Visualization of Uncertainty

Kristin Potter

PhD Defense

December 10, 2009

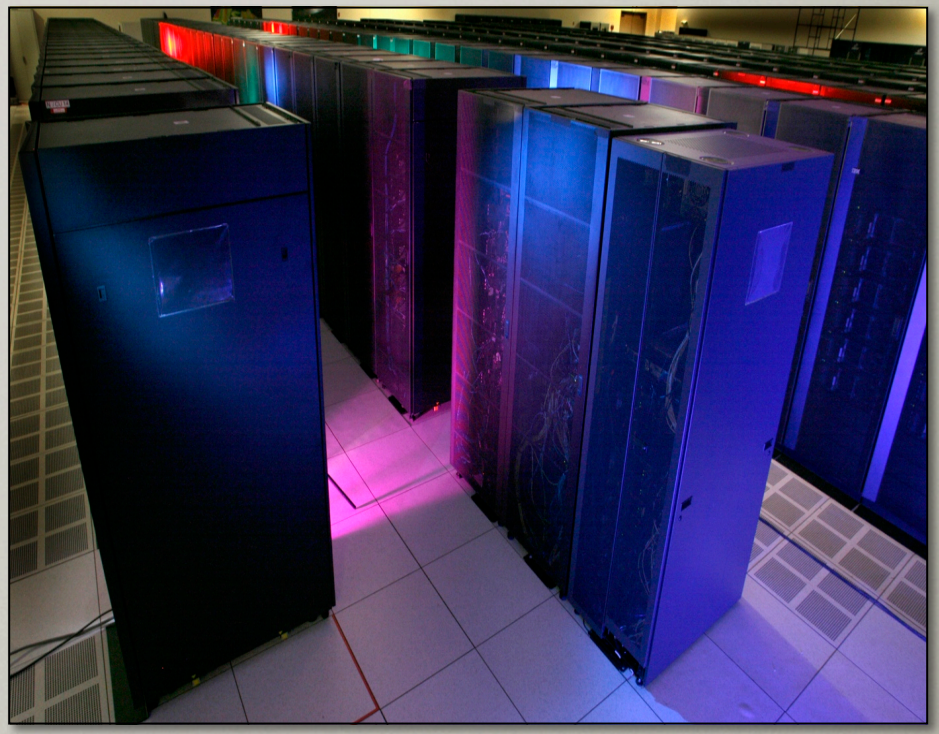
Committee:

Richard Riesenfeld, Chris Johnson,

Mike Kirby, Elaine Cohen, and Bruce Gooch

Advanced Computing and Scientific Data

- More bandwidth, storage, computational power
- Larger data sets:
  - Higher resolutions
  - Run for longer time
- More sophisticated models
- Leading 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
- Typically included as charts and graphs

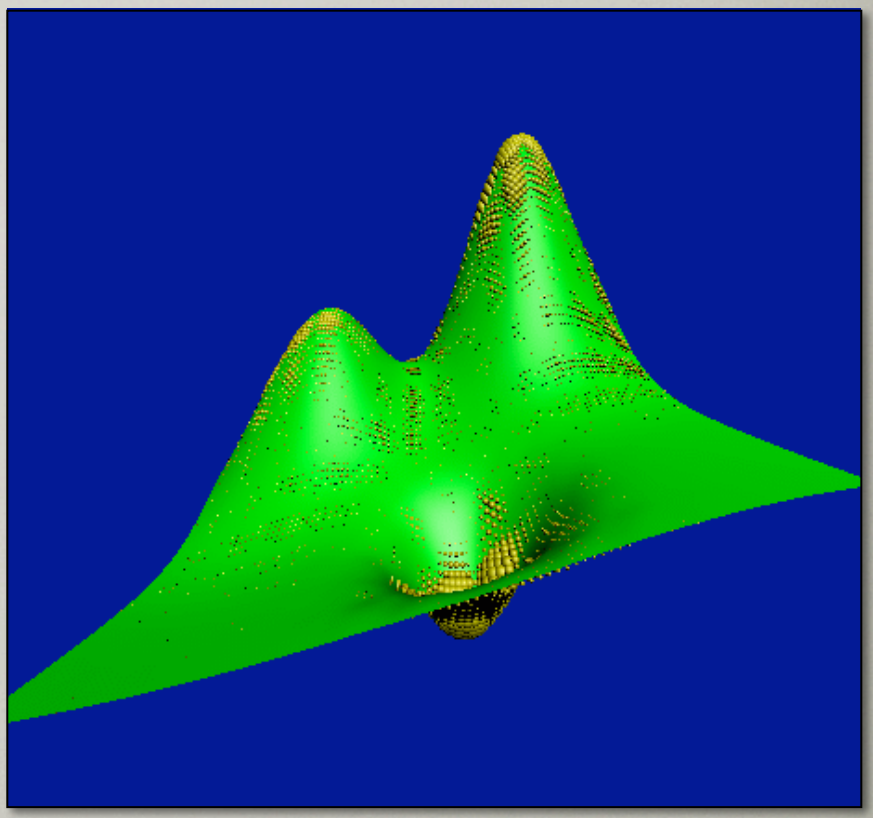
Types of Uncertainty

- Experimental Uncertainty
  - NIST defines uncertainty as standard deviation *
- 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*



* Suresh Lodha, Bob Sheehan, Alex Pang and Craig Wittenbrink. Visualizing geometric uncertainty of surface interpolants In Proceedings of the conference on Graphics interface '96 pp. 238--245. 1996.

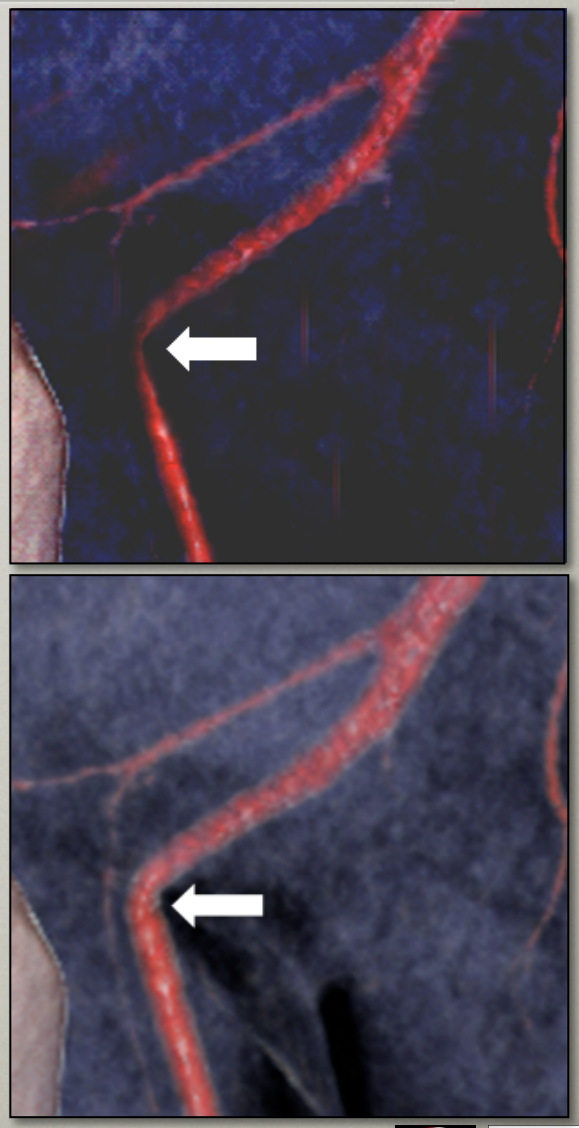
Types of Uncertainty

- Experimental Uncertainty
- Geometric Uncertainty
- Simulation Uncertainty
  - Multimodel or non-determininistic
- Visualization Uncertainty

	Perturbations									
Model	ctl1	ctl2	n1	n2	n3	n4	p1	p2	p3	p4
ETA	●		●				●			
EM	●	●	●	●	●	●	●	●	●	●
NMM	●		●				●			
RSM	●		●	●			●	●		

Types of Uncertainty

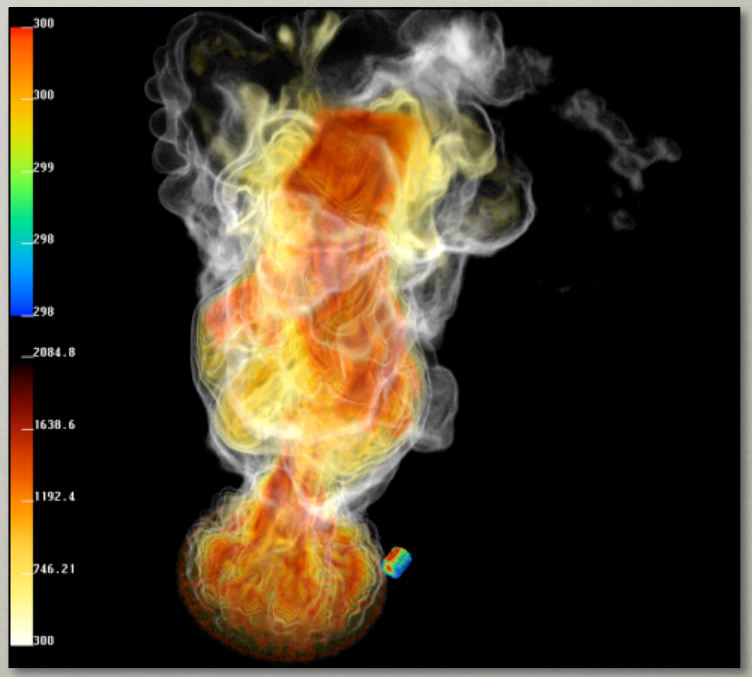
- Experimental Uncertainty
- Geometric Uncertainty
- Simulation Uncertainty
- Visualization Uncertainty
  - Parameters of vis technique lead to differences



* Claes Lundström, Patric Ljung, Anders Persson, and Anders Ynnerman, Uncertainty Visualization in Medical Volume Rendering Using Probabilistic Animation, In IEEE Transactions on Visualization and Computer Graphics, vol. 13, no. 6, pp. 1648-1655, Nov./Dec. 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



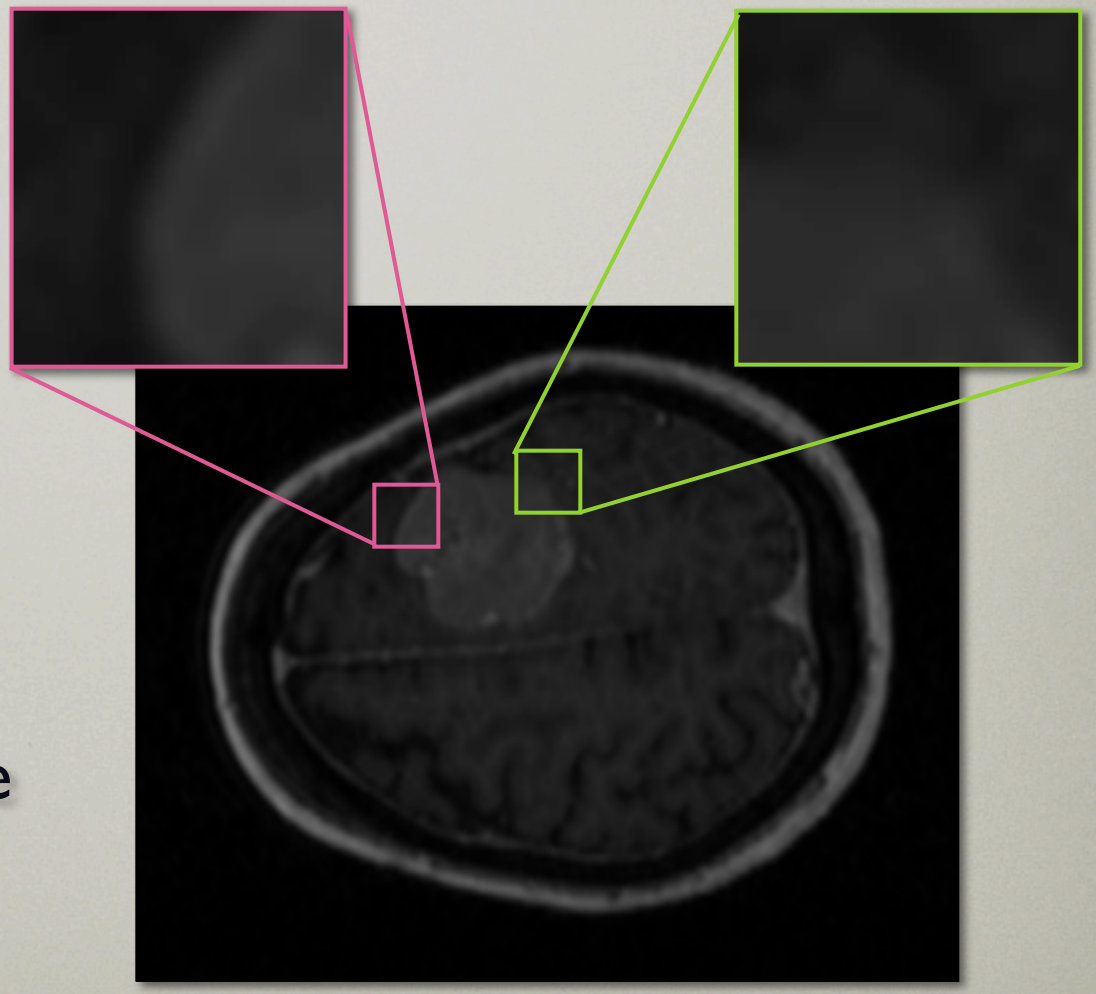
Uncertainty Visualization

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

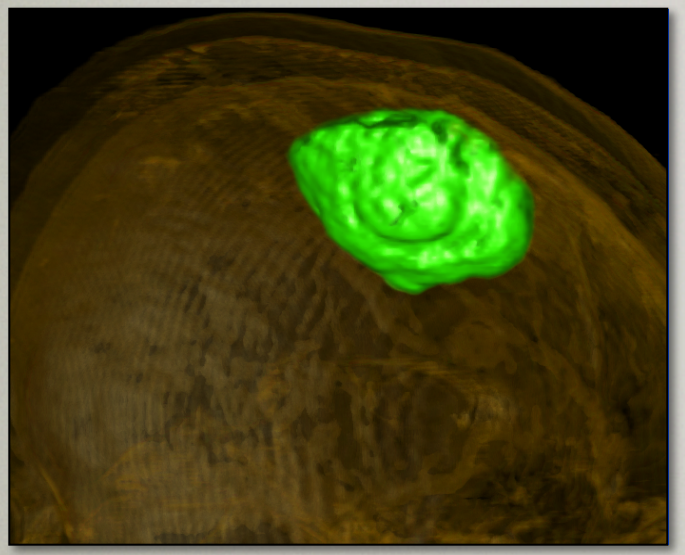
* Chris R. Johnson.
Top Scientific Visualization Research Problems,
In *IEEE Computer Graphics and Applications: Visualization Viewpoints*, Vol. 24, No. 4, pp. 13--17. July/August, 2004.

Brain Tumor Example

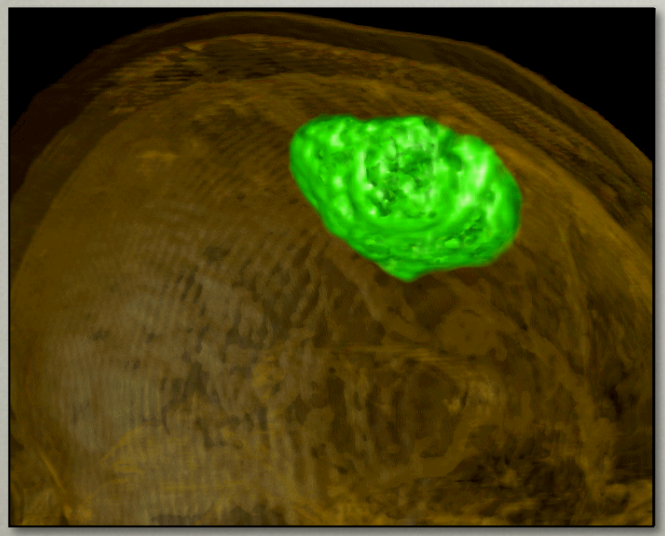
- Fuzzy boundaries exist in the data
- How to distinguish between tumor and gray matter?
- Pre-operative planning: Doctors (and patients!) need to know confidence of the line between tissue types

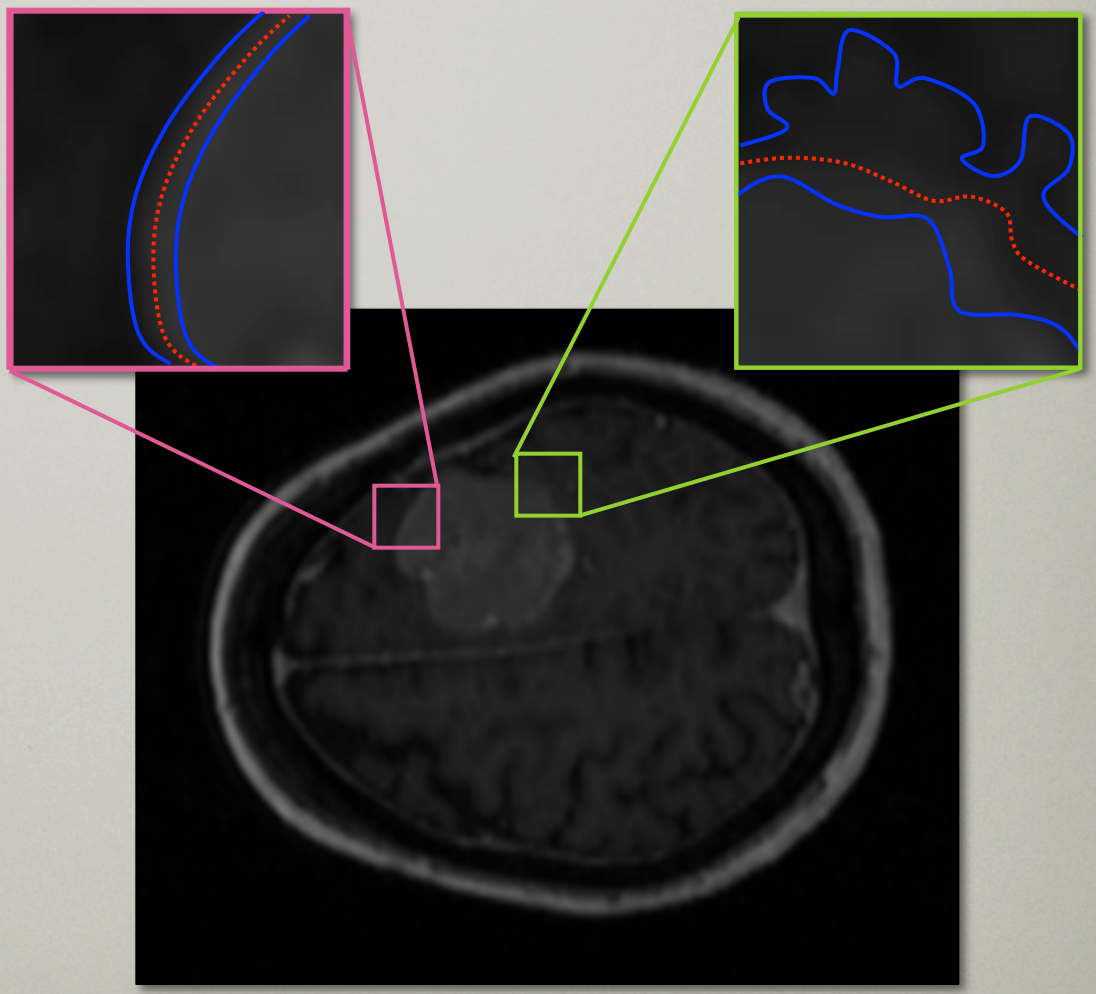


Brain Tumor Example - cont



Or





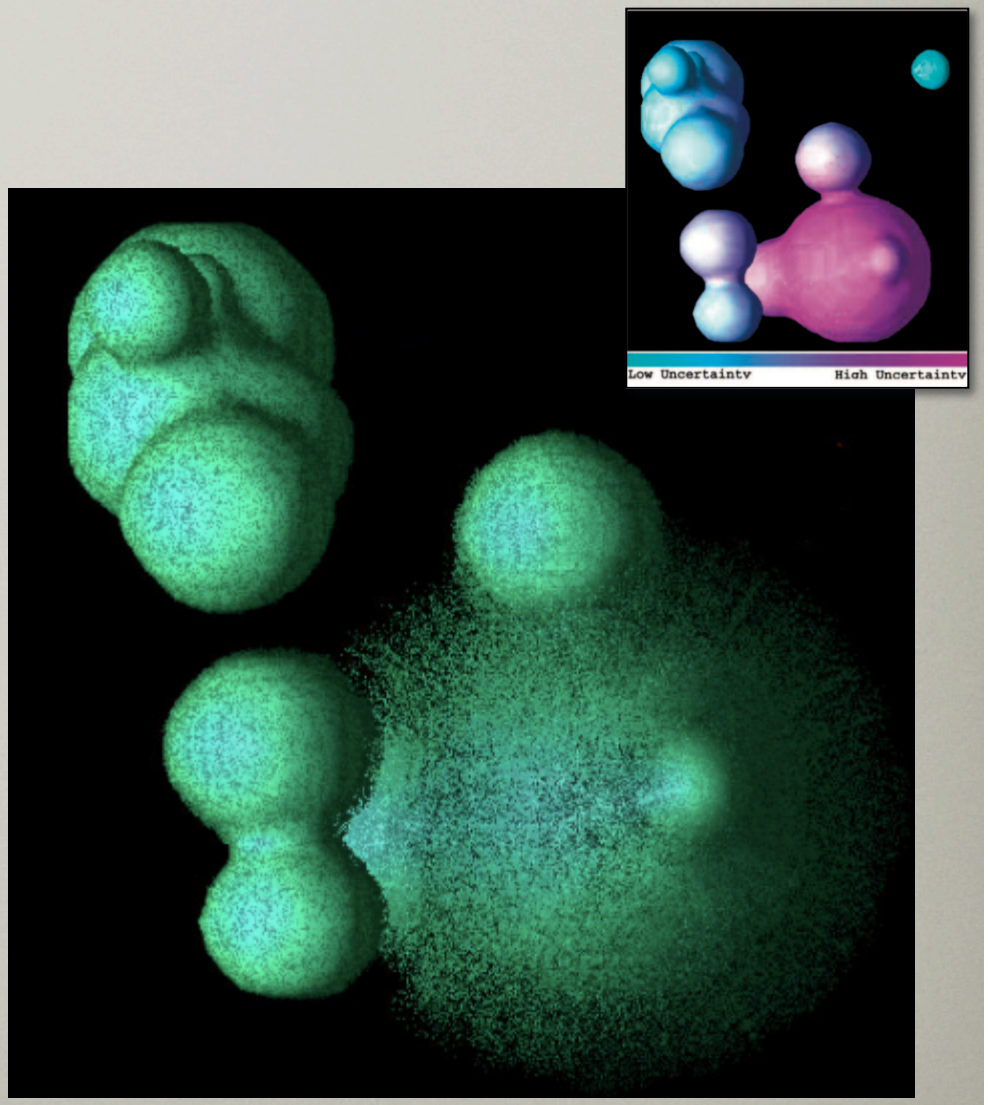
Whole Brain Atlas
http://www.med.harvard.edu/AANLIB/home.html

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



Gevorg Grigoryan and Penny Rheingans.
Point-Based Probabilistic Surfaces to Show Surface Uncertainty
In *IEEE Transactions on Visualization and Computer Graphics*,
Vol. 10, No. 5, pp. 546--573, September/October 2004.

Contributions / Roadmap

- New visualization technique for 1D distribution data
  - Combination of graphical data analysis methods
  - Highlight feature characteristics & uncertainty
- Visualization of multi-dimensional distribution data
- Ensemble-Vis Framework for visual data analysis

Contributions / Roadmap

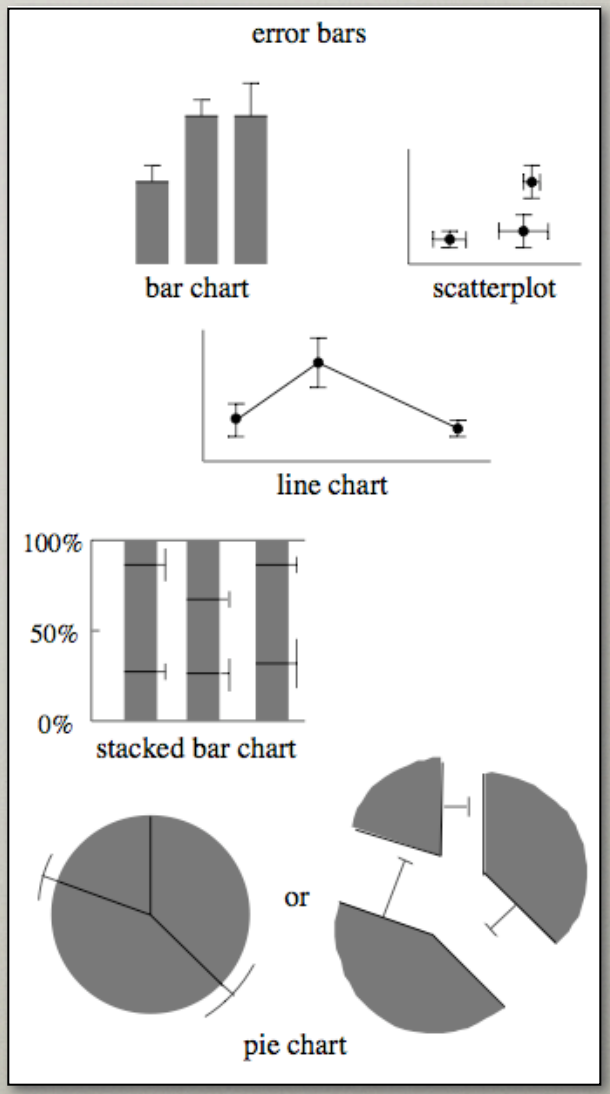
- New visualization technique for 1D distribution data
- Visualization of multi-dimensional distribution data
  - Explore relationship between input parameters and resulting outcomes
  - Global qualitative view, local quantitative
- Ensemble-Vis Framework for visual data analysis

Contributions / Roadmap

- New visualization technique for 1D distribution data
- Visualization of multi-dimensional distribution data
- Ensemble-Vis Framework for visual data analysis
  - User driven, component-based
  - Combine information and scientific visualizations

How is Uncertainty Usually Presented?

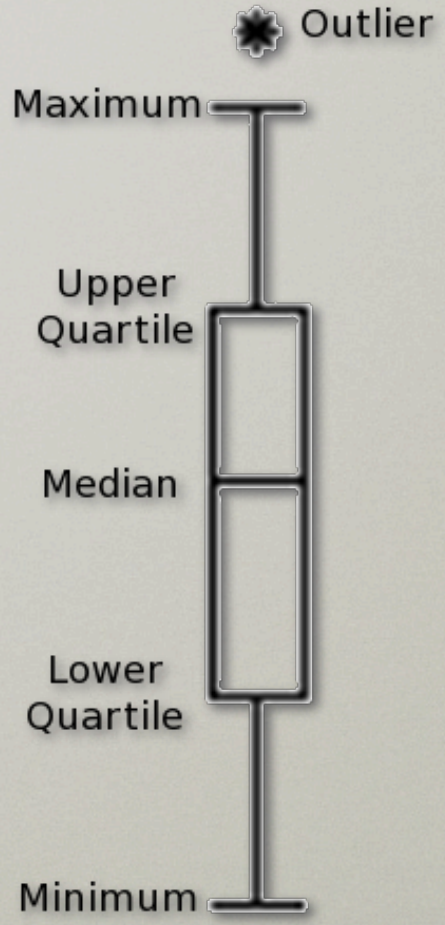
- Error bars
  - standard deviation
  - simple addition to display



Chris Olston and Jock D. Mackinlay.
Visualizing Data with Bounded Uncertainty.
In *Proceedings of the IEEE Symposium on Information (InfoVis'02)*, pp. 37-40, 2002.

How is Uncertainty Usually Presented?

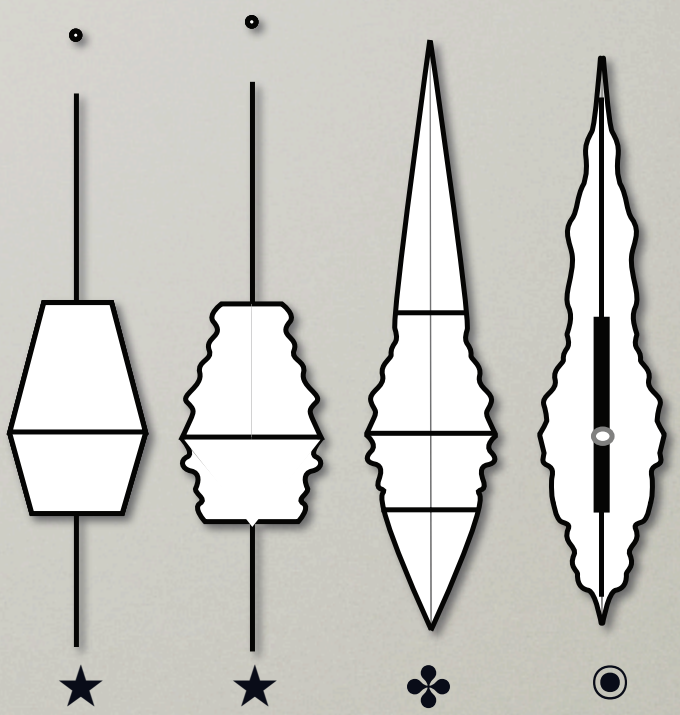
- Boxplots
  - Show full range of data
  - Quartile range, including median
  - Outliers



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

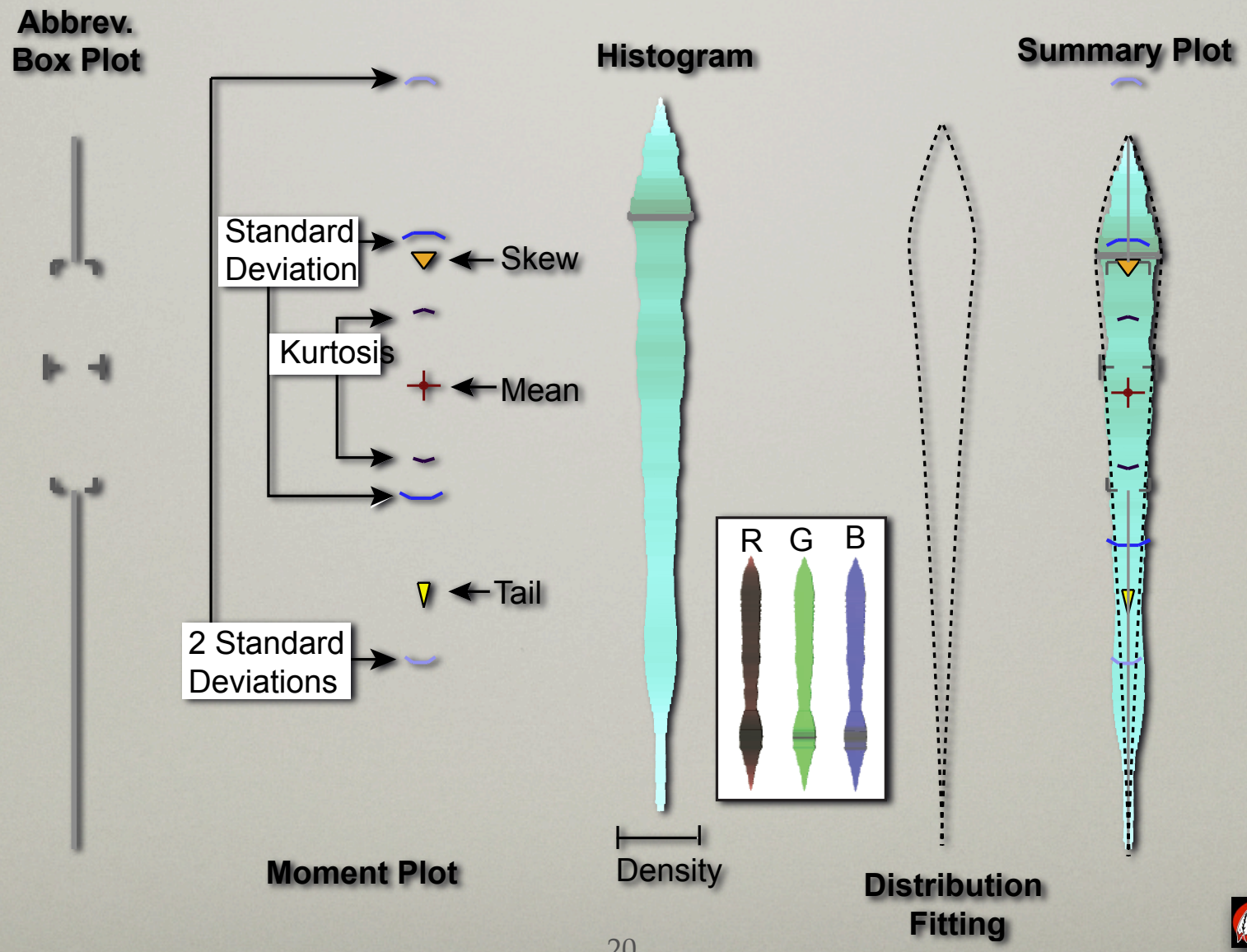


★ Yoav Benjamini. Opening the box of a boxplot. The American Statistician 42, 4, pp. 257–262, 1988.

✤ Warren Esty and Jeffray Banfield. The box-percentile pot. Journal of Statistical Software 8, 17, 2003.

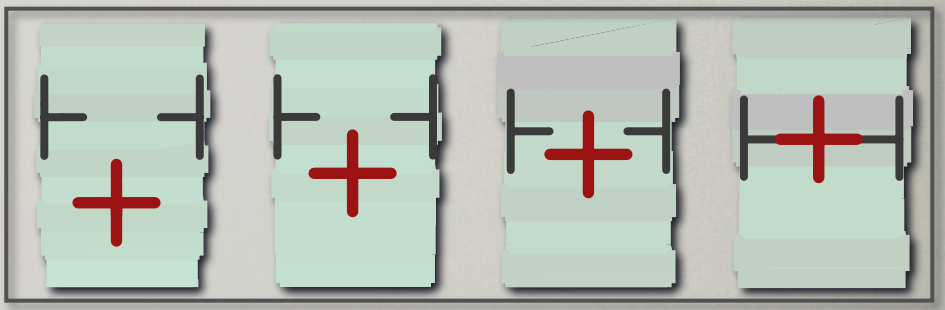
◉ Jerry L. Hintze, and Ray D. Nelson. Violin Plots: A Box Plot-Density Trace Synergism. The American Statistician 52(2):181-84, 1998.

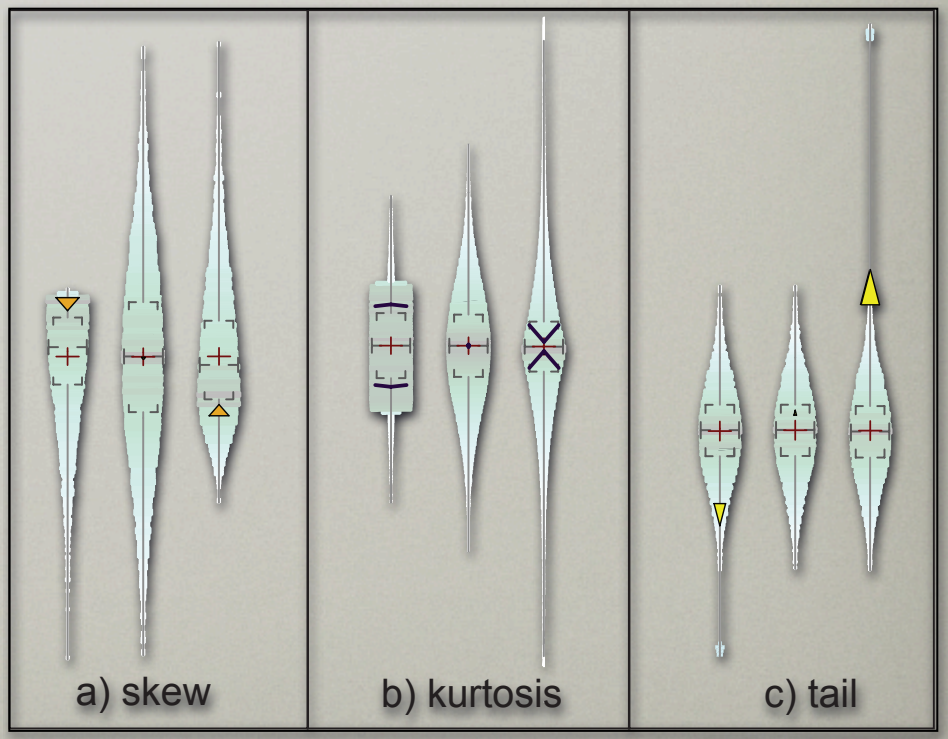
The Summary Plot



Moment Plot

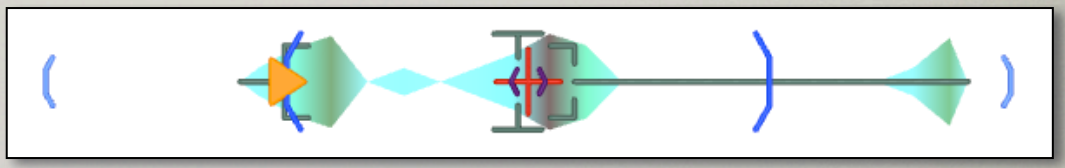
- Statistical measures of feature characteristics
- Signature similar to boxplot
- Can express features hidden by boxplot (e.g. asymmetry)



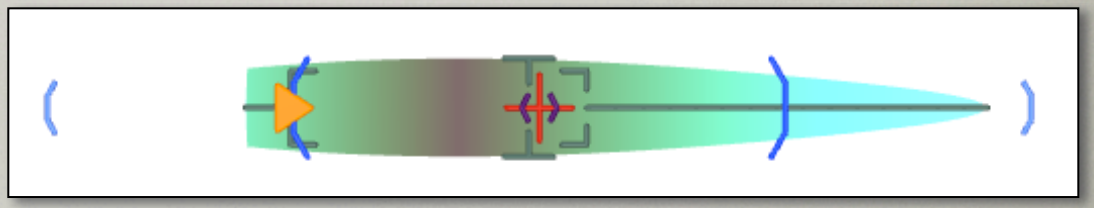


Density Estimate

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



Histogram, 20 bins, 84 samples

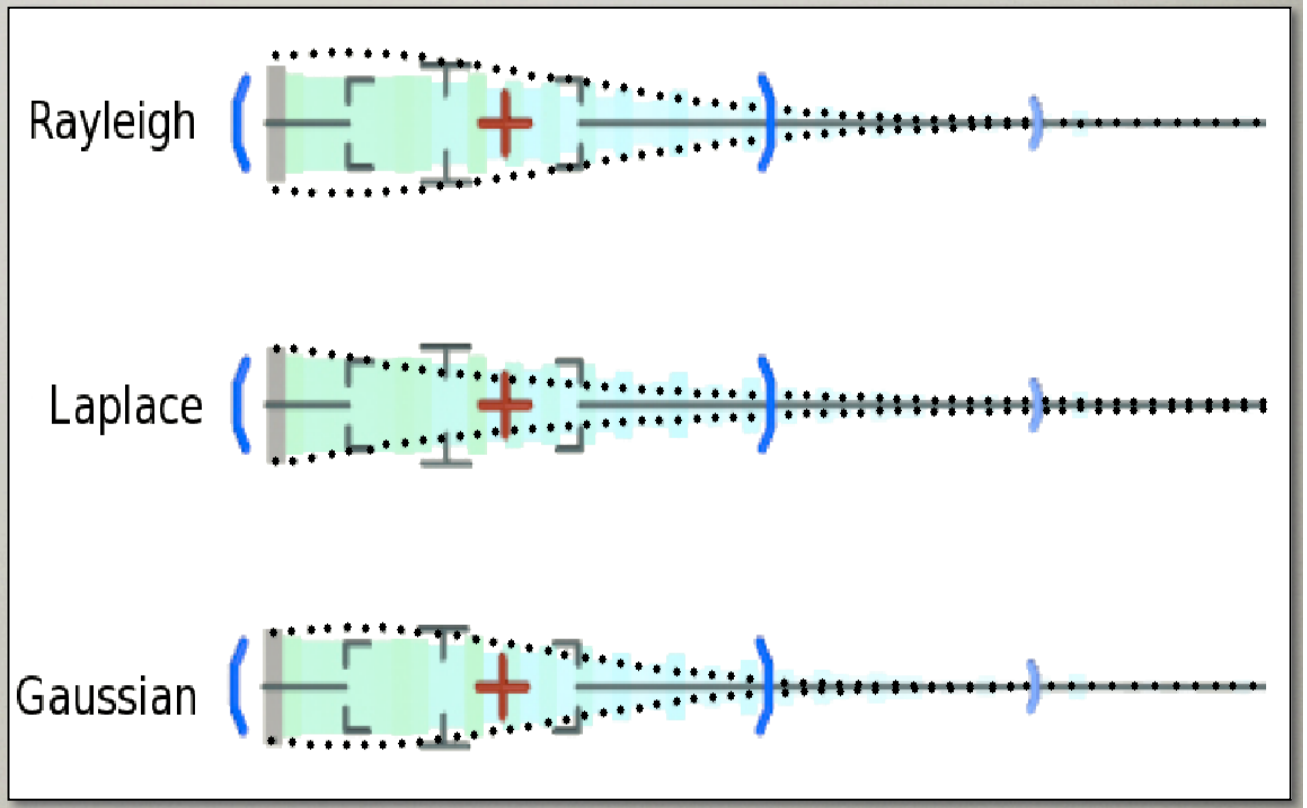


Kernel Density Estimation*

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

Distribution Fitting

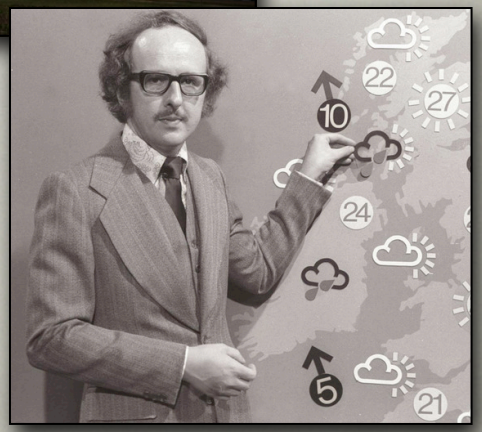
- Fit to canonical distributions from library
- Find a best fit
- Or fit to a chosen distribution



Using the Summary Plots

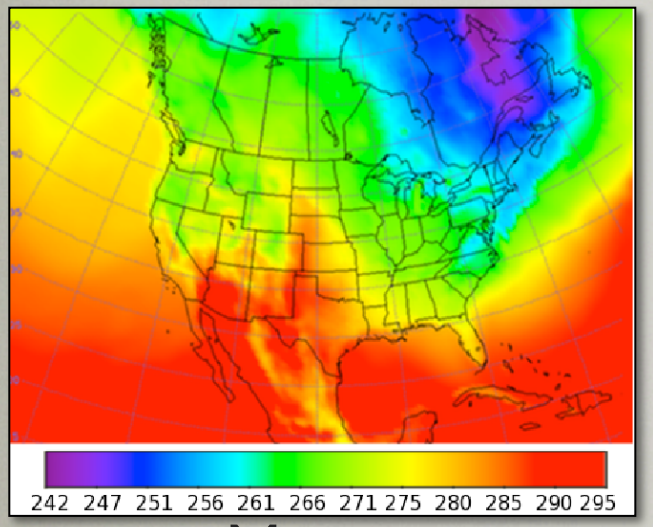
Short-Range Ensemble Forecasts (SREF)

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

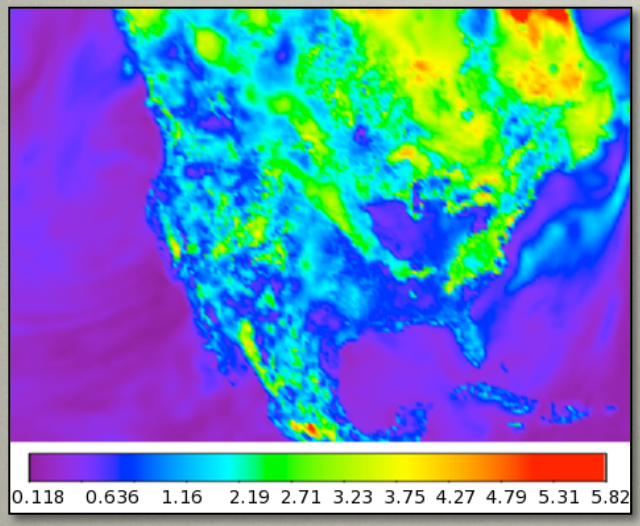


Short-range ensemble forecasting. http://wwwt.emc.ncep.noaa.gov/mmb/SREF/SREF.html.

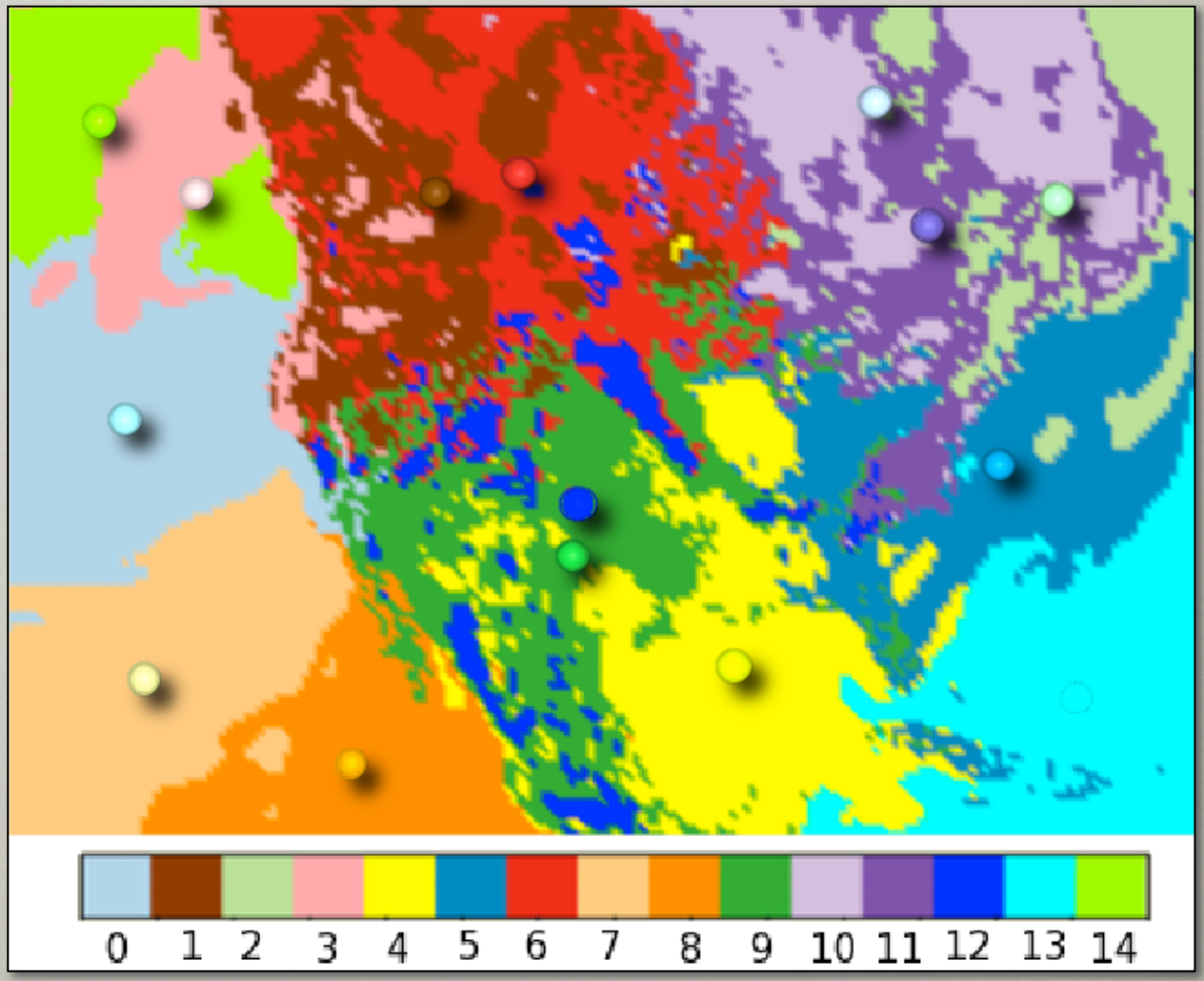
Choose Clusters Based on Variance



Mean

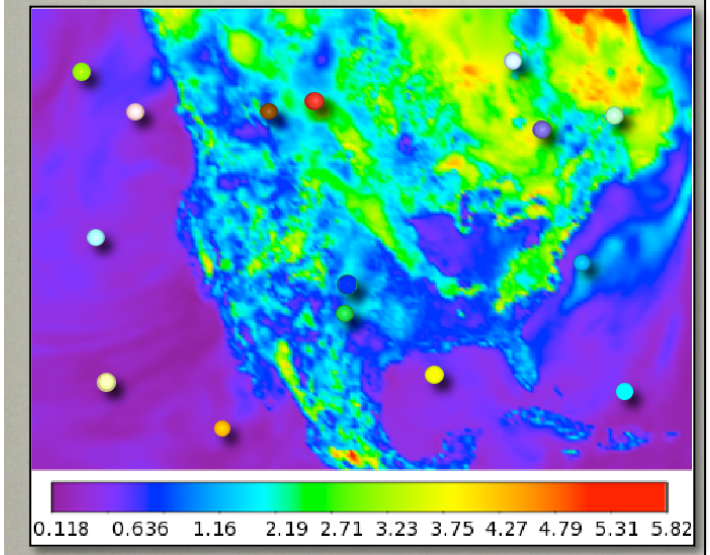


Standard Deviation

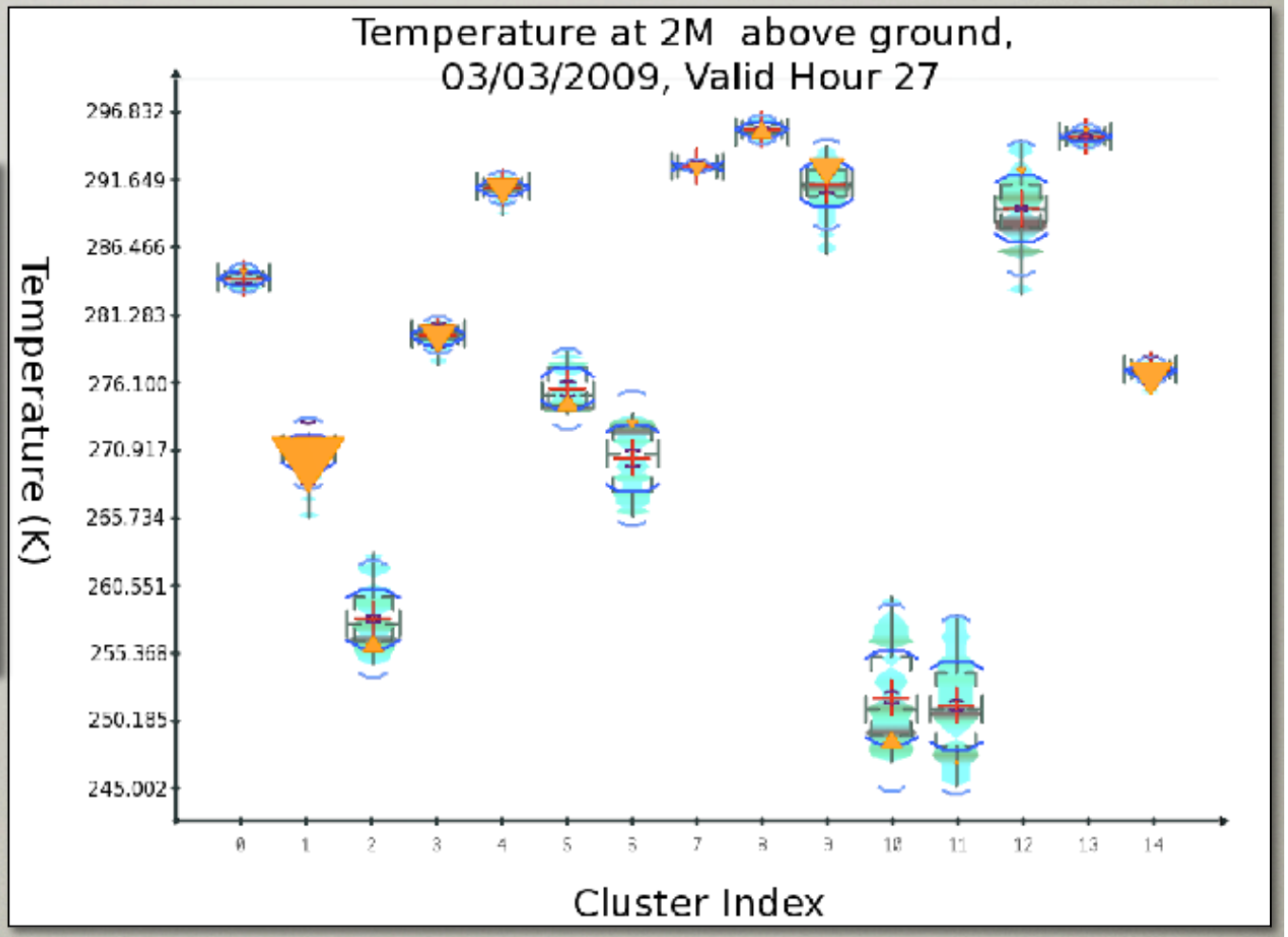


Clusters based on standard deviation and spatial position

Summary Plots on Clusters



Standard Deviation



In Summary

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

Future Work:

- Extend to 2,3+ dimensional distributions
- Other descriptive statistics

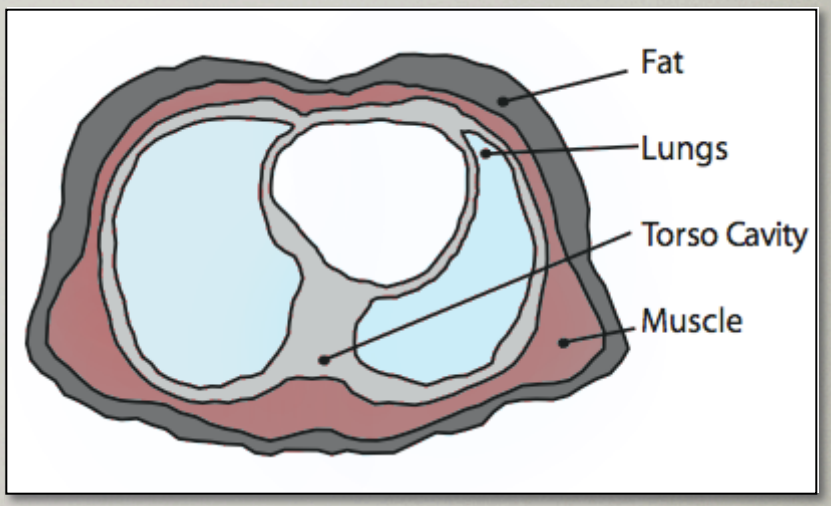
Kristin Potter, Joe Kniss, Richard Riesenfeld and Chris R. Johnson.
Visualization of Summary Statistics and Uncertainty.
Submitted to EuroVis 2010.

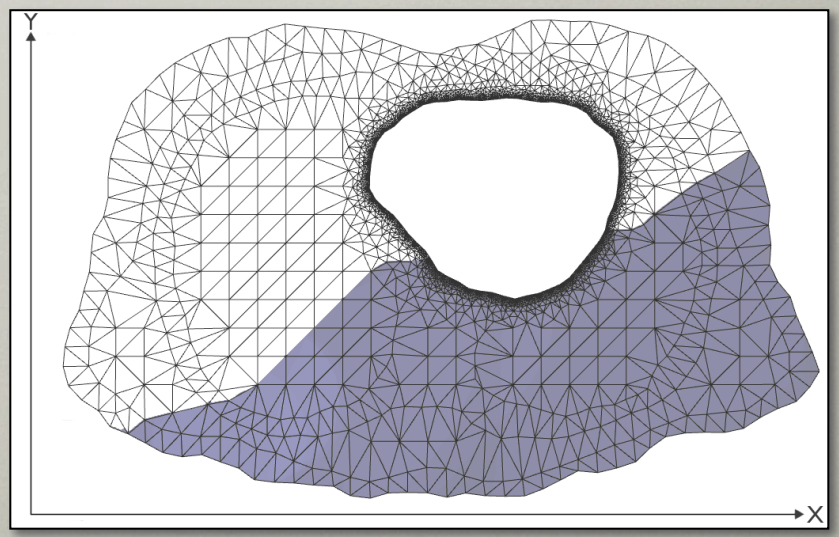
Multi-Dimensional Distribution Data

- Uncertainty across entire spatial domain, not just a single position
- Impact of input parameters on outcome

Electrical Conductivity of the Heart

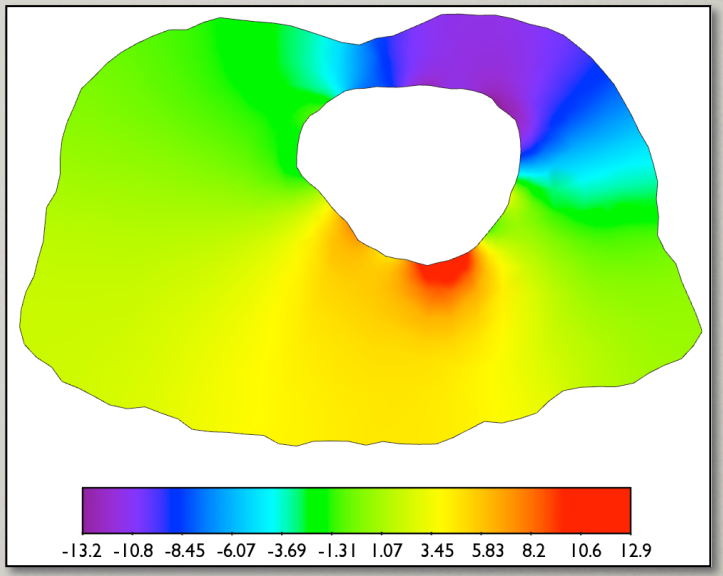
- Electrocardiogram
- Simulate how signals from the heart propagate across the torso
- Distinguish normal changes (breathing, movement) from abnormal heart function



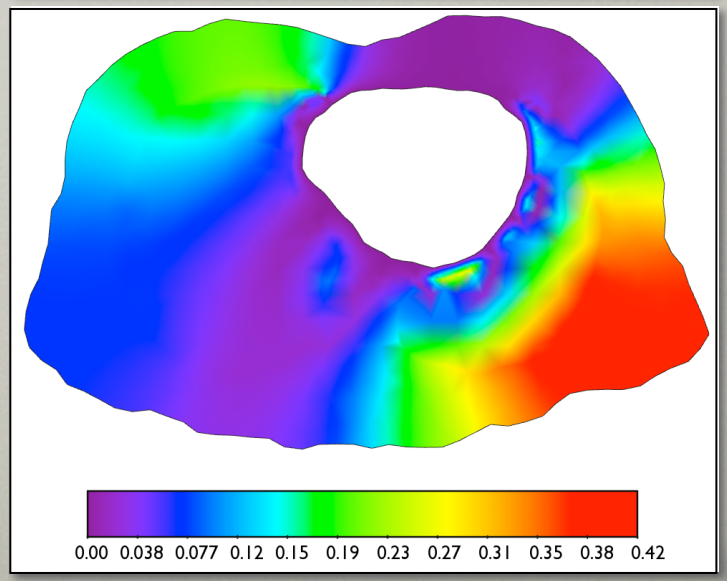


Potentials Data

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



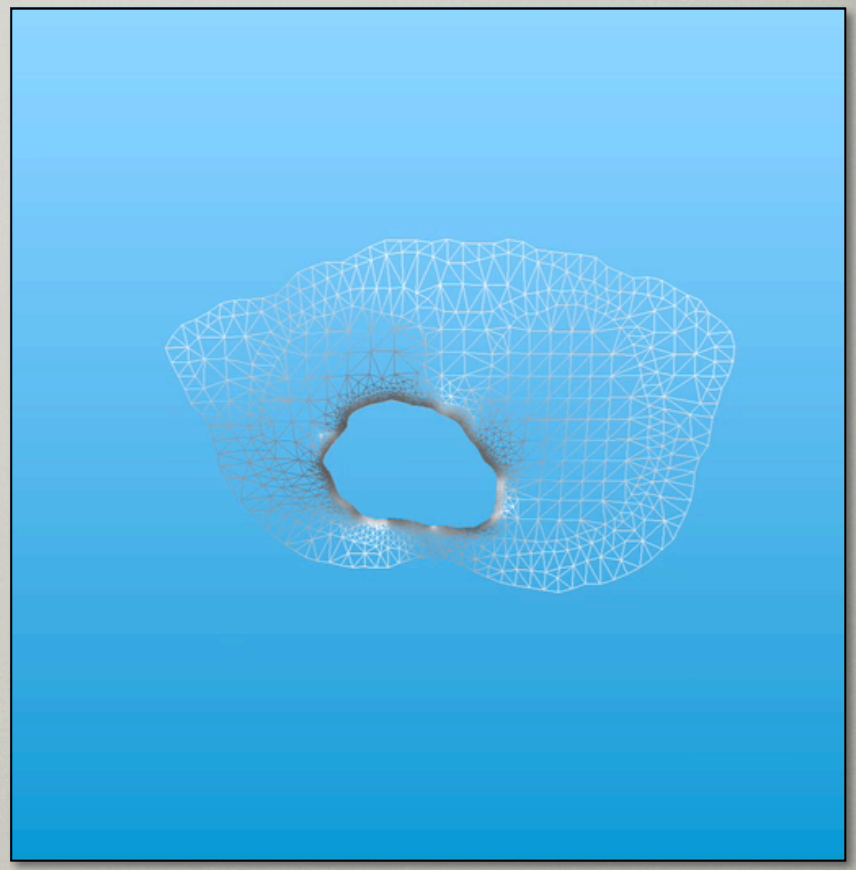
Mean

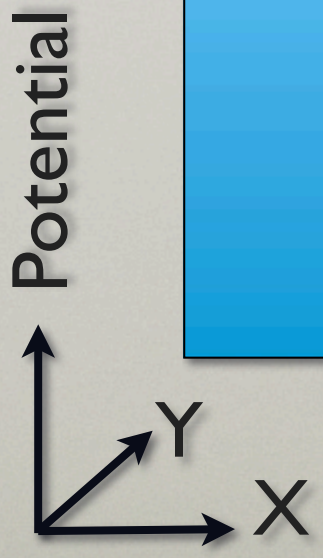


Standard Deviation

Displacement Mapping

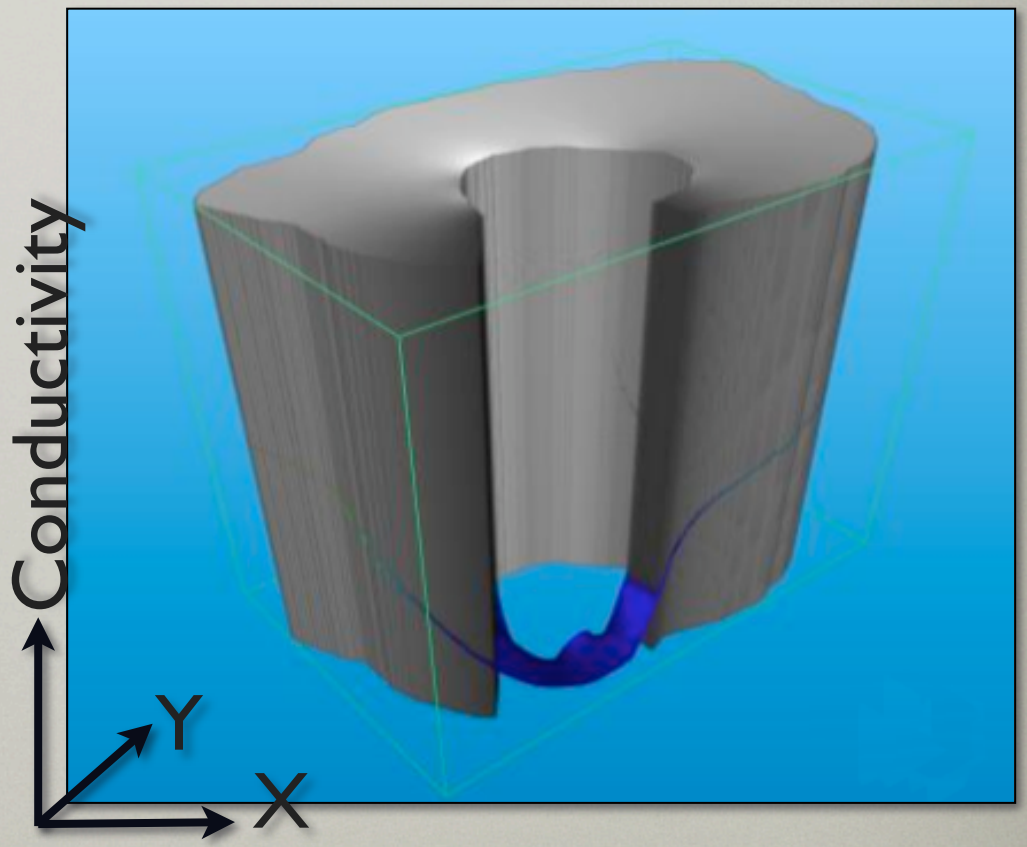
- Using the Z axis
- Height encodes mean
- Color encodes standard deviation
- Really want to see sensitivity of heart potentials to variations in lung conductivity





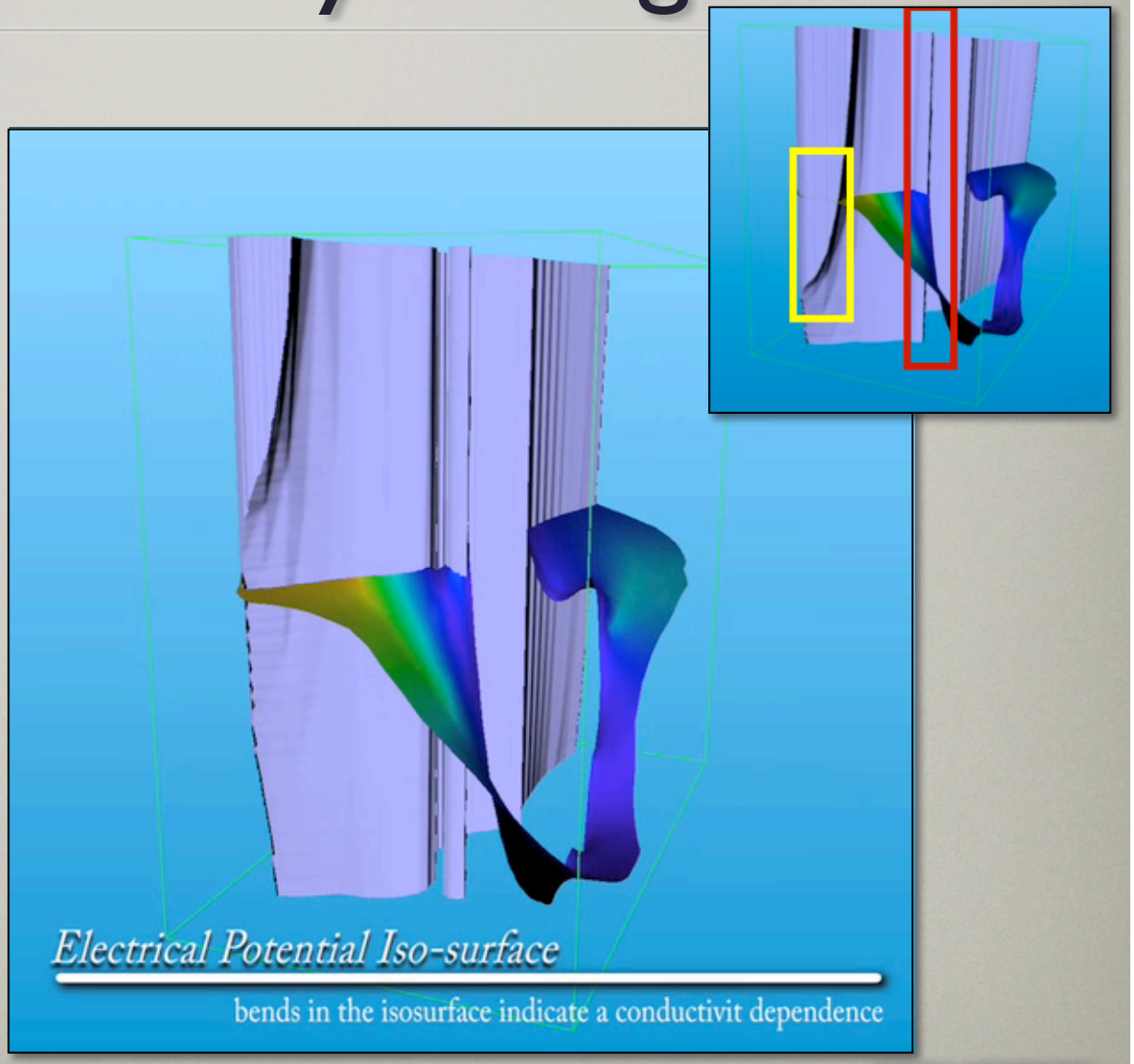
Direct Volume Rendering

- 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



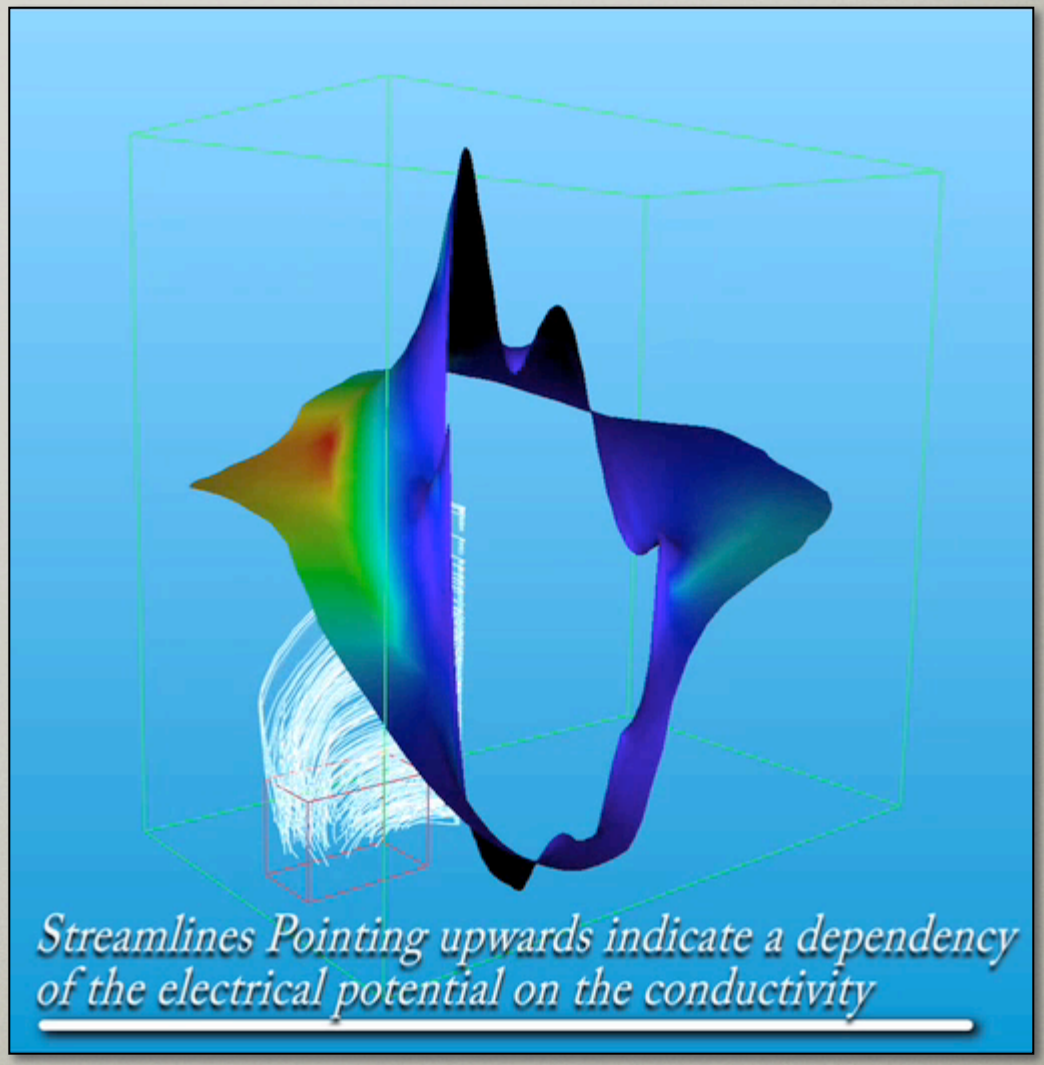
Isosurface Raycasting

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



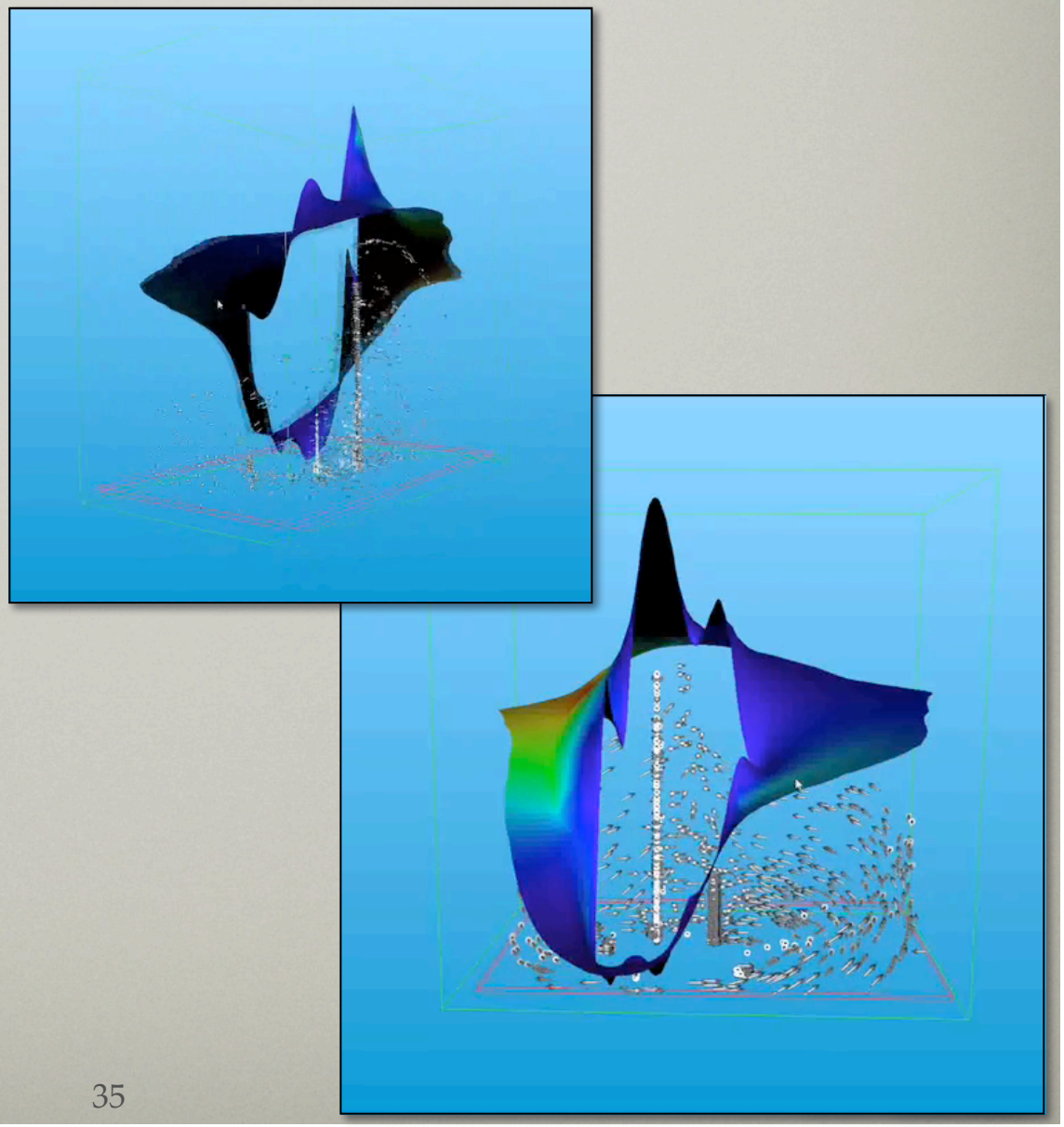
Streamlines

- 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

Future Work

- Higher order input parameters
- Integrate with information visualizations

Kristin Potter, Jens Krüger, and Christopher Johnson.
Towards the Visualization of Multi-Dimensional Stochastic Distribution Data.
In IADIS International Conference on Computer Graphics and Visualization, pp. 191-196, 2008.

Putting it All Together

- Variety of questions on the data, goals of scientist
- Some visual techniques preferable depending on context, locality
- Combine the techniques to allow users to drive the visual exploration

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

Climate Modeling

IPCC
Intergovernamental Panel On Climate Change

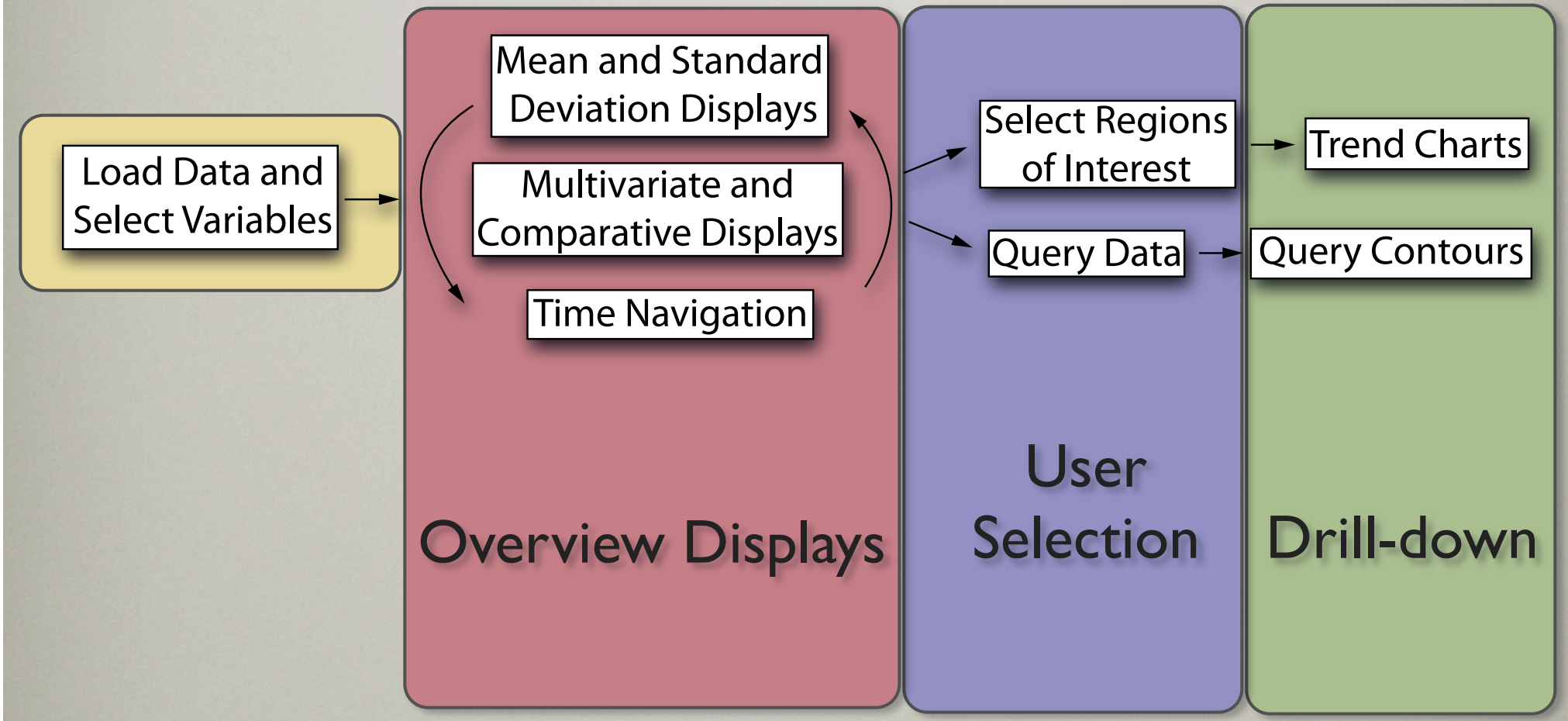
- IPCC Climate of the 20th century
- Spatial domain the whole globe
- Evolution over hundreds of years
- Impact of human activity, trends in natural disasters

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

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

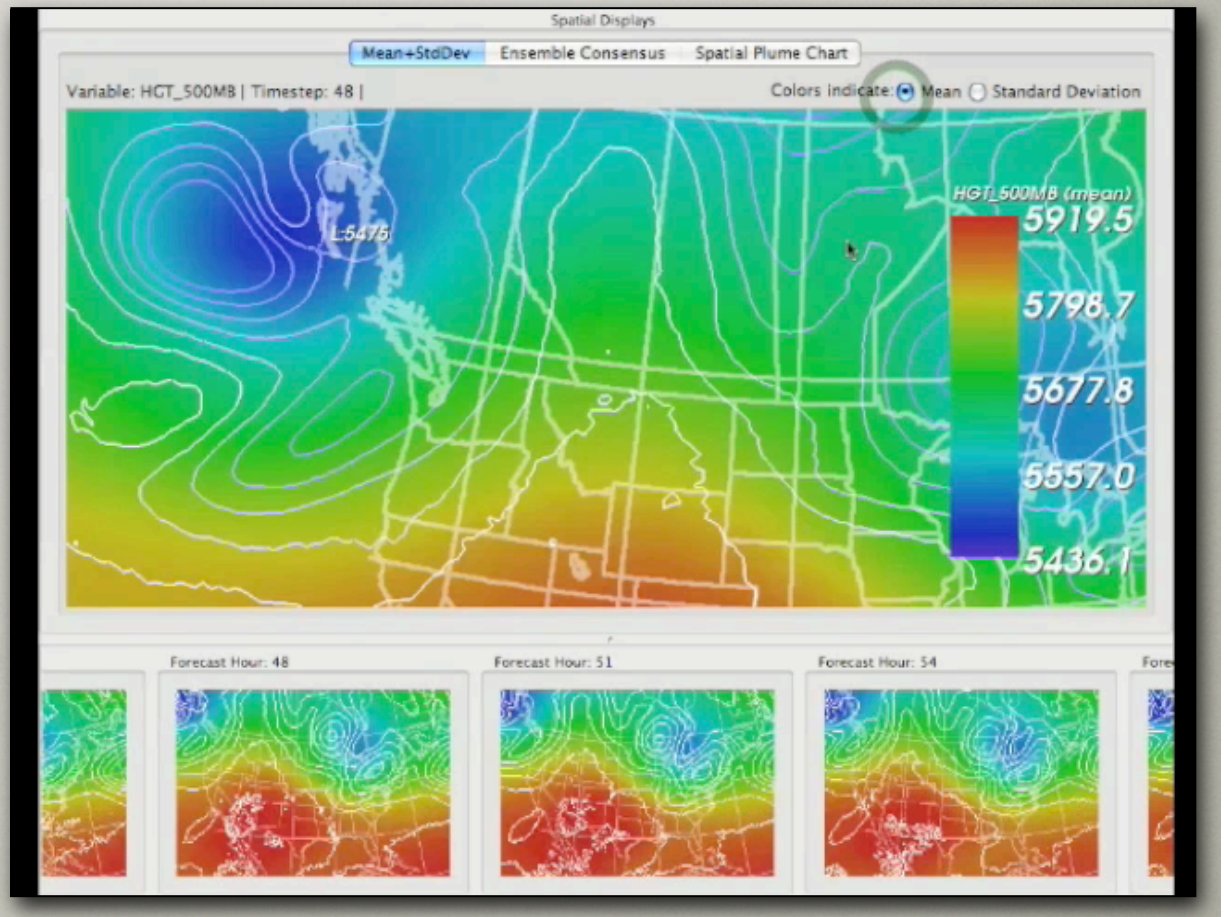
Ensemble-Vis Workflow



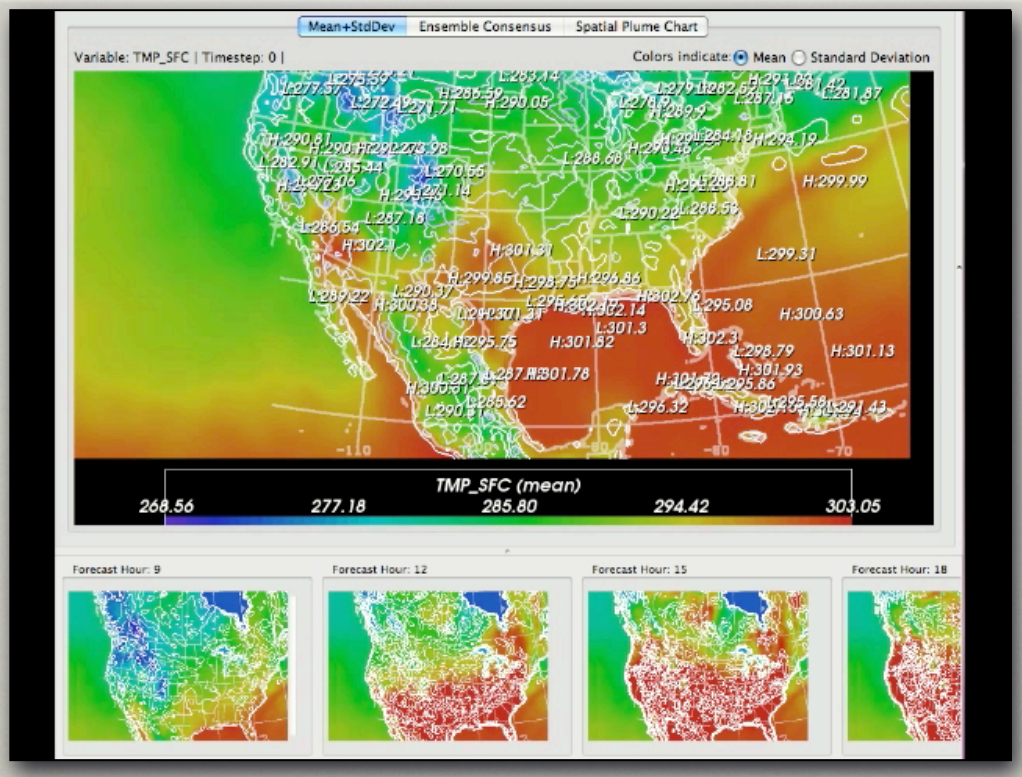
Ensemble Overviews

Mean & standard deviation

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



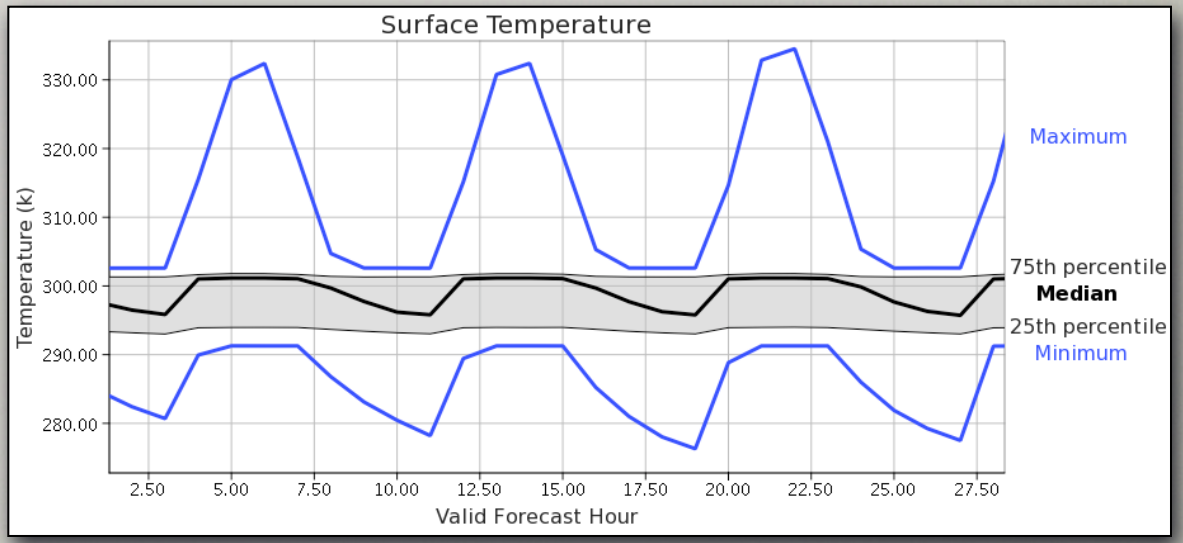
Time Navigation Overviews

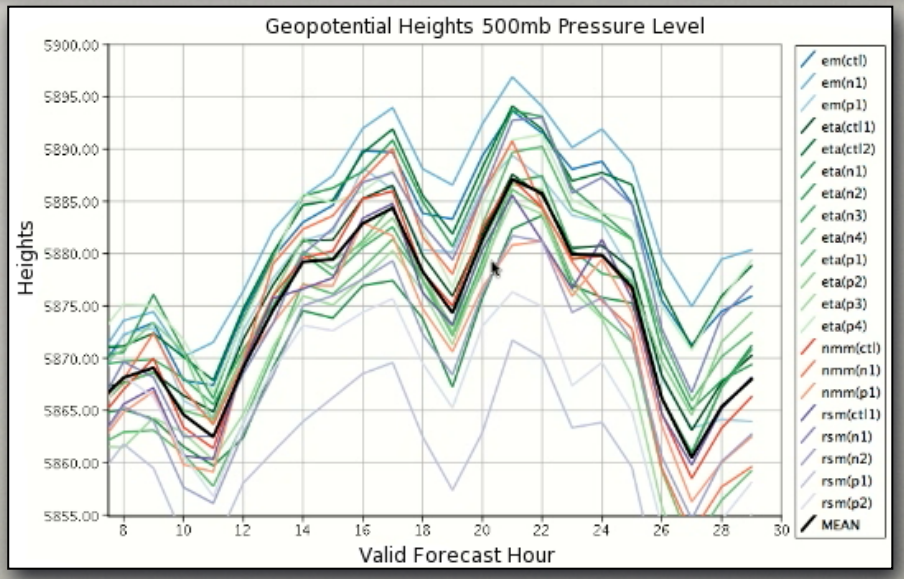


- Small multiples showing each time step across
- Quickly see evolution across time
- Choose time step of interest

Trend Charts

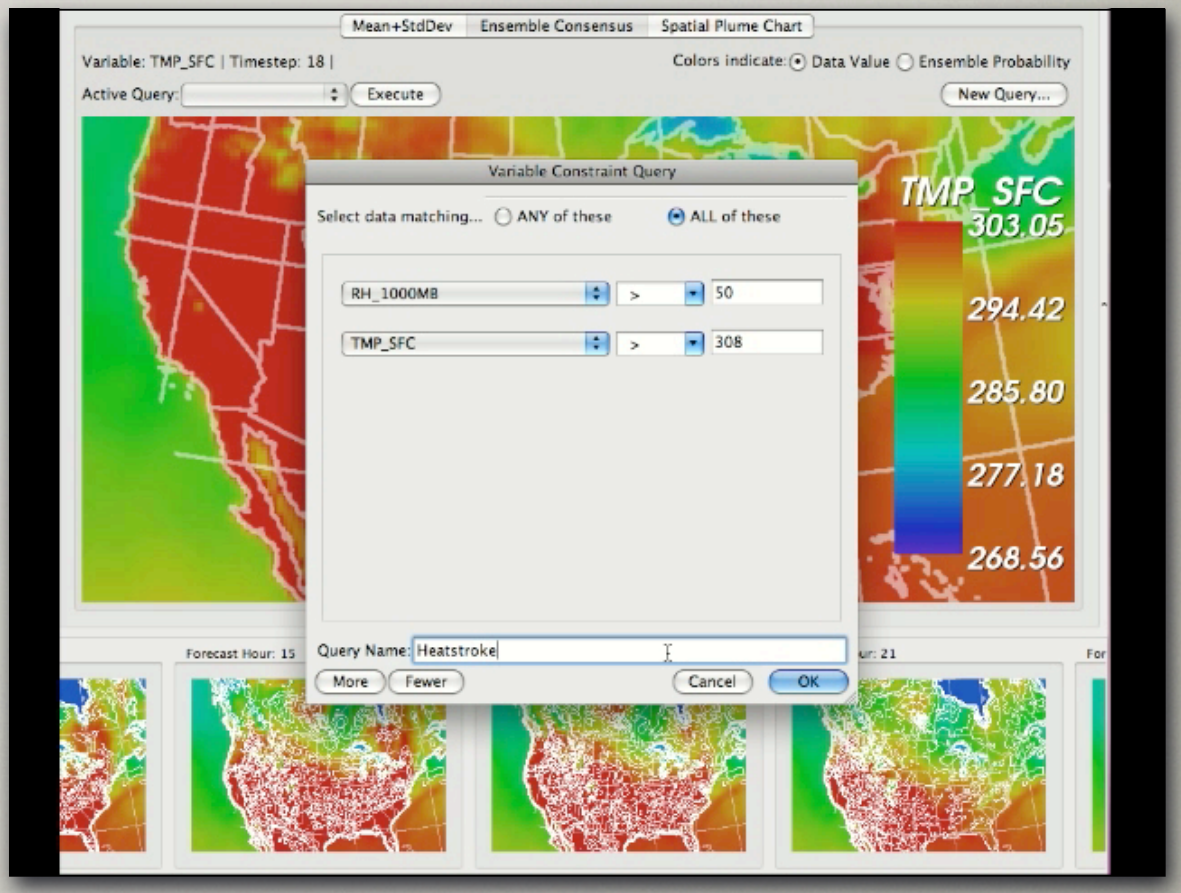
- Select region of interest
- Show statistics like mean, quartiles, etc
- Drill-down to direct data display





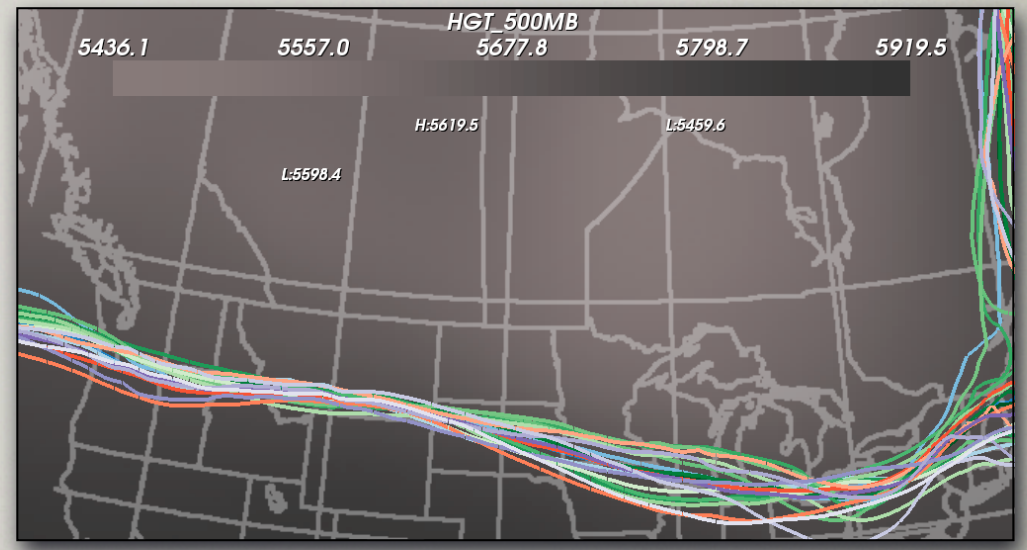
Query Contours

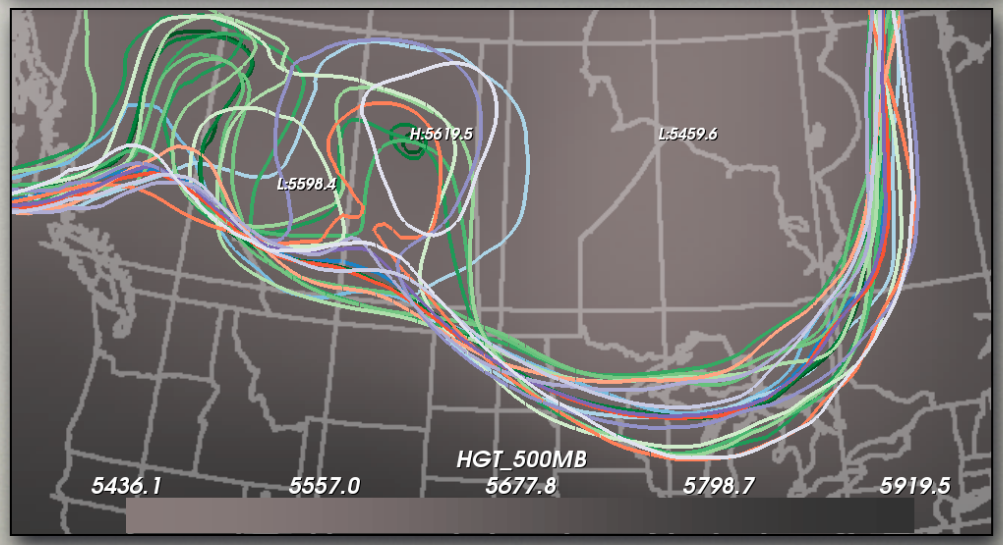
- User-driven query
- Select subset of data
- List of points where conditions are satisfied
- Scalar value at each point indicates number or percentage of satisfying members



Spaghetti Plots

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

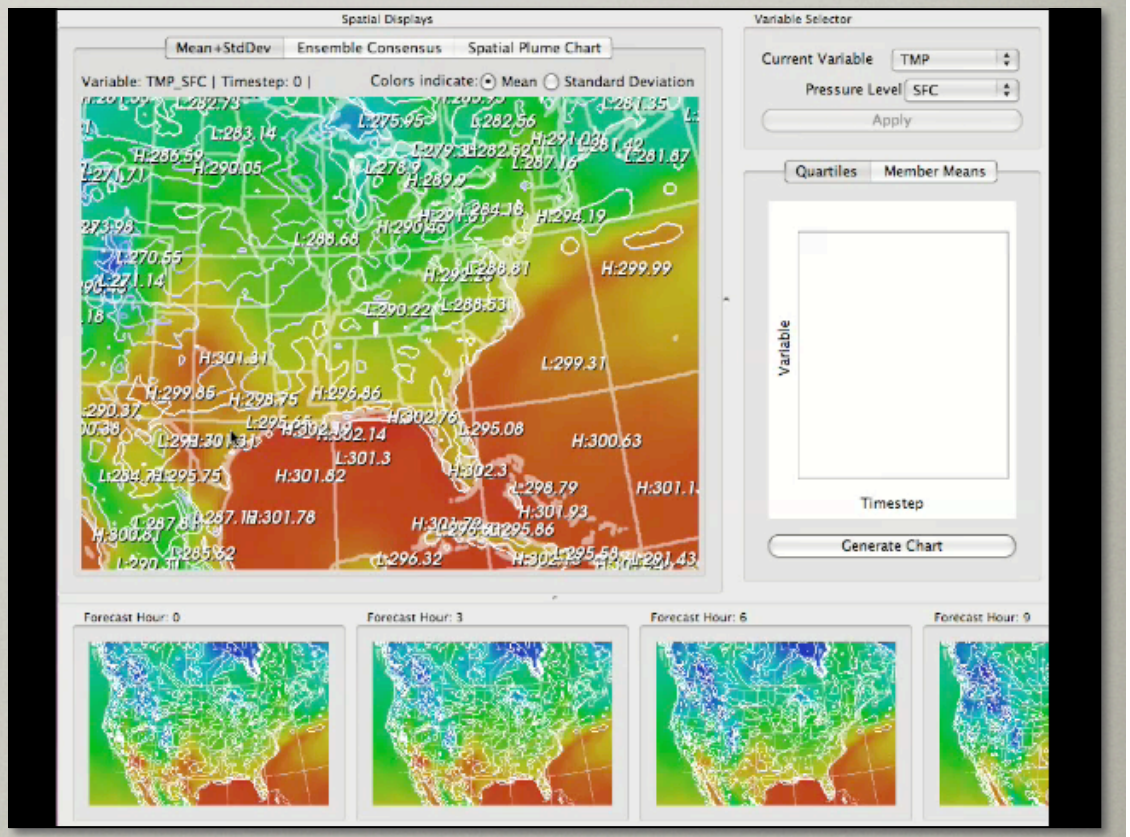




Implementation

Two Prototypical Systems:

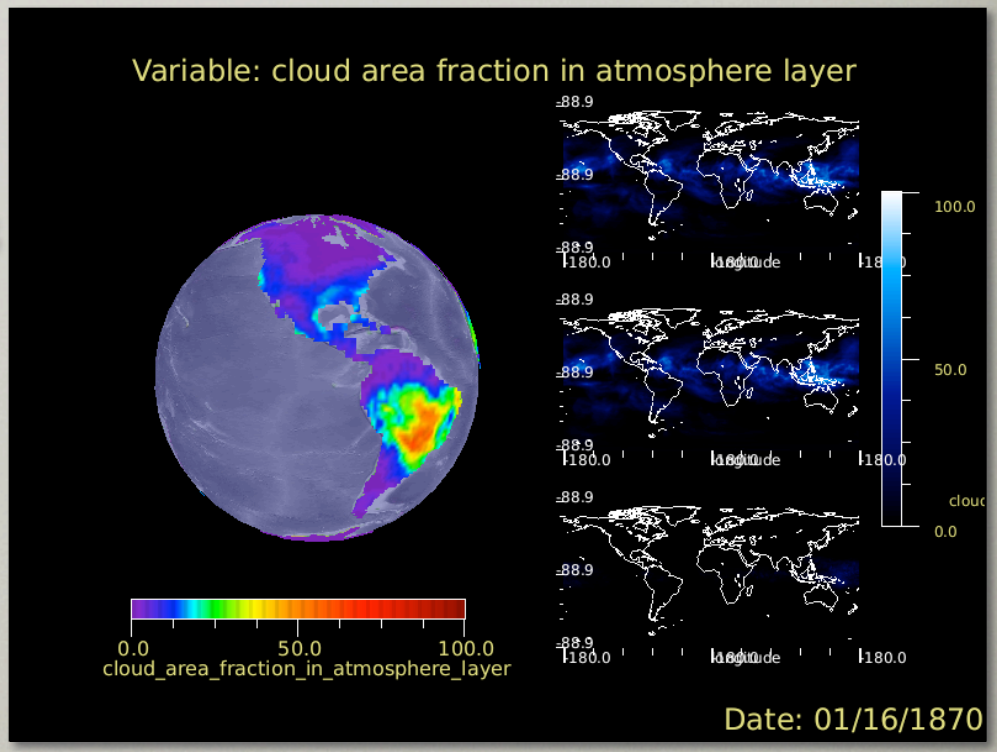
- SREF Weather Explorer
  - VTK filters, Qt widgets
  - Relational database backend MySQL & parallel Netezza



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

Future Work:

- Integration of this framework with higher dimensional data
- Feature detection algorithms

Kristin Potter, Andrew Wilson, Peer-Timo Bremer, Dean Williams, Charles Doutriaux, Valerio Pascucci, and Chris 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,* to appear.

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

Future Work

- Methods for higher spatial dimensions
- In situ data processing
- Feature detection algorithms

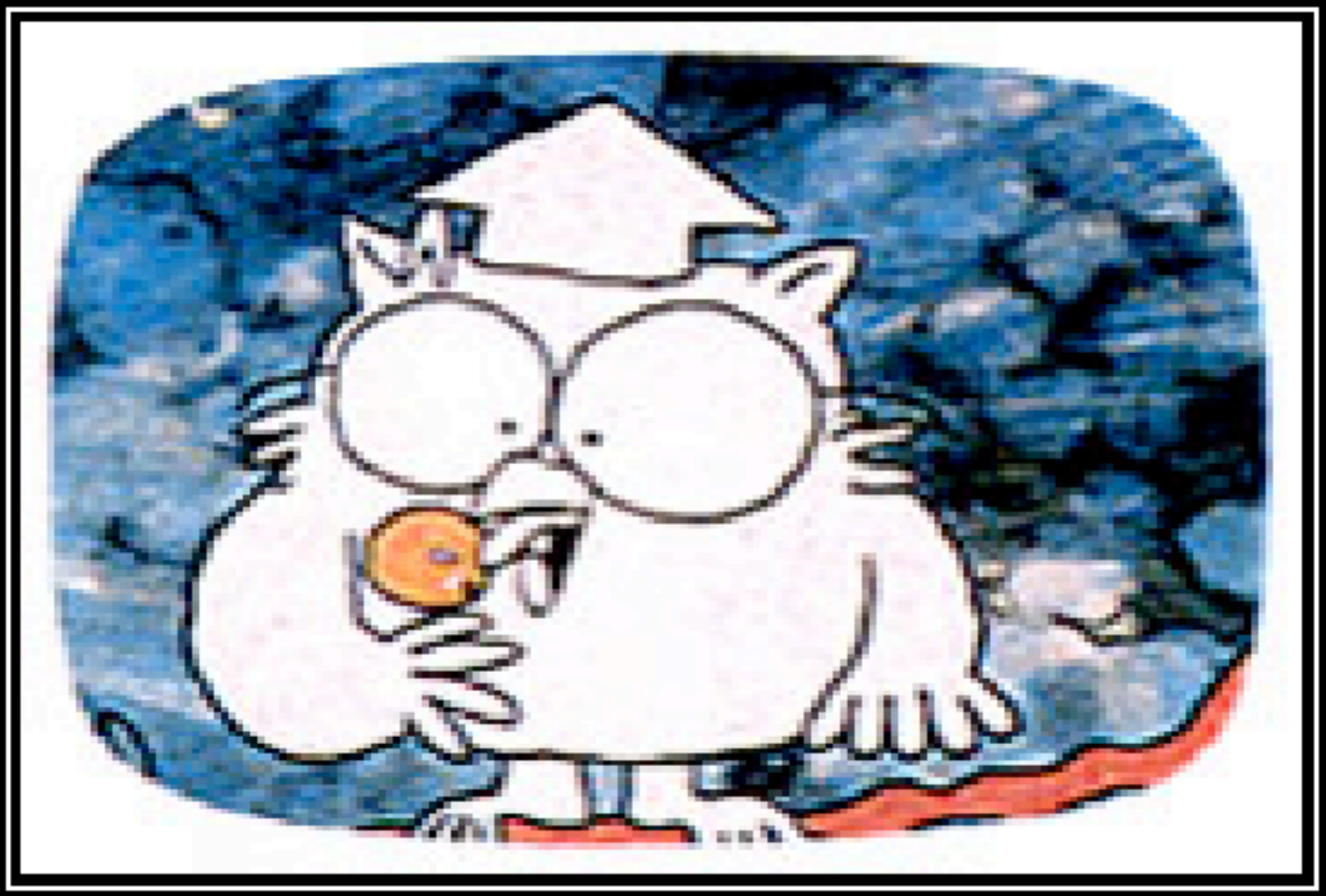
General approaches can only get you so far, specialization needed

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 Energy's 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

Thank You!

- My committee
- Collaborators: Andrew Wilson, Peer-Timo Bremer, Valerio Pascucci, Jens Krueger, Joe Kniss, Sarah Geneser, Roni Choudhury, Erik Anderson
- Scientific Computing and Imaging Institute, School of Computing
- Karen Feinauer, aka wonder woman
- Nathan, Chems and Erik for making me look good
- The ladies: Liz, Miriah, Betty (ladies night is cheaper than therapy!)
- The old school: Amy G, Aaron L, Shaun R, Chris W, Erik R, Gordon, Xaiver, Mike S, Ramy S, Taylor E, Joel D, Dave D
- The new school: Sam G, Roni, Sylvain, Austin, Jacob, Sam P, Erin P, P. Tom F, Abe, Thiago, Nellans, Mike S(& Kelly & Audrey!), Ryan, Chelsea, Sarah G, Erik A
- Unwavering support from Mom and Dad



Uncertainty

Face it kid, Not even Mr. Owl knows how many licks it takes.