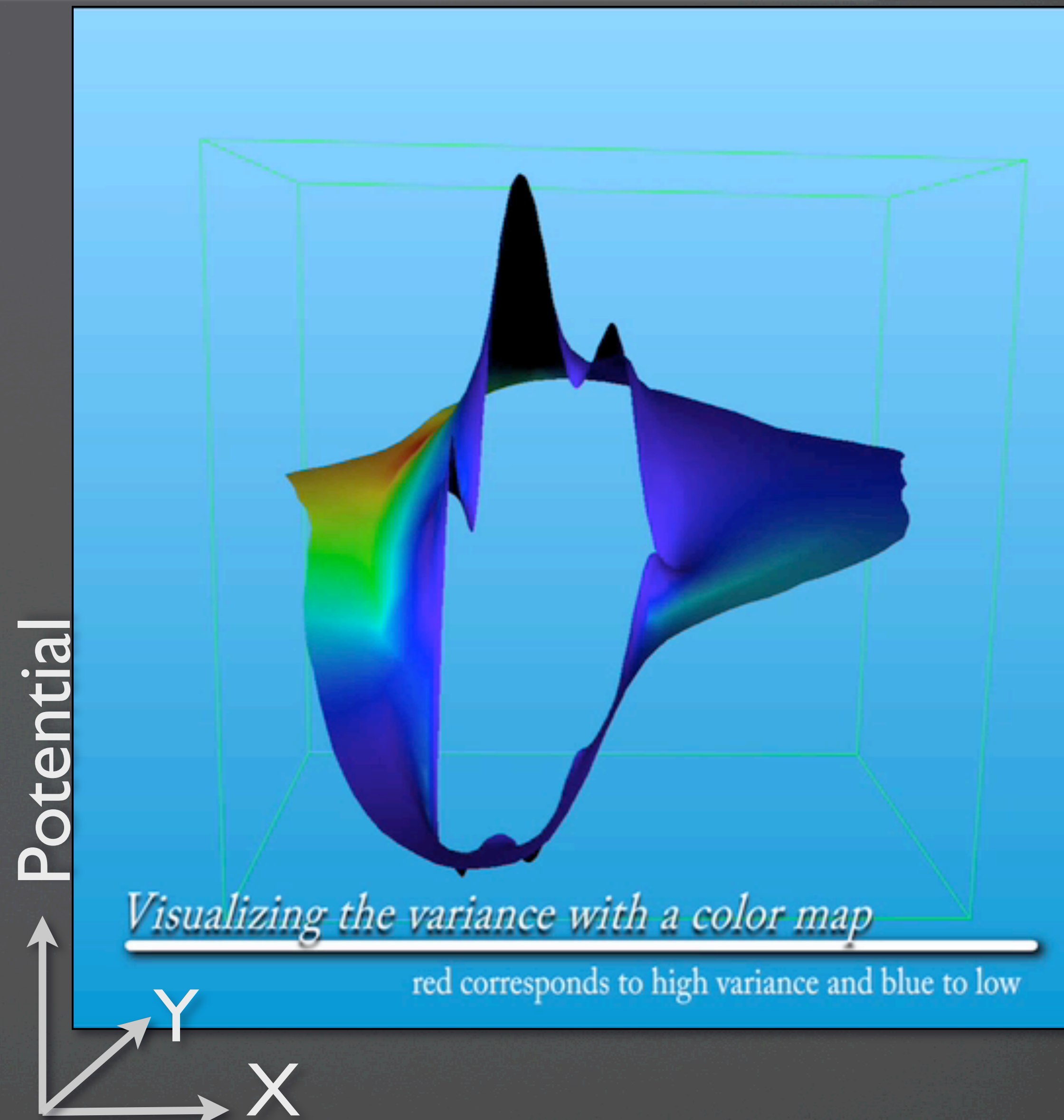
- Study the impact of variation on input conductivity
- Vary lung conductivity uniformly +/- 50% from the reference
- 10,000 realizations, estimate a PDF





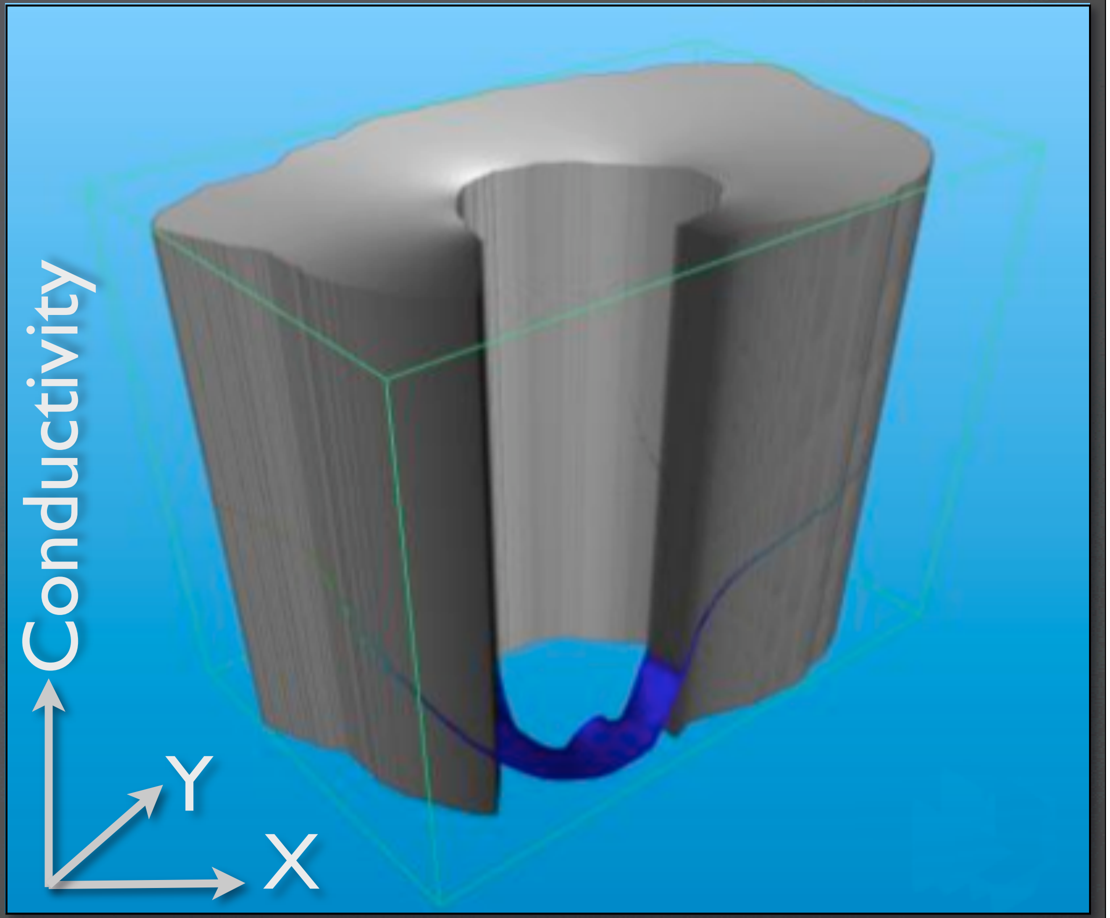
Displacement Mapping

- Using the Z axis
- Height encodes mean
- Color encodes standard deviation
- Really want to see sensitivity of heart potentials to variations in lung conductivity



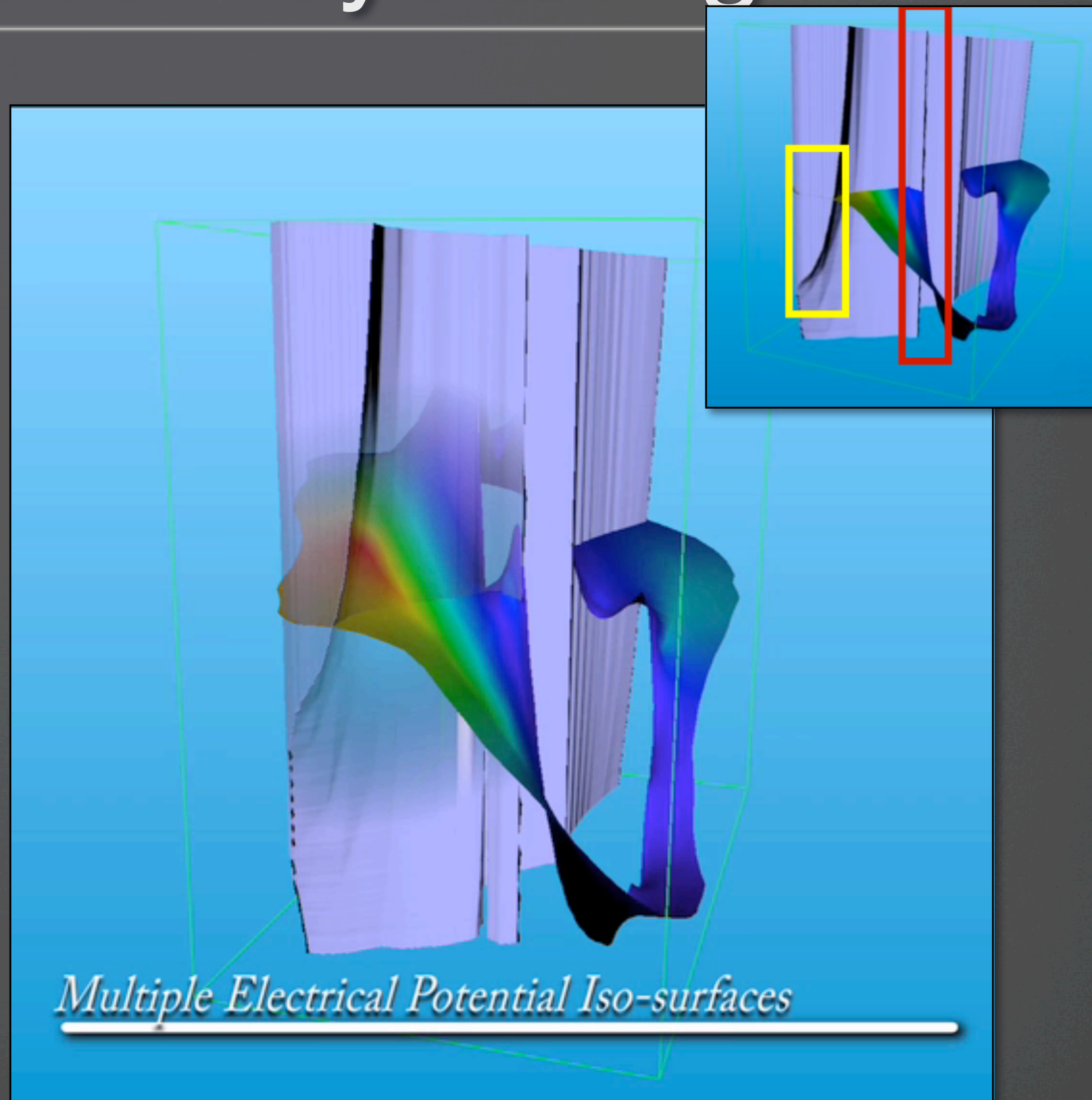
Direct Volume Rendering

- Change z-axis to represent input conductivity (low to high)
- stack realizations/slices (subsample to 512 for texture limitations)
- 2D slices occlude each other
- not clear how to use transparency
- overall dvr not effective



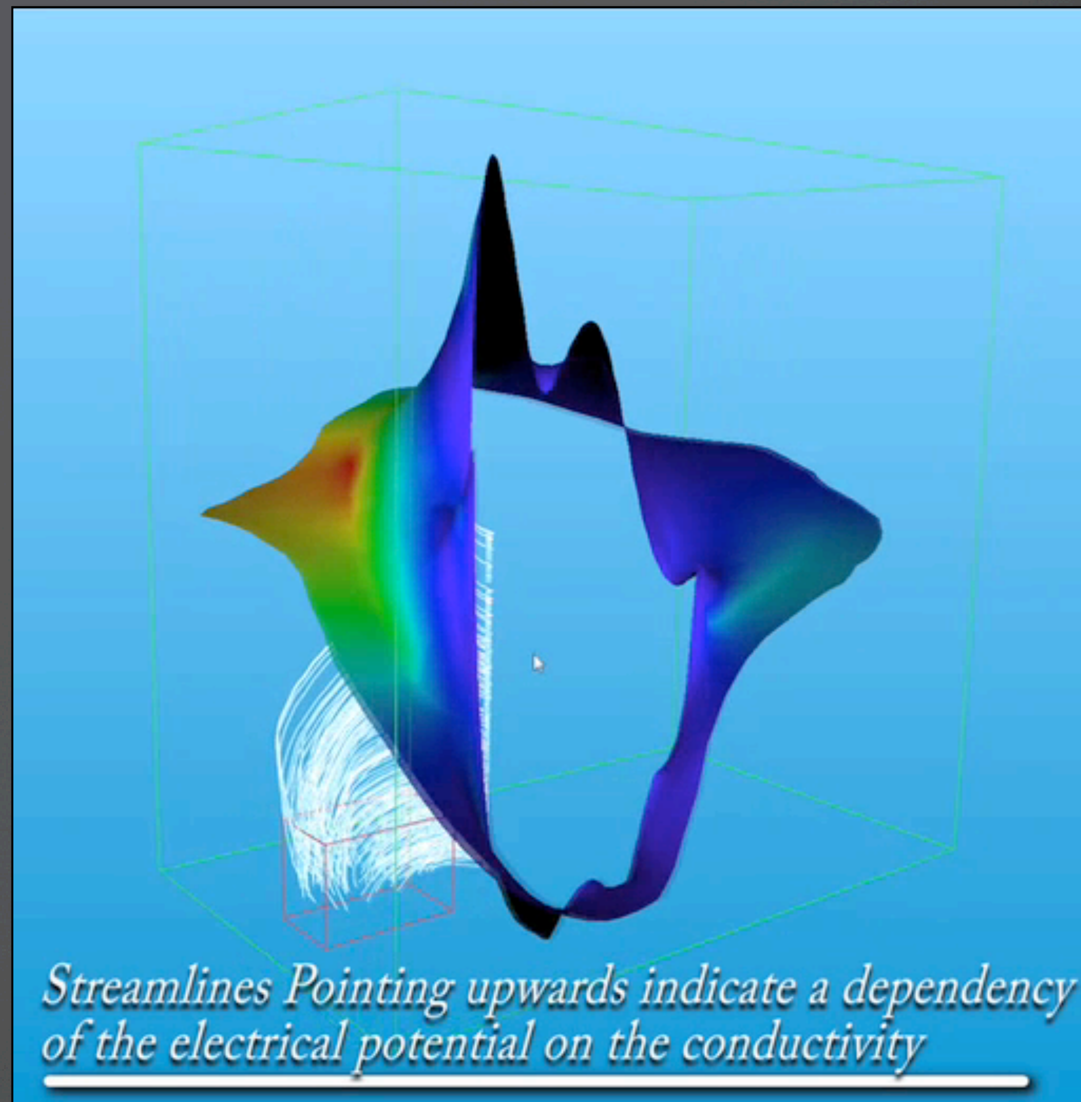
Isosurface Raycasting

- Isovalues of input conductivities
- Structure of isosurface more important than value
- Curves in isosurface indicate dependence on input
- Multiple isosurfaces



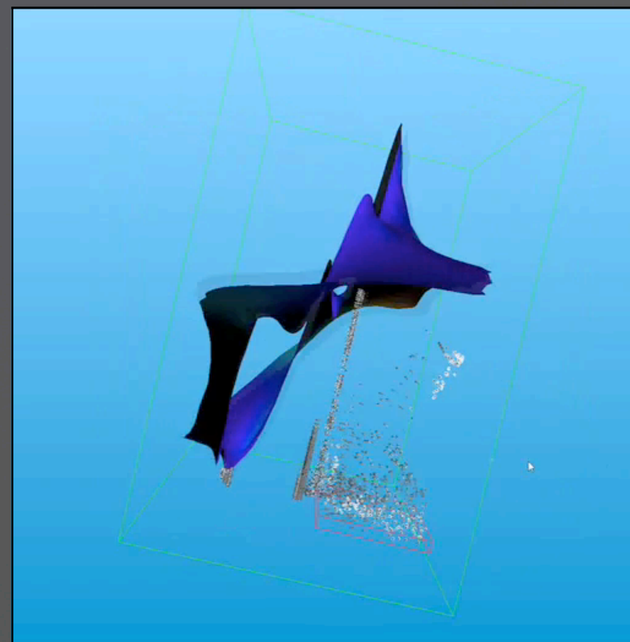
Streamlines

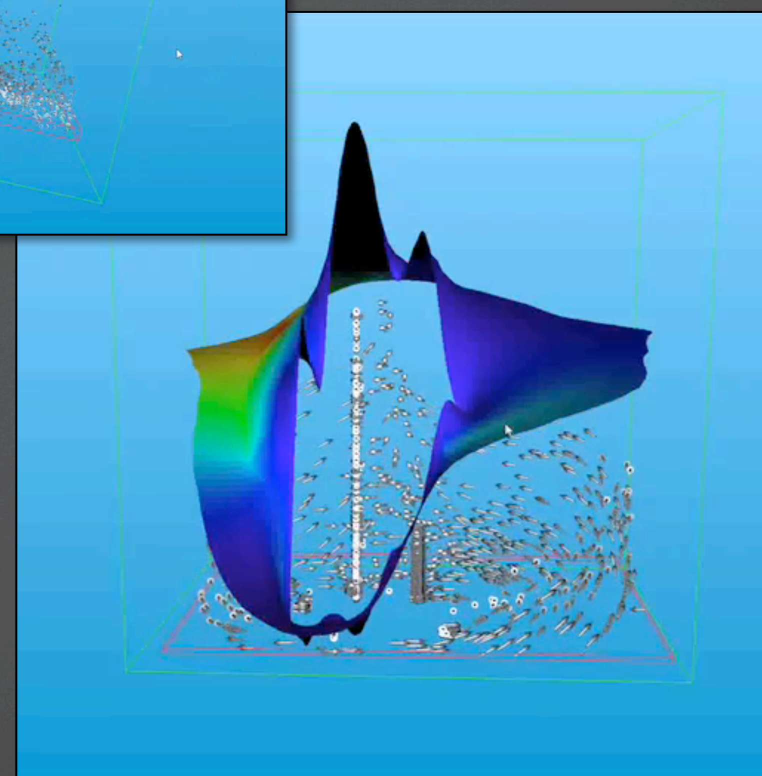
- Gradient field of the output potentials
- Further investigate the changes in potentials
- Streamlines follow the change in potential
- Horizontal streams show independence
- Length indicates strength of change



Particle Tracing

- Seeded particles follow gradient
- Similar to streamlines
- Faster speed indicates greater dependence
- Arrow glyphs better for images, 2D presentation





In Summary

- Visualization techniques for the exploration of the relationship between input and output
- Novel visualization method for multi-dimentional distribution data
- Global qualitative and local quantitative

K. Potter, J. Krüger, and C.R. Johnson.
Towards the Visualization of Multi-Dimensional Stochastic Distribution Data.
In IADIS International Conference on Computer Graphics and Visualization, pp. 191-196, 2008.

The Ensemble-Vis Framework

- User-driven, component-based framework
- Combine various visualization paradigms
- Explore the range of possible predictions
- Show probability of outcomes
- Interrogate the ensemble

What is Ensemble Data?

Collection of data sets (*members*) generated by computational simulations.

- Multidimensional
  - 2D or 3D spatial domain plus
  - time component
- Multivariate
  - simulations predict for numerous variables (i.e. temperature, humidity, etc)
- Multivalued
  - several values for each variable at each point

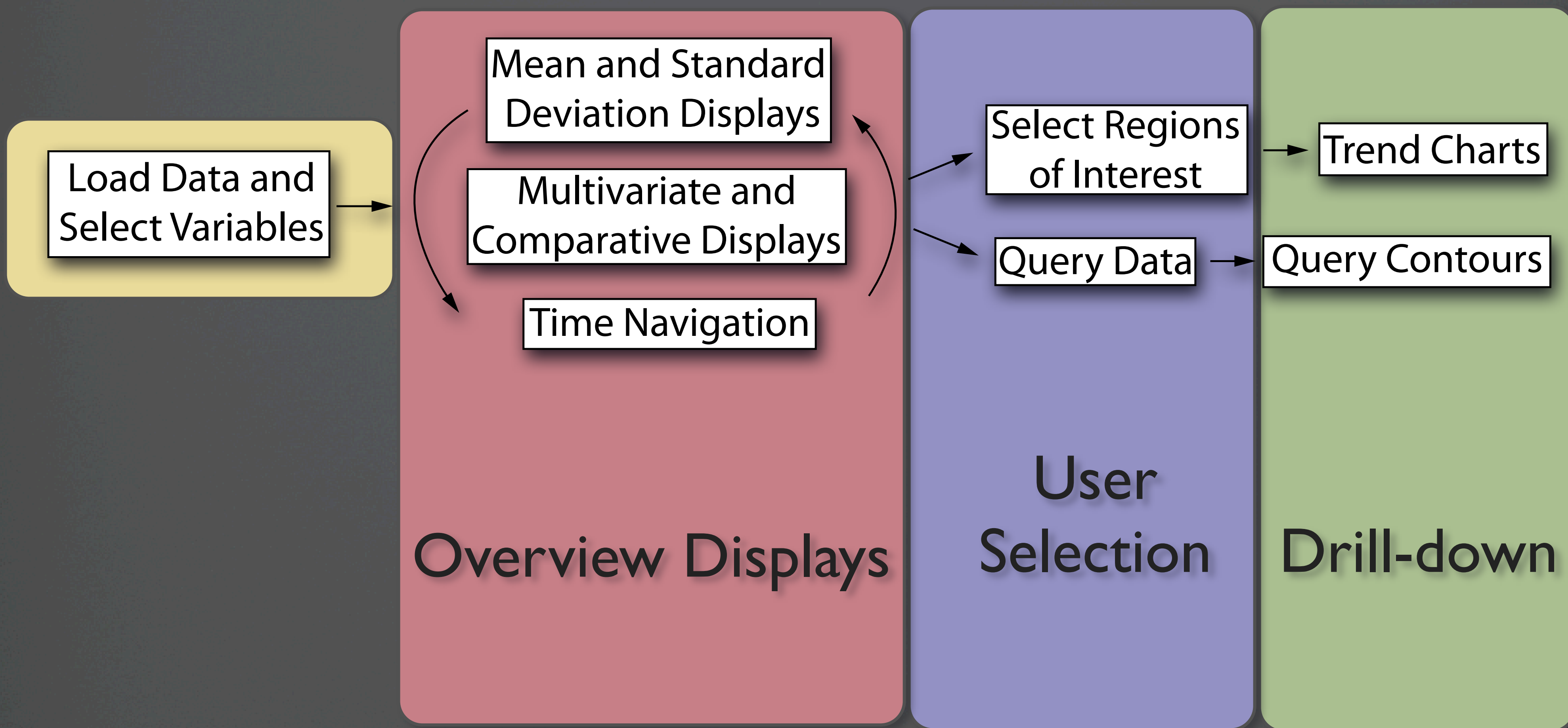
IPCC
Intergovernmental Panel On Climate Change

Climate Modeling

- IPCC Climate of the 20th century
- Spatial domain the whole globe
- Evolution over hundreds of years
- Impact of human activity, trends in natural disasters

Climate of the 20th century experiment (20c3m).
https://esg.llnl.gov:8443/index.jsp.

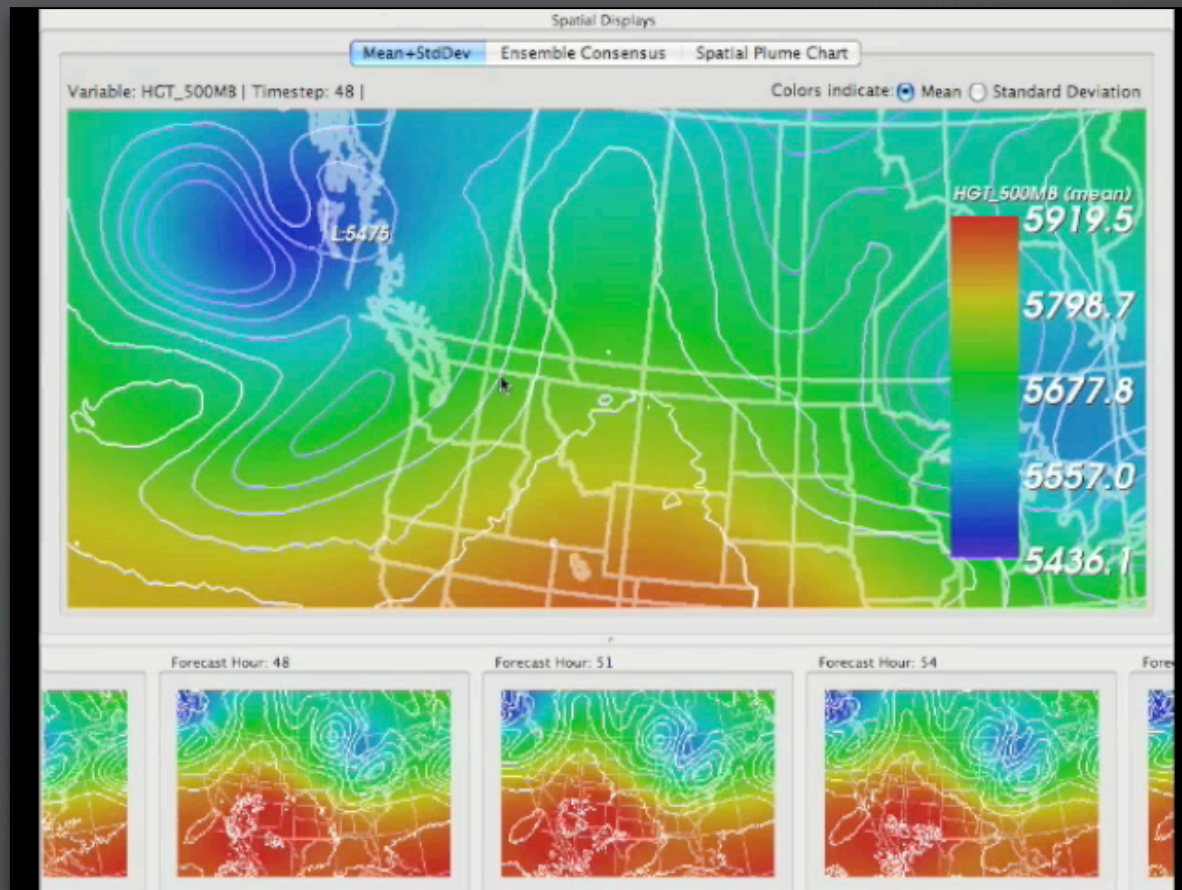
Ensemble-Vis Workflow



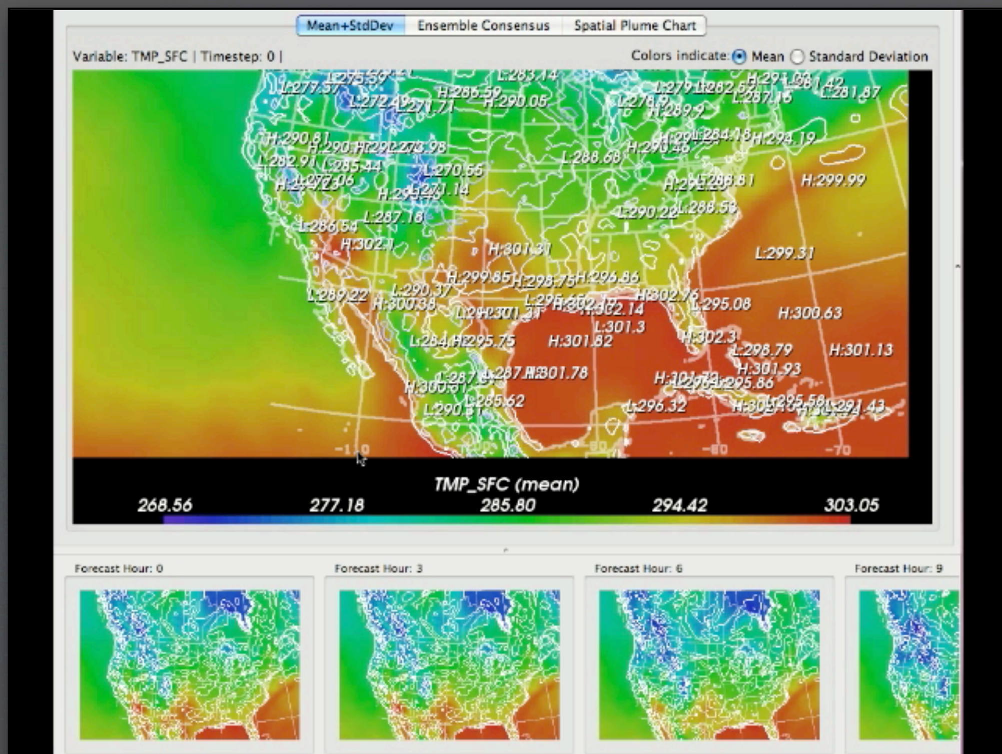
Ensemble Overviews

Mean & standard deviation

- roughly indicate value
- highlight areas of variation
- single time step across spatial domain



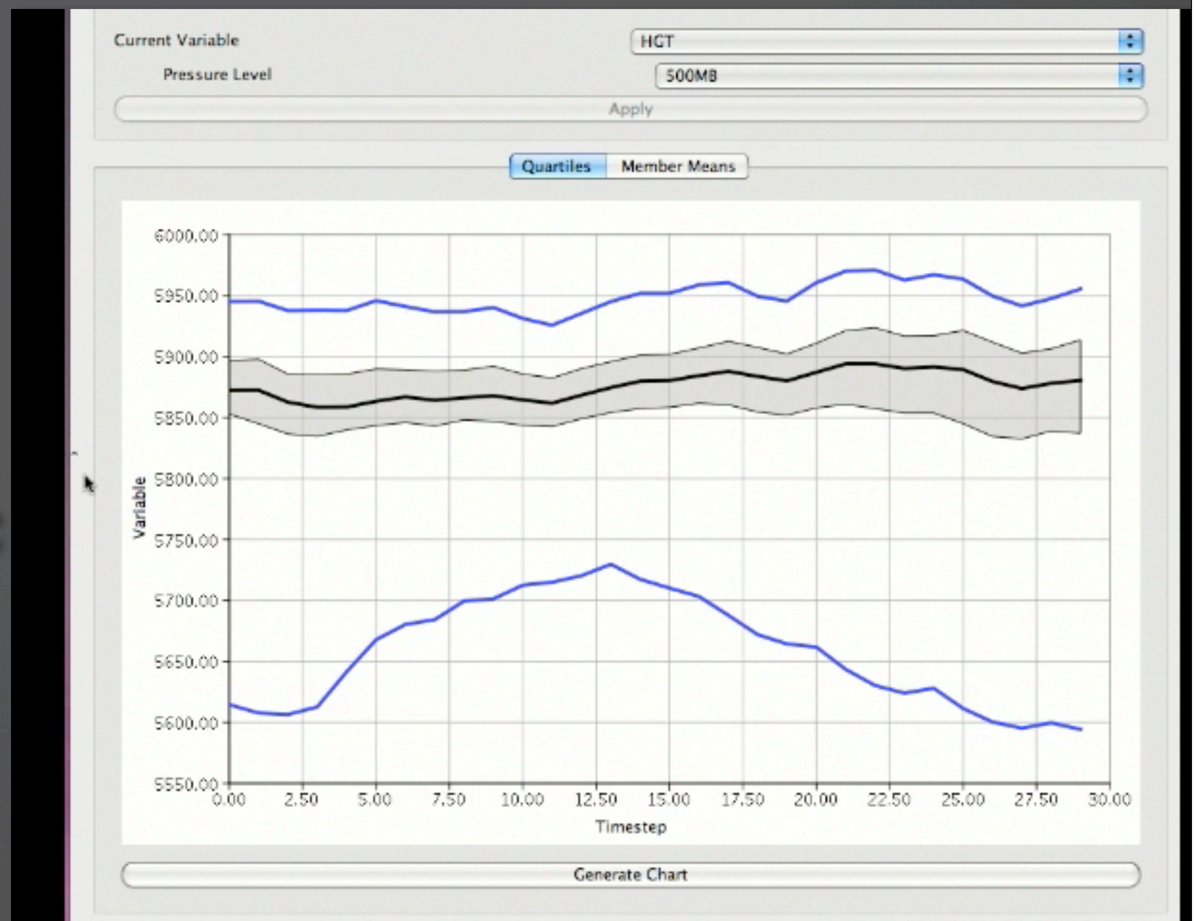
Time Navigation Overviews



- Small multiples showing each time step across
- Quickly see evolution across time

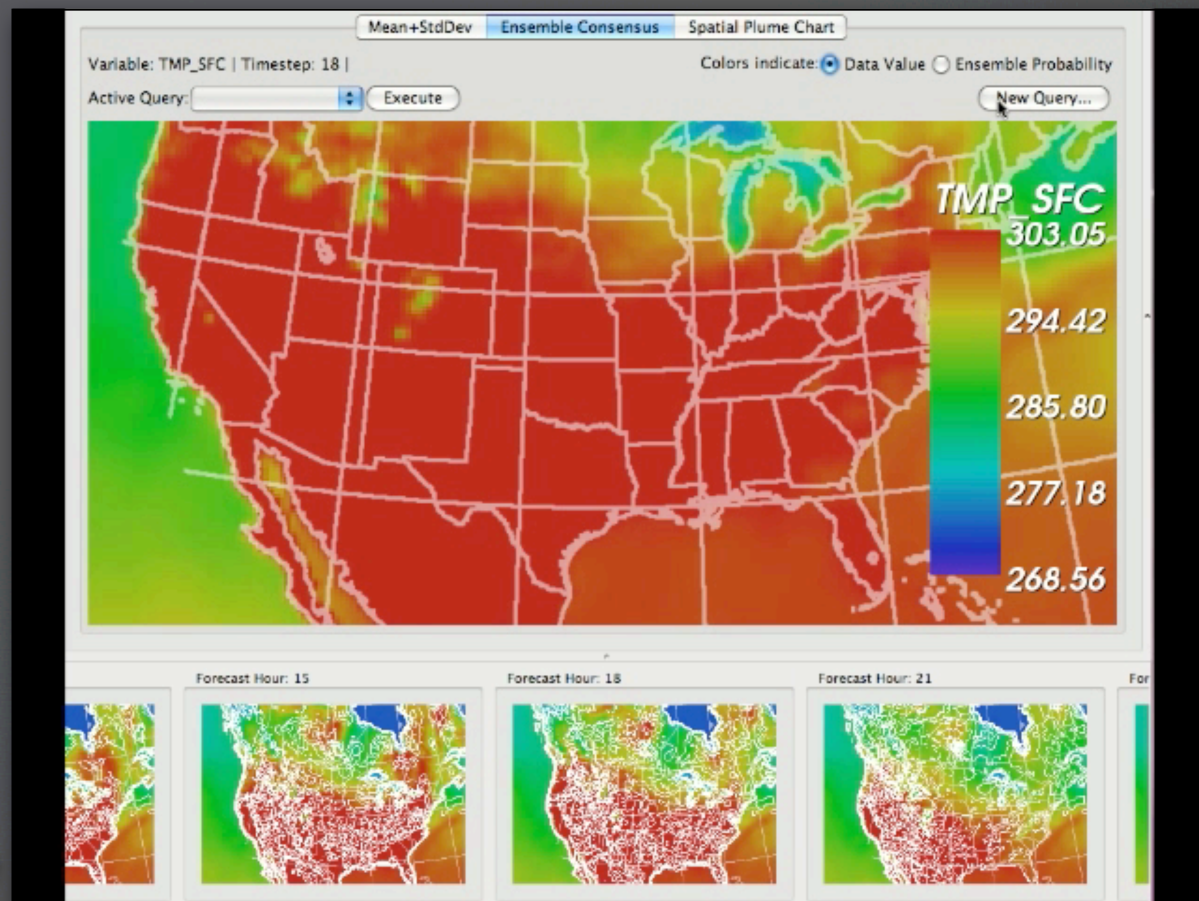
Trend Charts

- Select region of interest
- Show statistics like mean, quartiles, etc
- Drill-down to direct data display



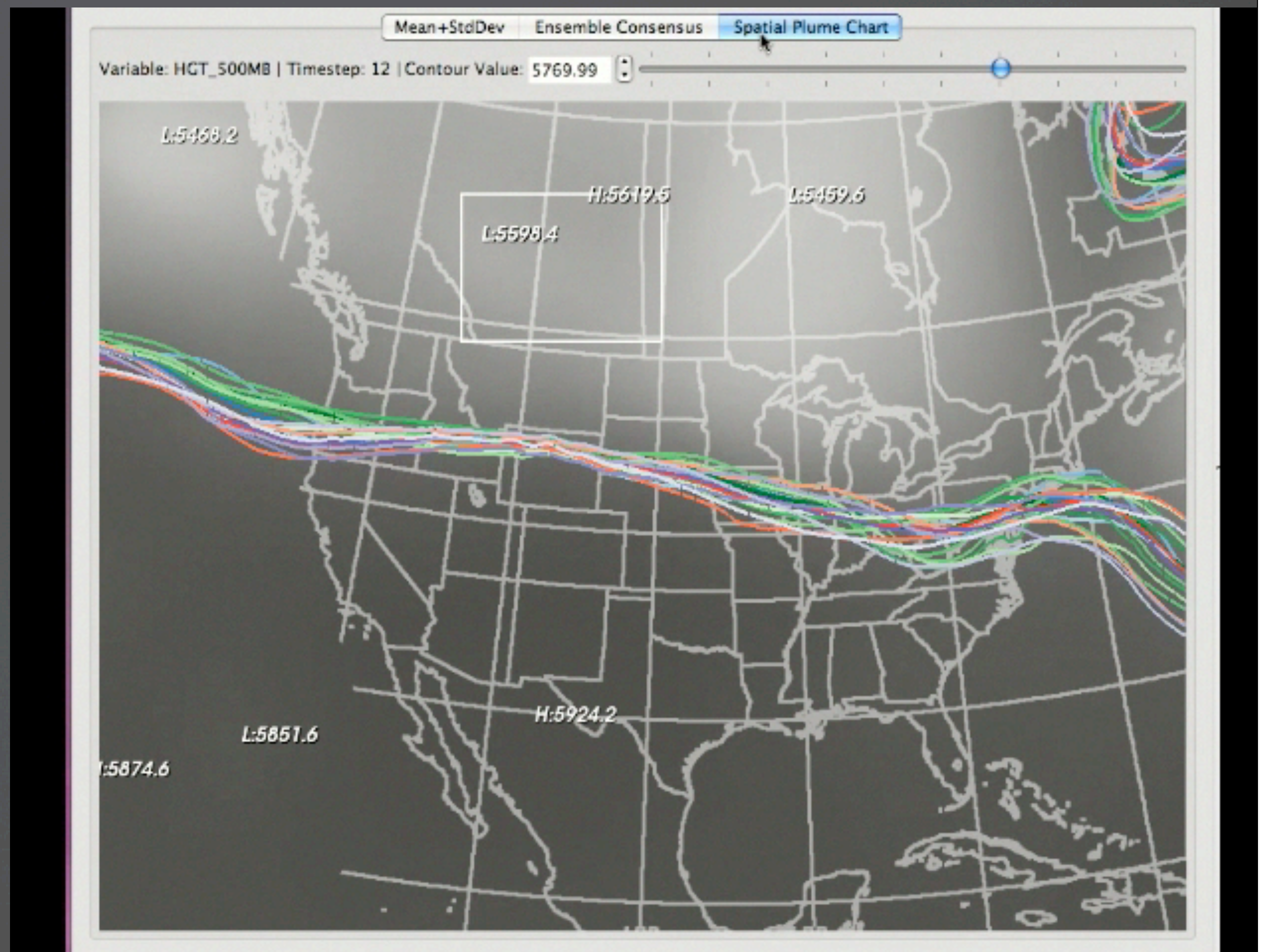
Query Contours

- User-driven query
- Select subset of data
- List of points where conditions are satisfied
- Scalar value at each point indicates number or percentage of satisfying members



Spaghetti Plots

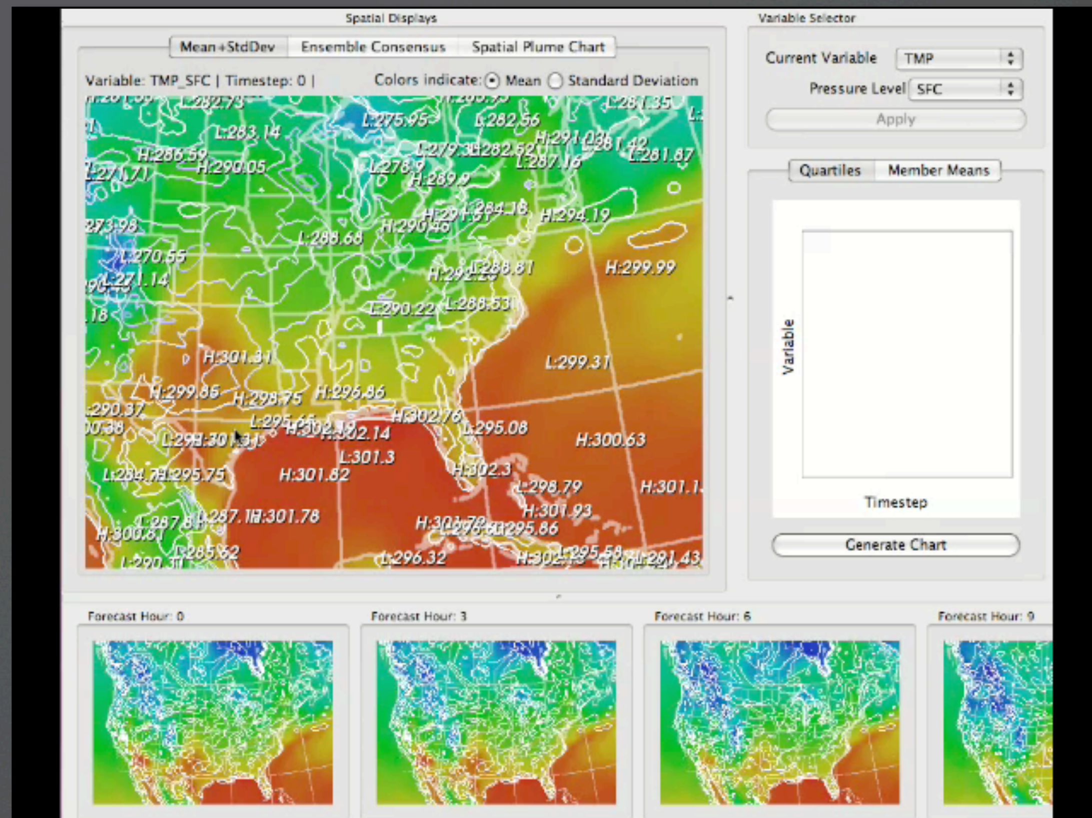
- Show variation across ensemble over space
- User selected contour value
- Isocontour for each member
- Highlights outliers and divergence



Implementation

Two Prototypical Systems:

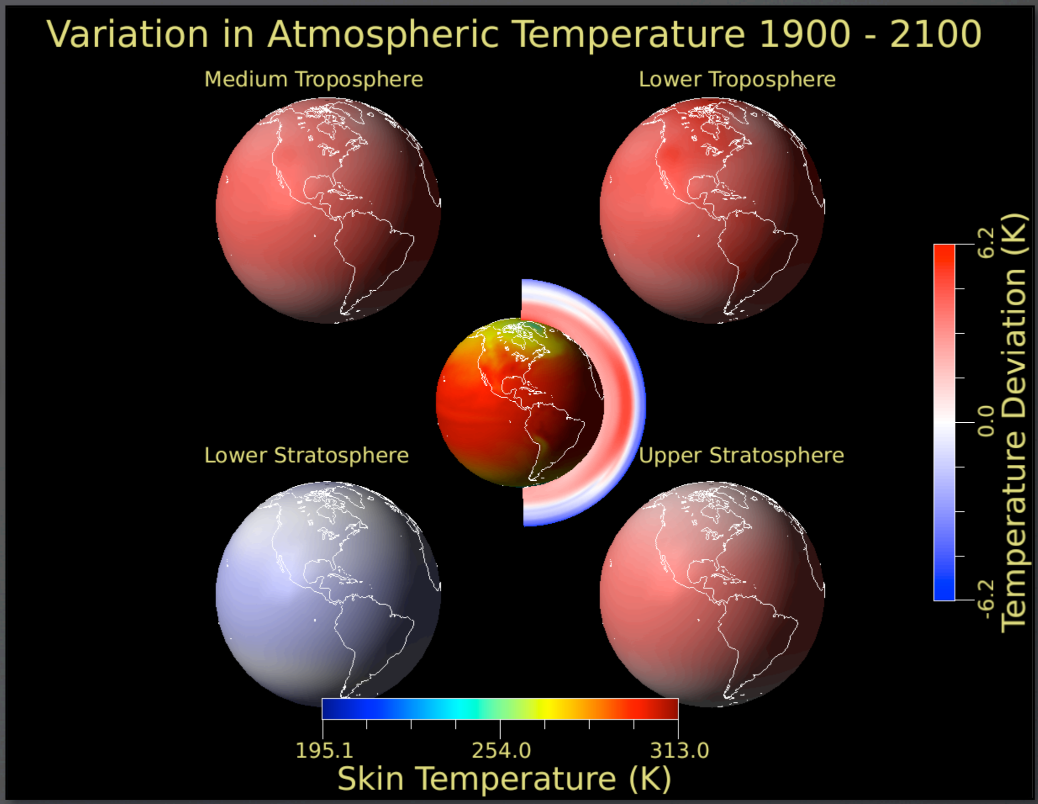
- SREF Weather Explorer
  - VTK filters, Qt widgets
  - Relational database backend MySQL & parallel Netezza



Implementation

Two Prototypical Systems:

- ViSUS
  - Climate Data Analysis Tools (CDAT) integration
  - C++, OpenGL, Python, FLTK
  - Out-of-core streaming



Climate data analysis tools.
http://www2-pcmdi.llnl.gov/cdat

In Summary

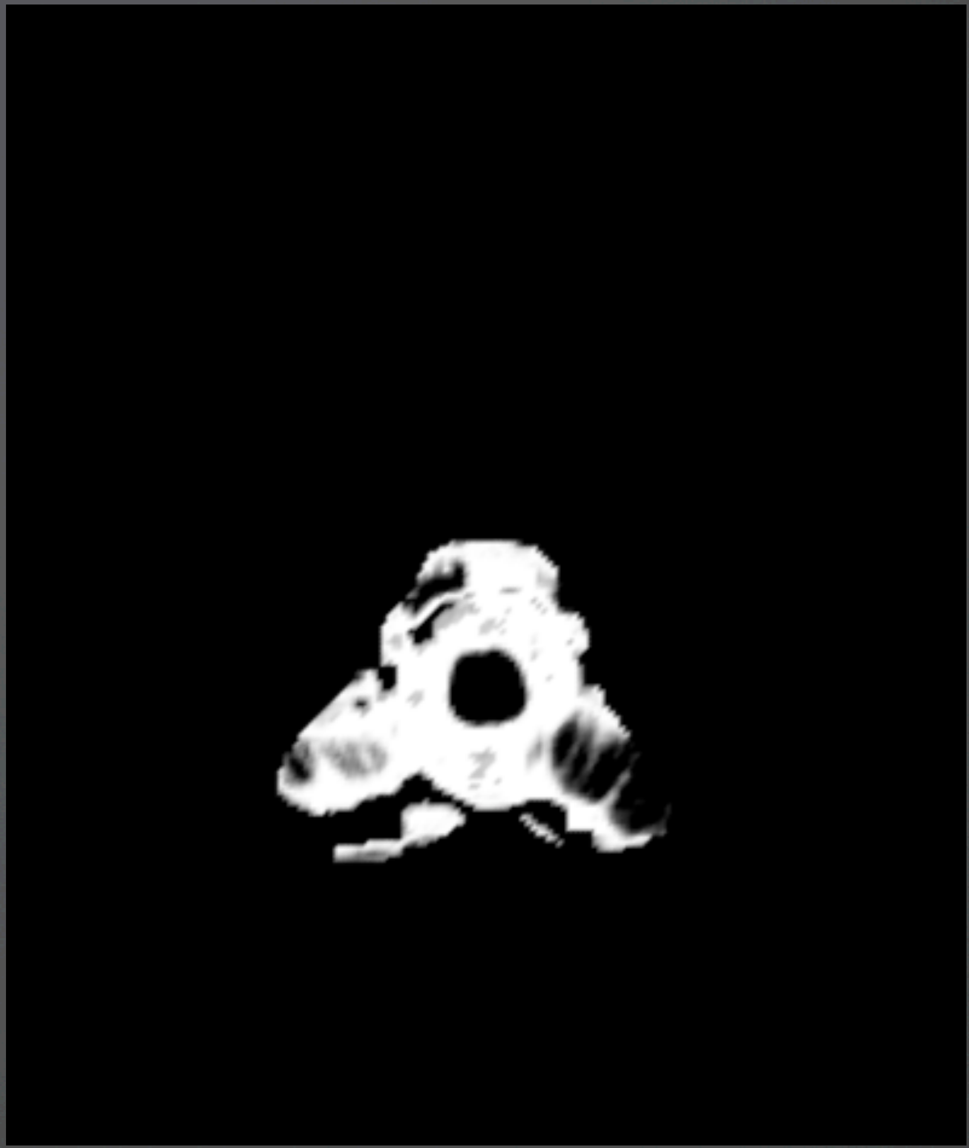
- Framework to let users drive visualization
- Combine various representations to highlight different aspects of the data
- General approach can be applied to numerous other fields

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, and Impacts, Dec 2009*.

Current Work

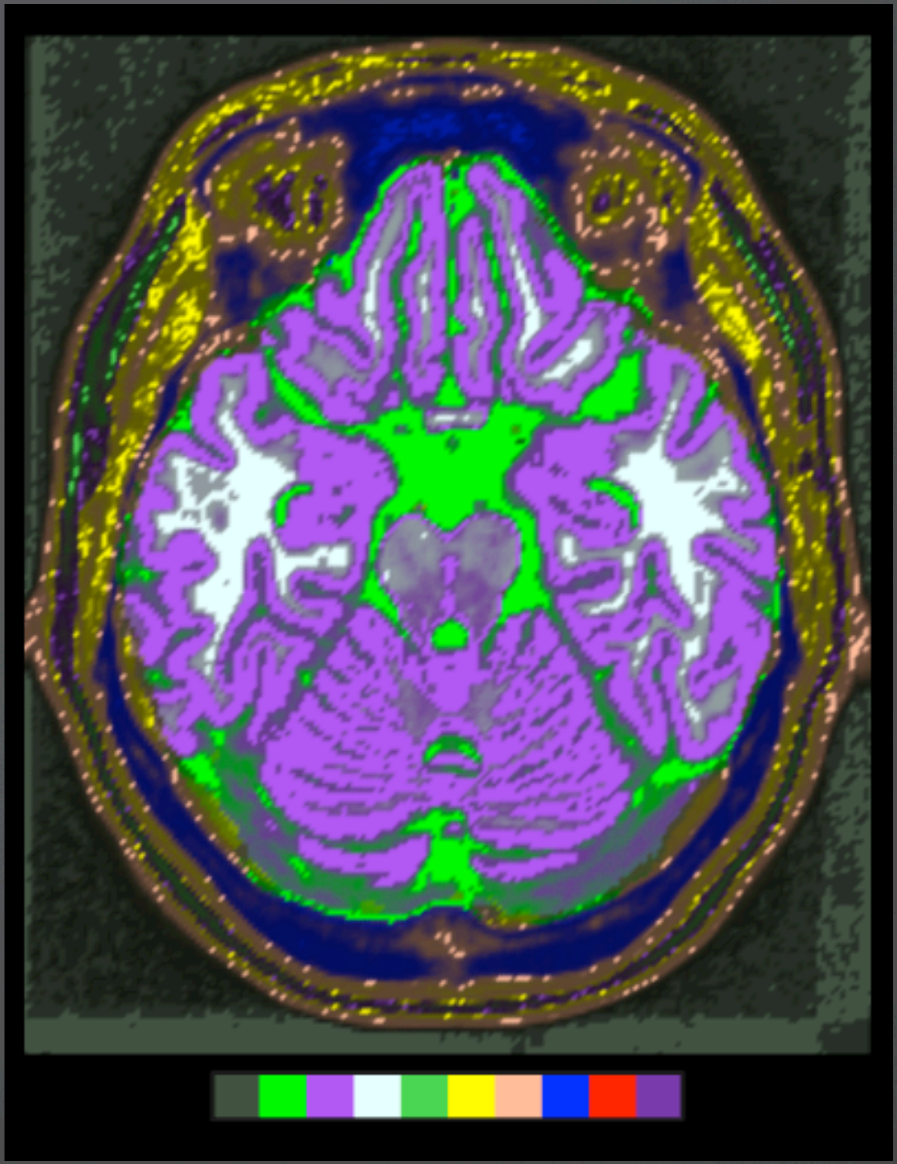
- BrainWeb Data *
  (freely available)
- Classification of data into 10 distinct types (9 tissue + background)
- 10 volumes, 0-255 probability of each tissue type
  (0-1 in colorspace)

http://mouldy.bic.mni.mcgill.ca/brainweb



Current Work

- Colormapping
  - distinct color for each tissue type
  - opacity also encodes probability
- Not too bad, but...
  - hard to tell difference in blended color types
  - limited to 2D slices



Future Work

- 3D+ techniques
- Parameter investigations
- Enhanced user interface & interactivity
- Large datasets may require *In situ* data processing, data reduction, etc

General approaches can only get you so far, eventually specialization is needed

Thank You!

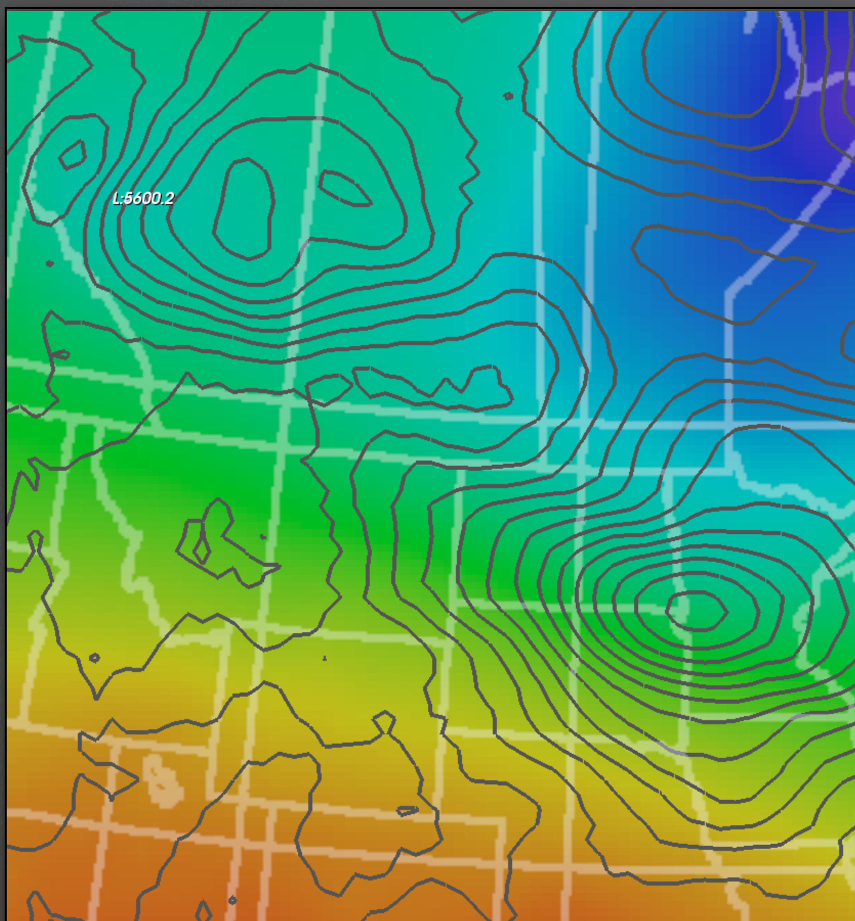
Funding Agencies

- This work was funded in part by the DOE SciDAC Visualization and Analytics Center for Enabling Technologies (www.vacet.org) and the NIH NCRR Center for Integrative Biomedical Computing(www.sci.utah.edu/cibc), NIH NCRR Grant No. 5P41RR012553-02.
- Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy¹s National Nuclear Security Administration under contract DE-AC04-94AL85000.
- This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344

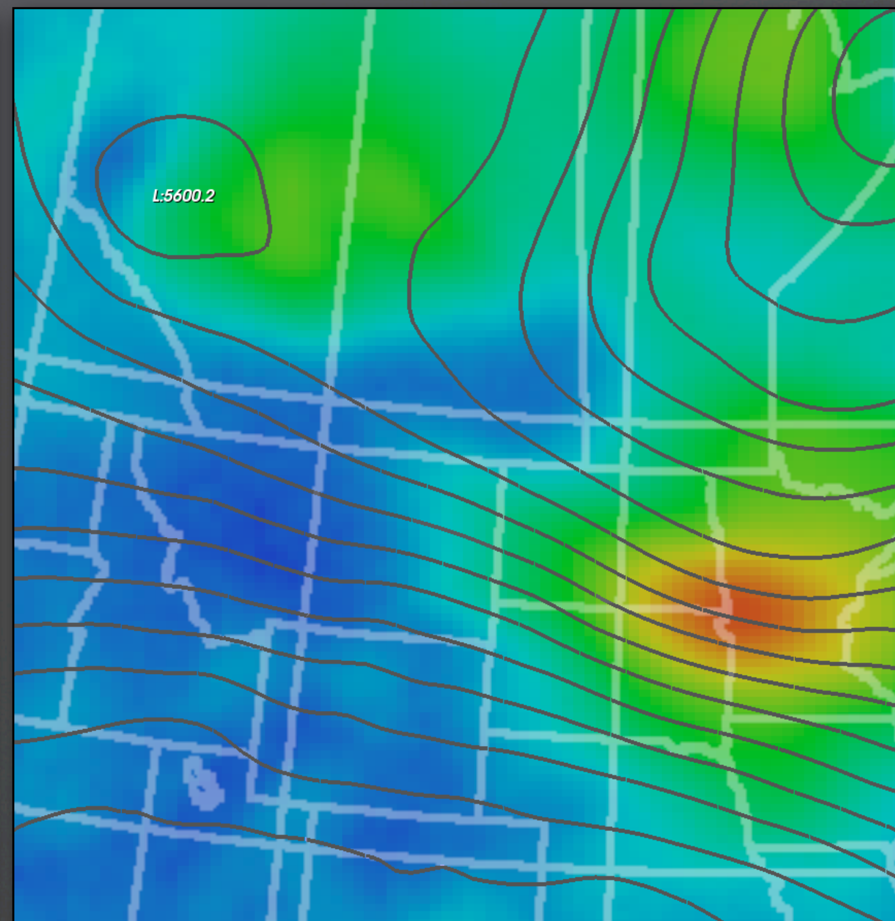
Conclusion

- Uncertainty visualization needs to be a priority
- But it is a hard problem
- Large data increasing the demand
- Presented two novel visualization schemes and a framework for visual data analysis

Contours and Colormapping



Mean = colormap,
Standard Deviation = contours



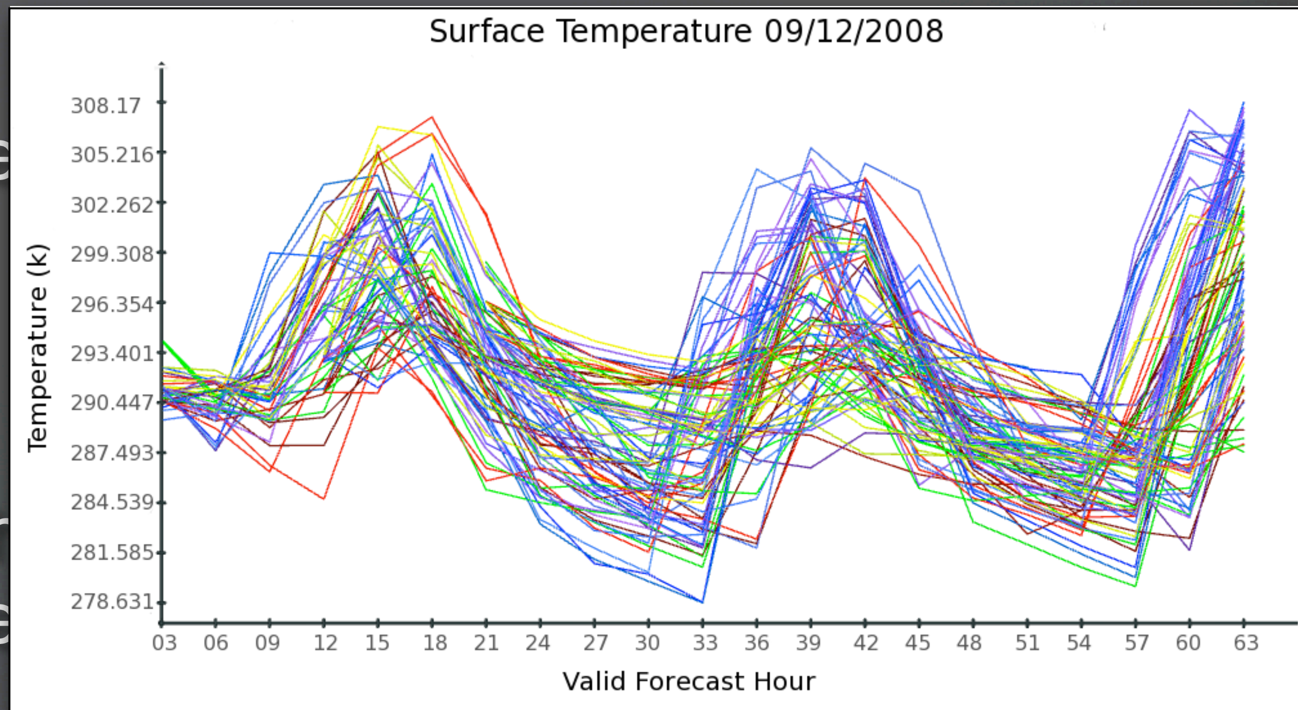
Mean = contours,
Standard Deviation = colormap

Why use ensemble data?

- Simulate complex systems
- Handle unknowns in initial conditions
- Investigate sensitivity to parameters
- Mitigate uncertainty

Ensemble data is

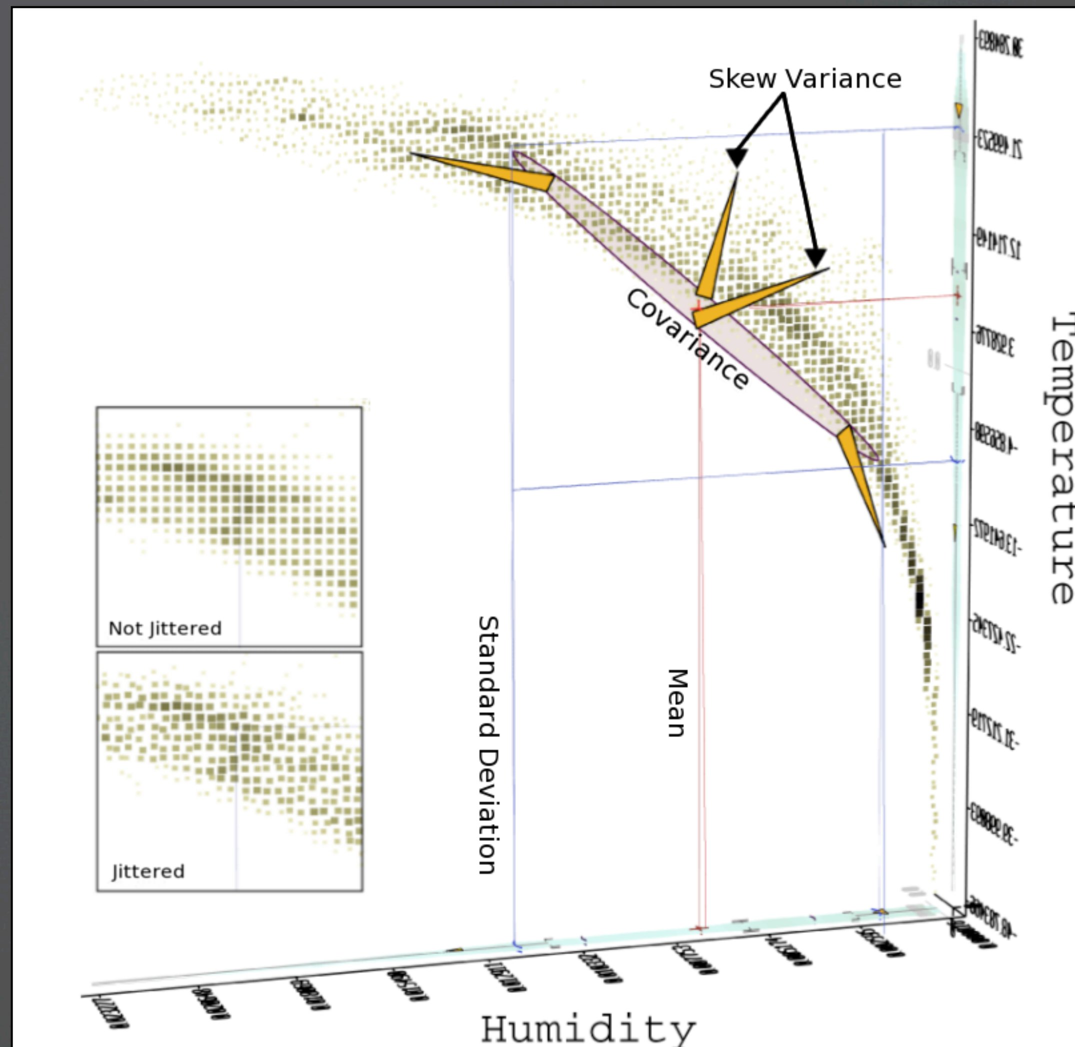
- Information rich
- Can get very large
- Not clear how or what to visualize
- Need an approach that handles these issues



Single weather station, single variable, all runs, across all valid forecast hours

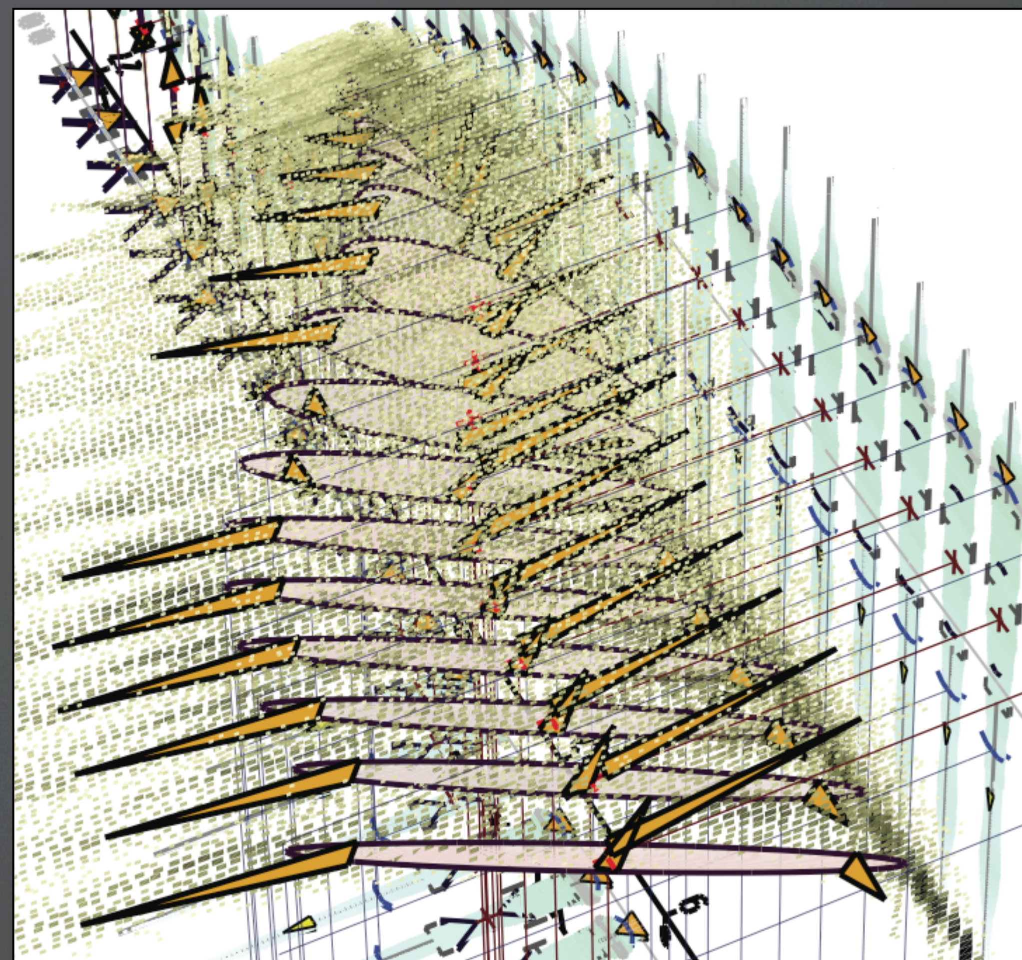
2D Summary Plots

- Joint density
- Covariance
- Skew variance



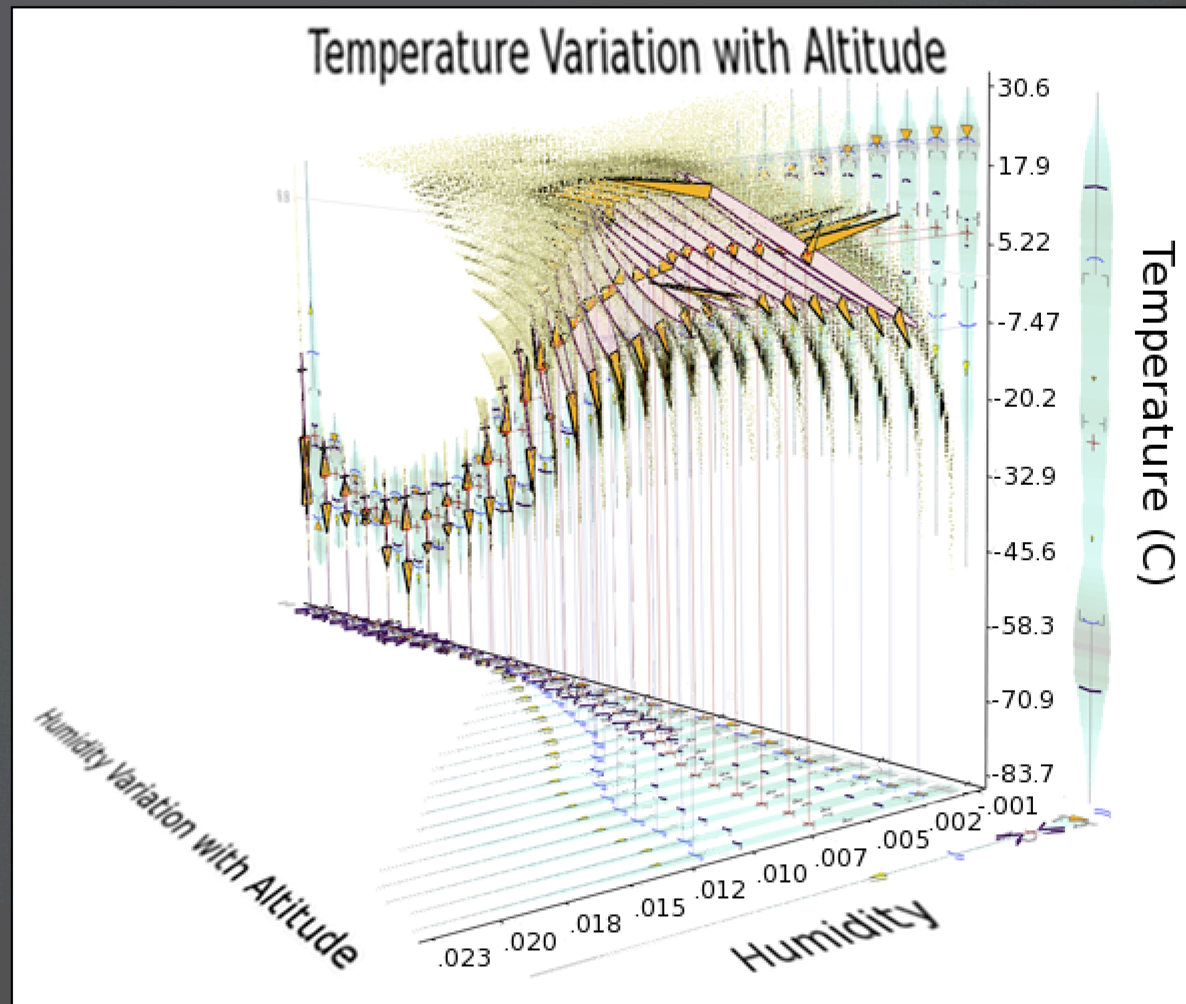
Multiple 2D Plots

- Show trends in data
- Covariance and skew variance distinguish between plots

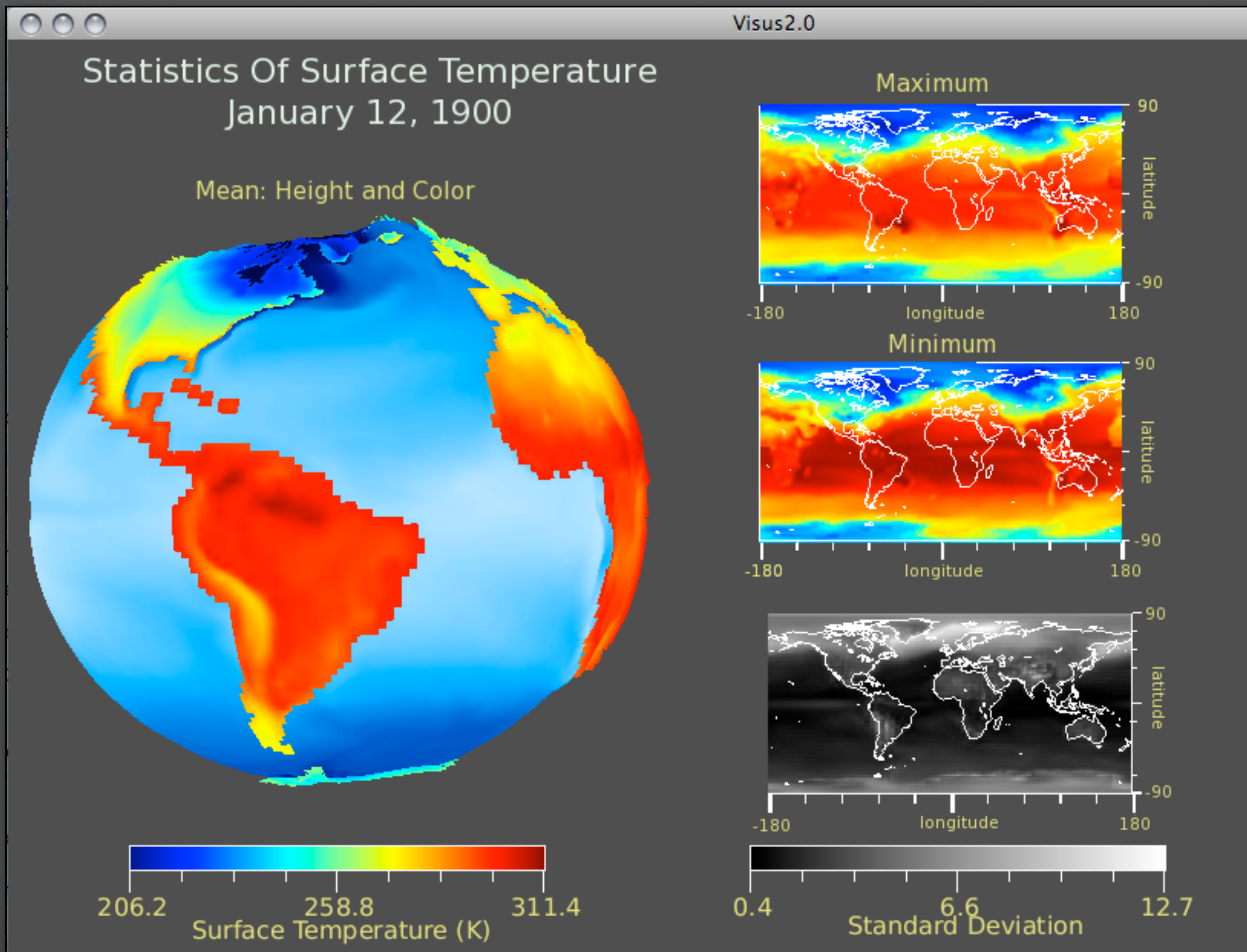


2D Summary Plots

- Average across spatial domain
- Slice for each altitude level
- Highlight trends in both variables

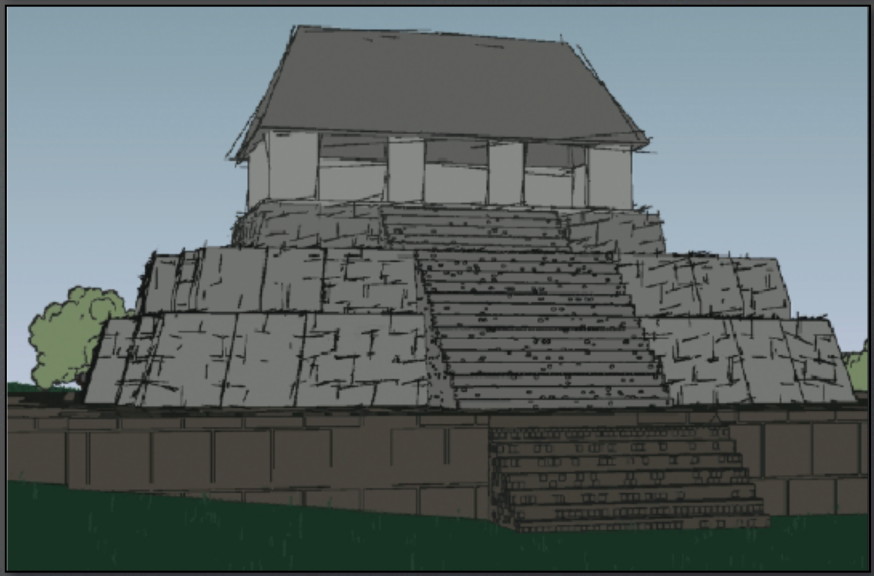


Height and Comparative Displays



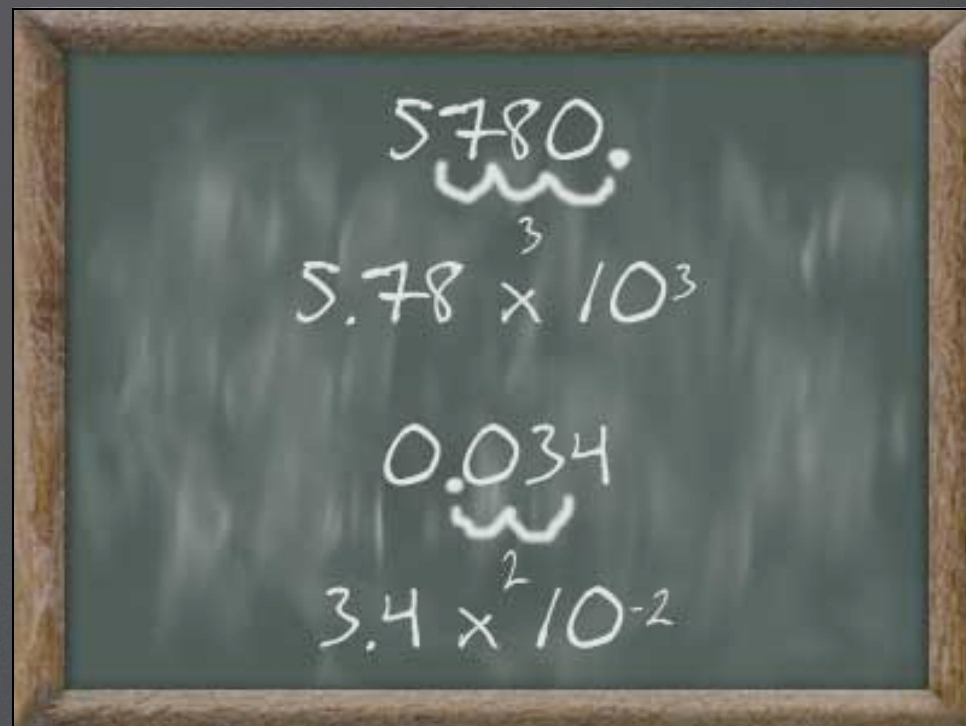
Motivation - 2

- 3D reconstruction of Mayan temple
- Express levels of confidence through rendering style
- The less confident, the more sketchy



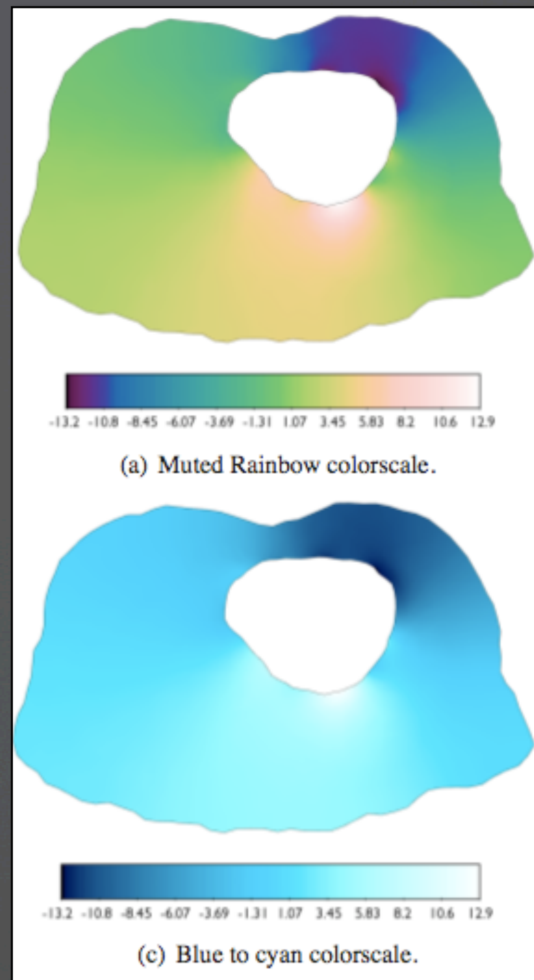
Uncertainty is everywhere

- Talk about how we are familiar with it, not new
- For example sig figures
- Contemporary problem in vis
- Remove insignificant zeros
- Use only the important



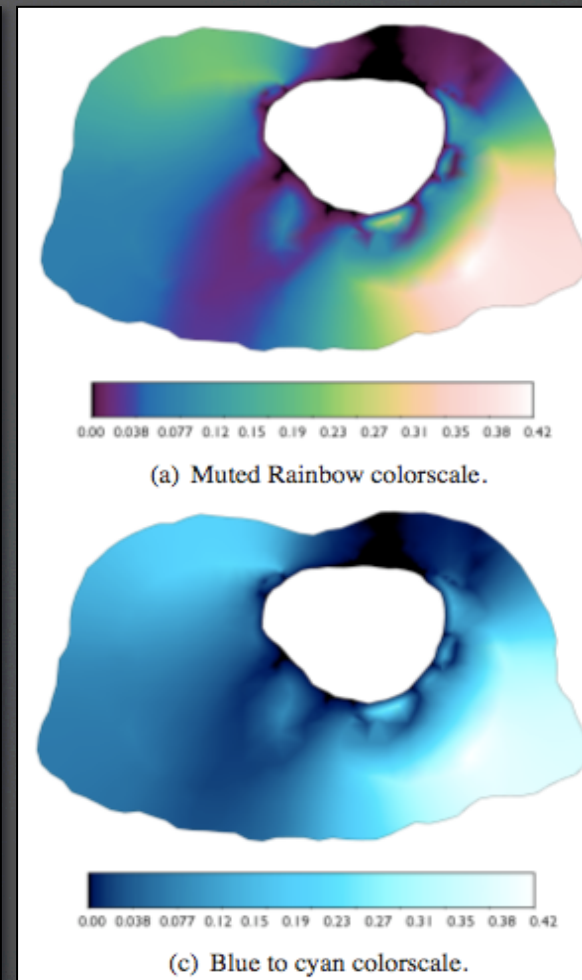
Colormapping

- Typically first step in data presentation
- Choice of colormap influences understanding of variable
- Low to high, rainbow, or two color



(a) Muted Rainbow colorscale.

(c) Blue to cyan colorscale.

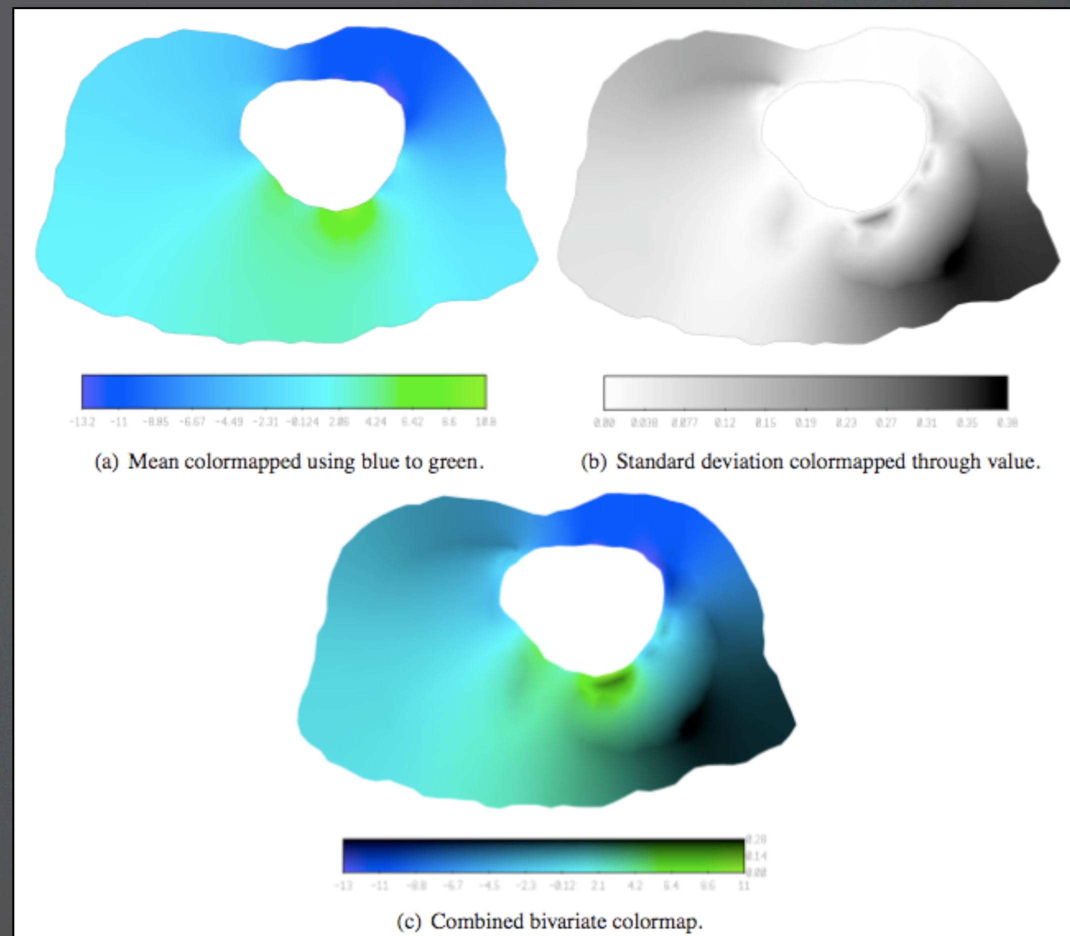


(a) Muted Rainbow colorscale.

(c) Blue to cyan colorscale.

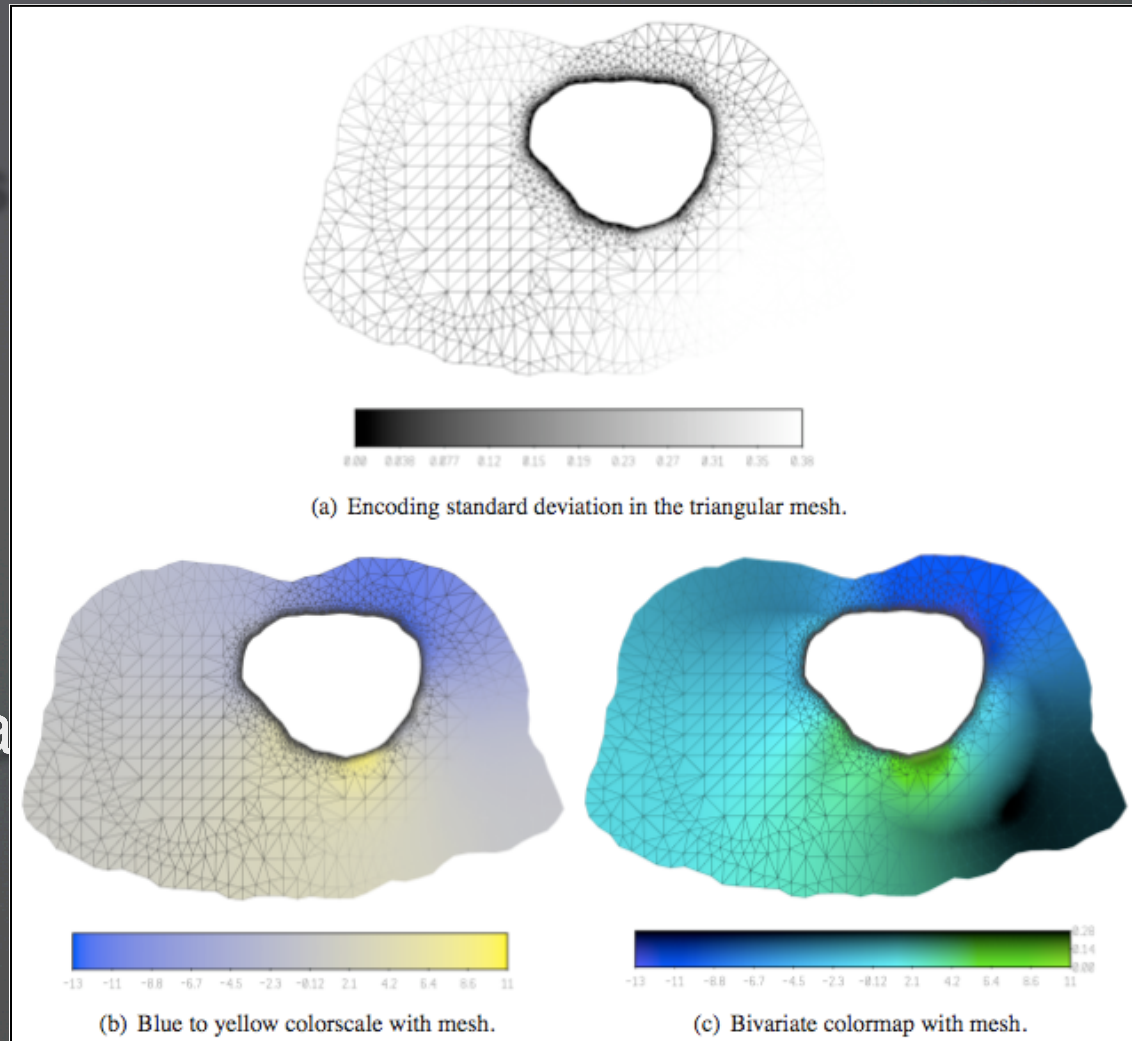
Bivariate Colormaps

- Encode two variables in a single map
- HSV colorscale intuitive
- Should uncertainty be encoded as light or dark?



Annotation

- Encode uncertainty as transparency in the wireframe.
- Redundant using 2D colormap
- Does not obscure data presentation



(a) Encoding standard deviation in the triangular mesh.

(b) Blue to yellow colorscale with mesh.

(c) Bivariate colormap with mesh.