Uncertainty Visualization: Beyond Mean and Standard Deviation

IAMCS-KAUST Workshop on Computational Biomedicine and Geophysics

Kristi Potter
Research Scientist, SCI
kpotter@sci.utah.edu
April 5, 2012

SCI INSTITUTE

Technological Advances in Science

Advances in Computing

- increases in storage
- more processors
- higher resolution displays
- greater bandwidth

Advances in Simulation

- higher resolution grids
- longer runs
- more sophisticated models
- better understanding of parameters

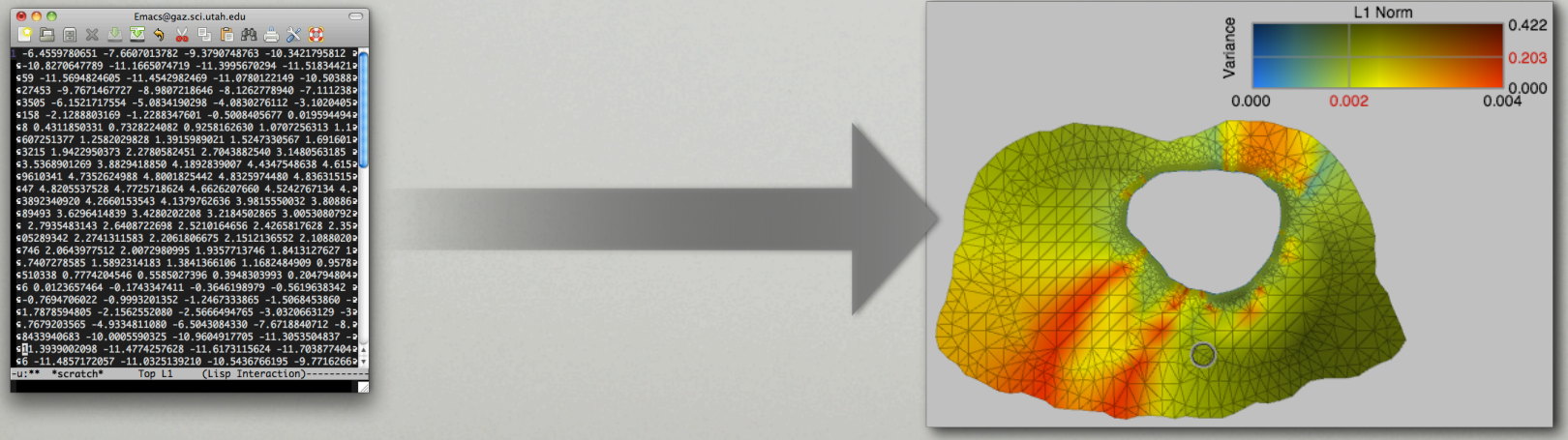
Now we have LOTS of complex data!

Uncertainty in Data

- Scientific data sets are incomplete without indications of uncertainty
- Umbrella term for error, accuracy, confidence level, missing data, inconsistencies, etc
- Multiple definitions depending on field or application
- Fundamental in science, why not in vis?

Visualization is Communication

- Mechanism for understanding large amounts of data
- Easily convey information
- Translate data into visual metaphors
- Expose relationships, patterns, & features



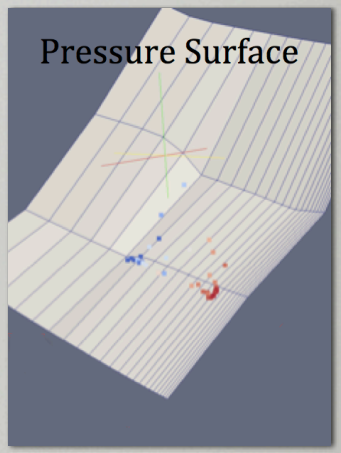
SCI INSTITUTE

Challenges to Uncertainty Visualization

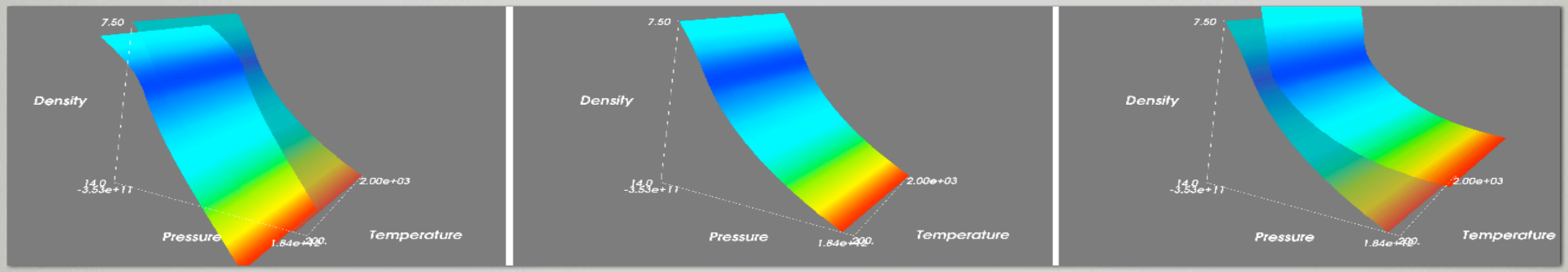
- Adding more info to already large data
  - Visual complexity and clutter
  - Can obscure data
  - Increasing visual "uncertainty" can decrease understanding
- What is an appropriate visual metaphor?
- Is the uncertainty primary or secondary?
- No singular definition, no singular solution

An Example Application - Material Models

- Equation of State (EOS) Model
  - characteristics of materials (state changes, stresses, etc)
- Inputs to many simulations
  - material interactions, failures, etc
- Often source of uncertainty
  - quality of simulation results tied to model
  - many models for 1 material - which to choose?

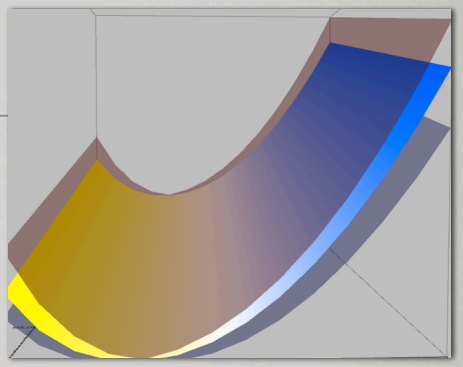


Visualization of EOS Surfaces

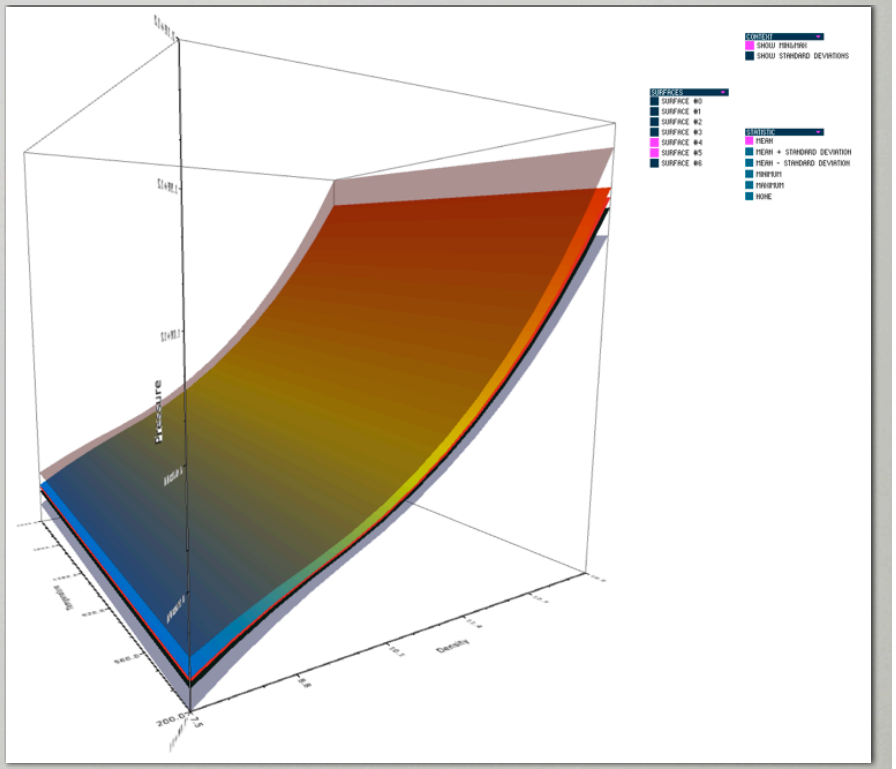


- Multiple surfaces in thermodynamic space
  - (temperature, pressure, density)
- How to display for modelers, analysts, code developers?
  - Series of focus groups, 4 rapid prototypes
- Main goal: display data to spur discussion

A Simple Prototype



- Display collection of surfaces
- Express uncertainty through statistical surfaces
  - (min, max, mean, std. dev)



- point-wise statistics NOT valid realizations!

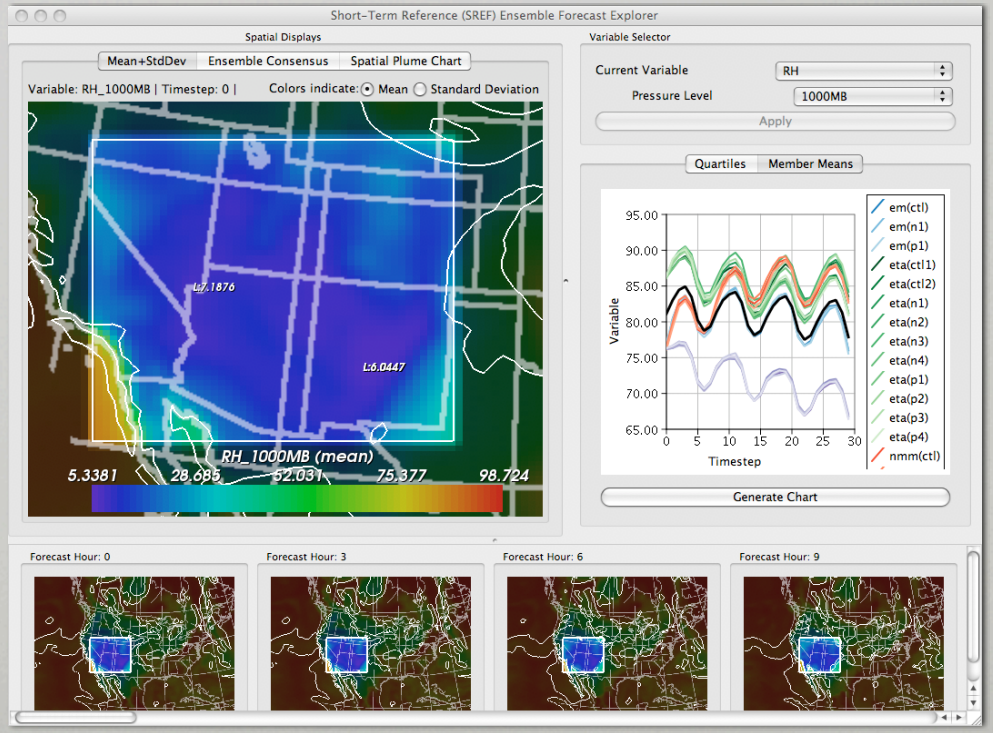
General Approach to Uncertainty Visualization

Reduce large amounts of data using *summaries*

- In what dimension do we summarize?
- Is mean/standard deviation appropriate?
- Do we need multiple summarizations?

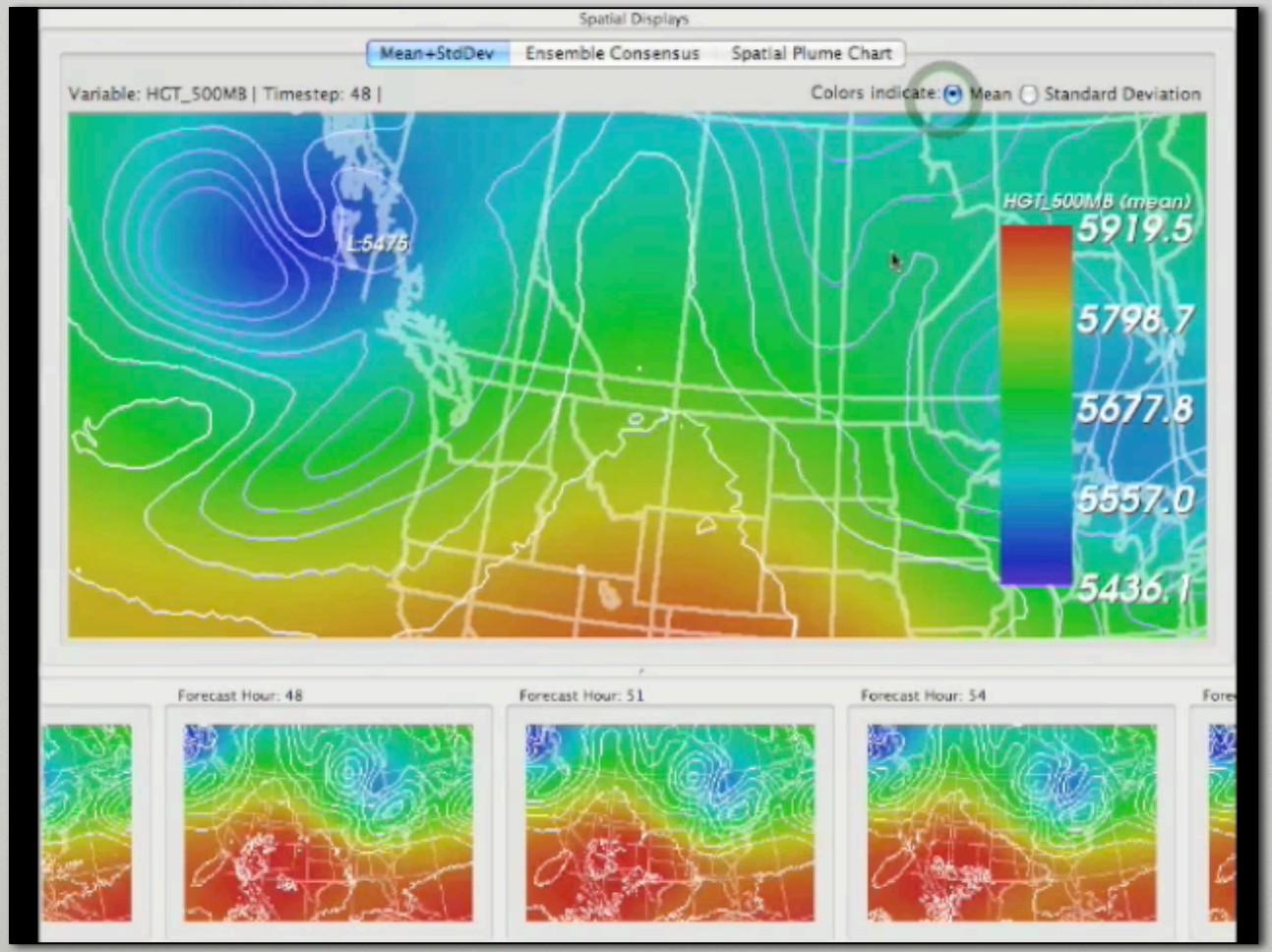
Example 1 - Ensemble-Vis

- Application to weather forecasting
- Multiple windows/linked views
- Summarize across space, time, data



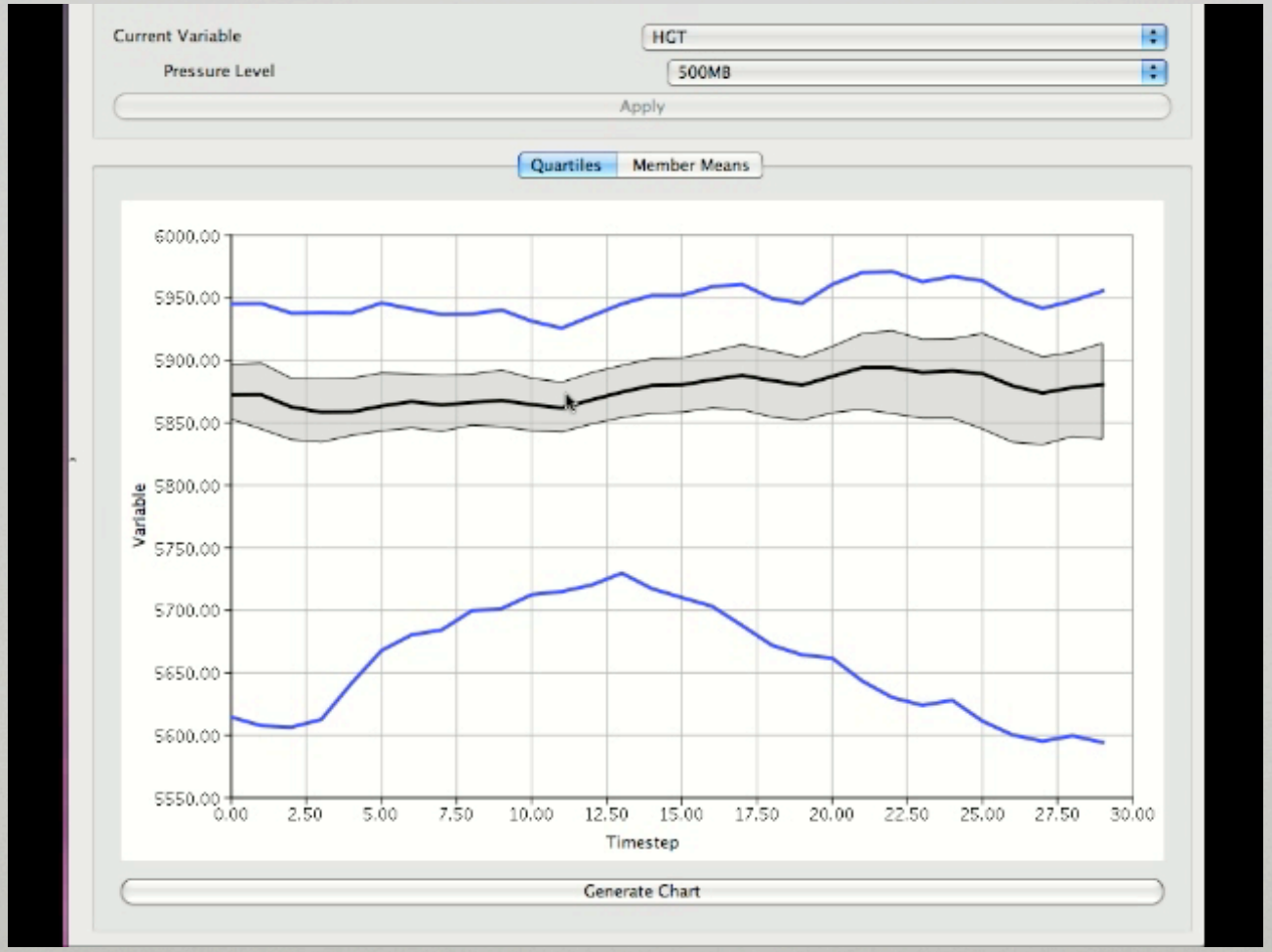
Grid Point Summaries

- Mean & standard deviation at each grid point



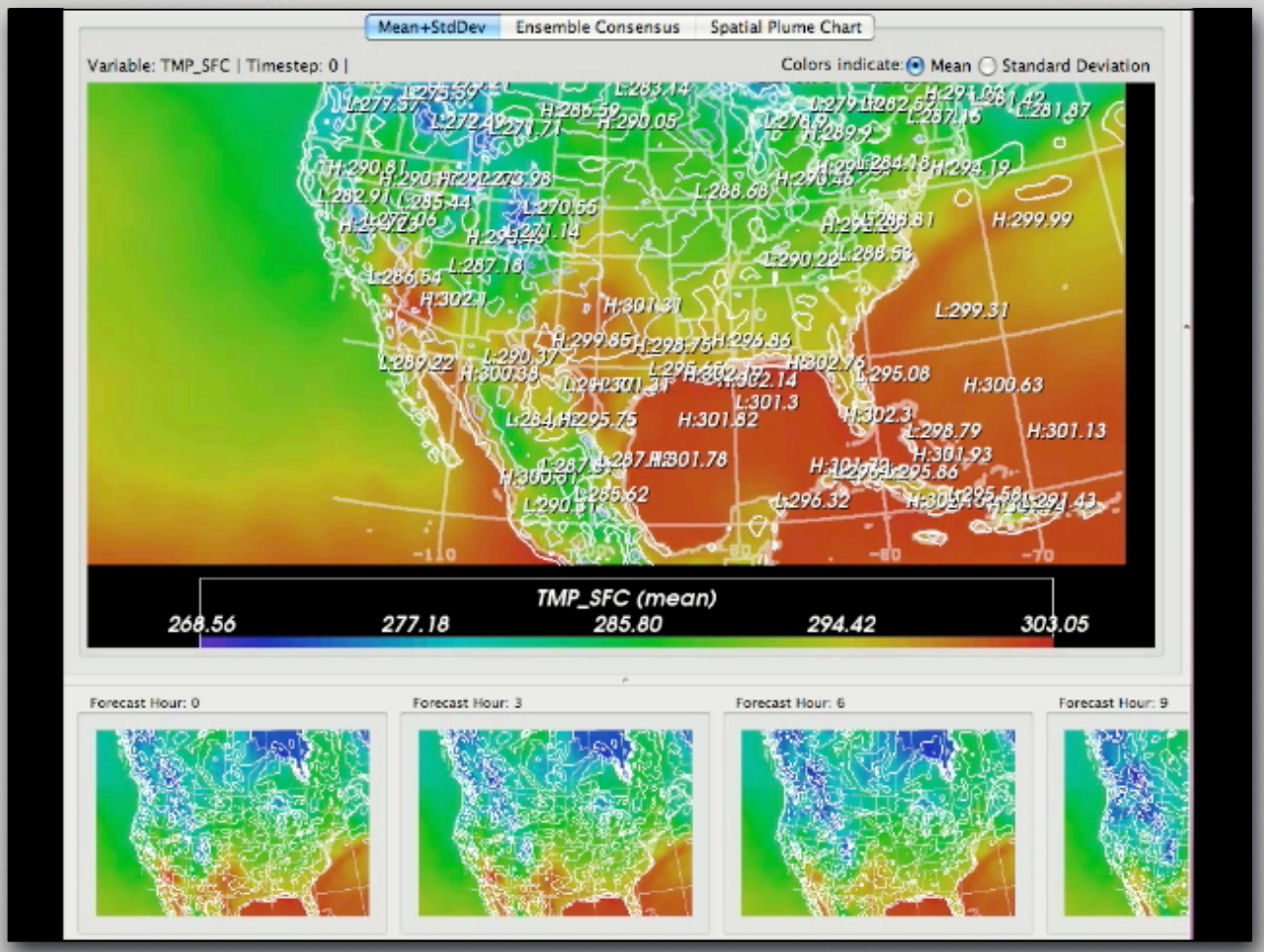
Statistical Summaries

- User selected region, boxplot type summary or summary across model types



Time Summary

- Filmstrip-style small multiples

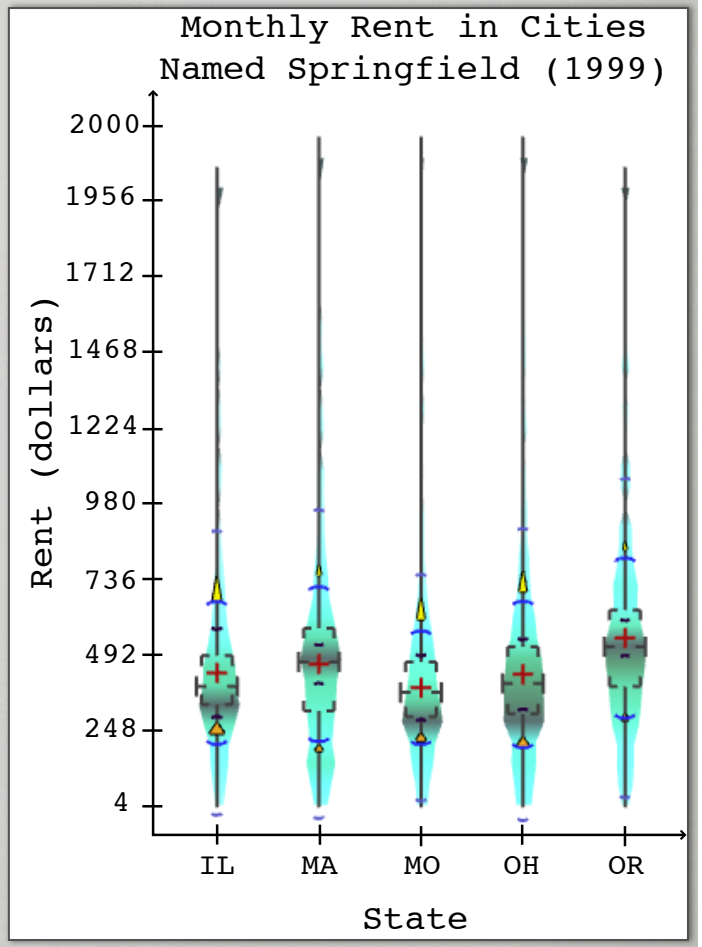


Ensemble Statistics

- Multiple stochastic models
- Spread of ensembles mitigates uncertainties in parameters, error in model output
- Ensemble data tolerates point-wise summaries
- Assume Gaussian distribution across points - can we do better?

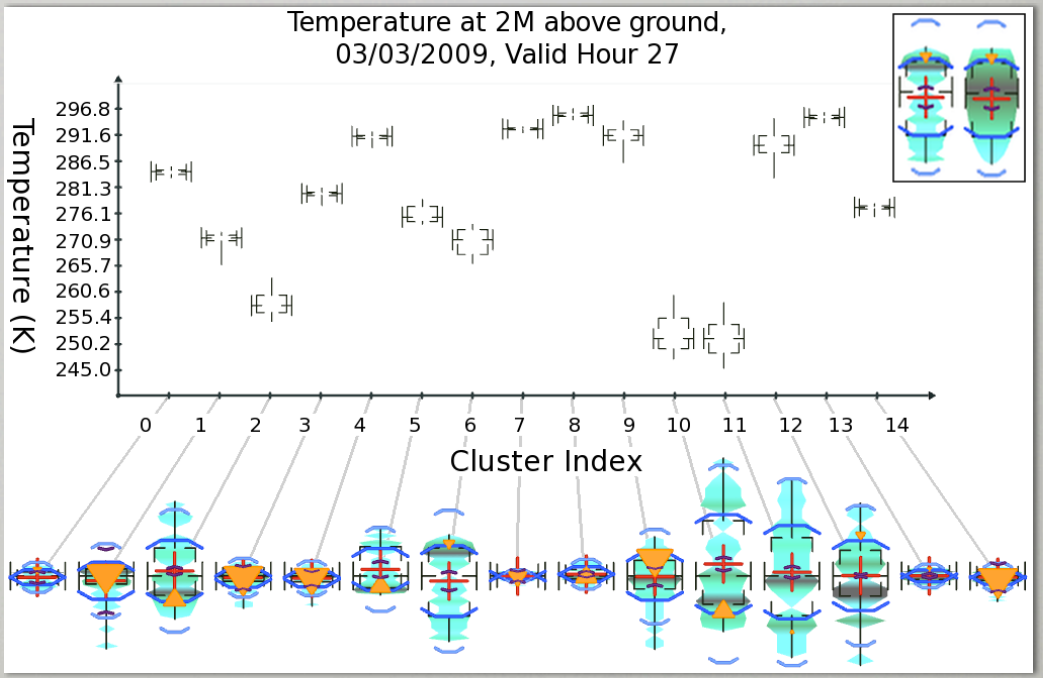
Example 2 - The Summary Plot

- Uncertainty of a single variable
- Extension of Tukey's Boxplot
- Often used in a collection of points

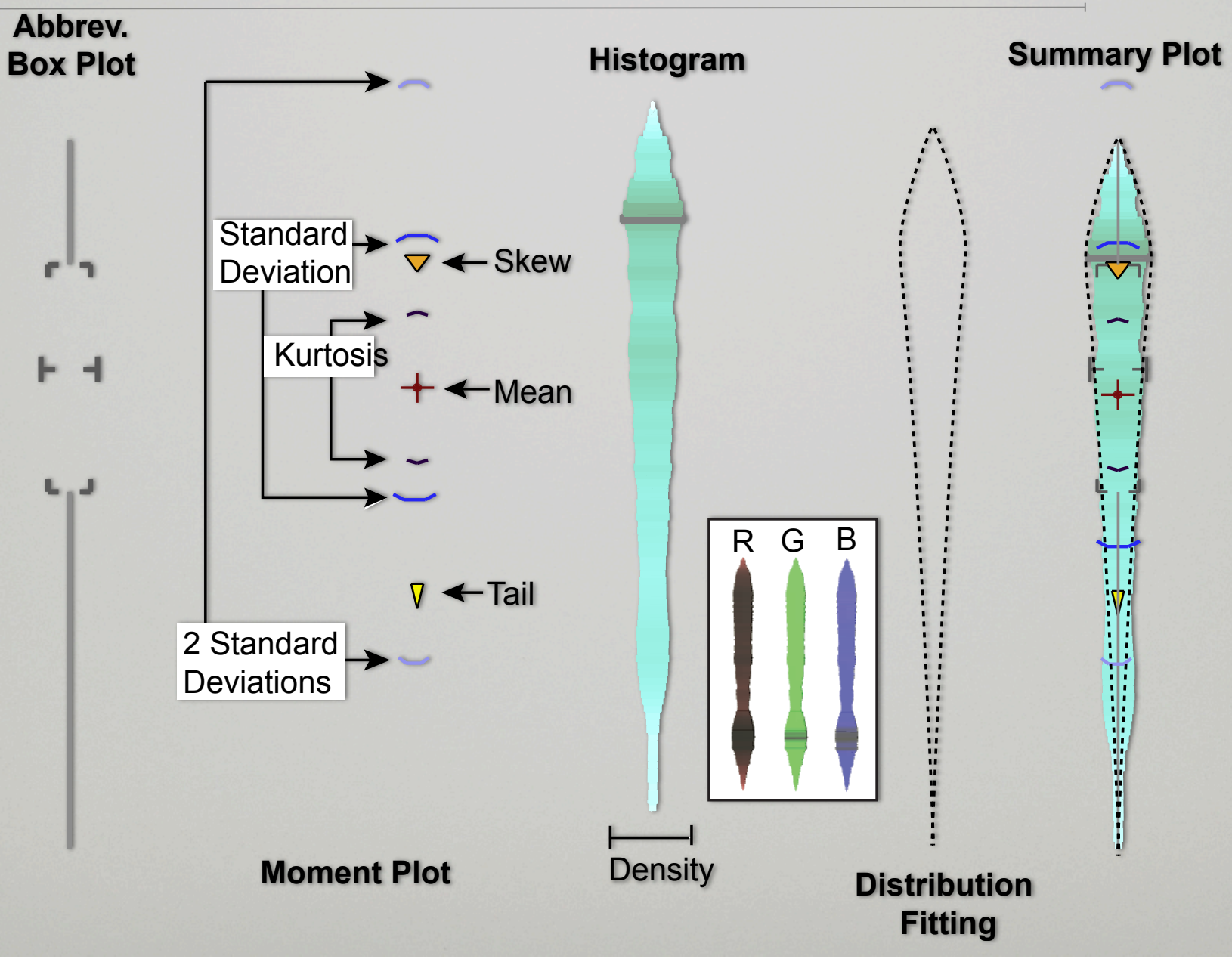


Highlights variance from normal

- Graphical representation of 1D PDF
- Overlap boxplot with more descriptive statistics
- Higher order moments



Anatomy of the Summary Plot

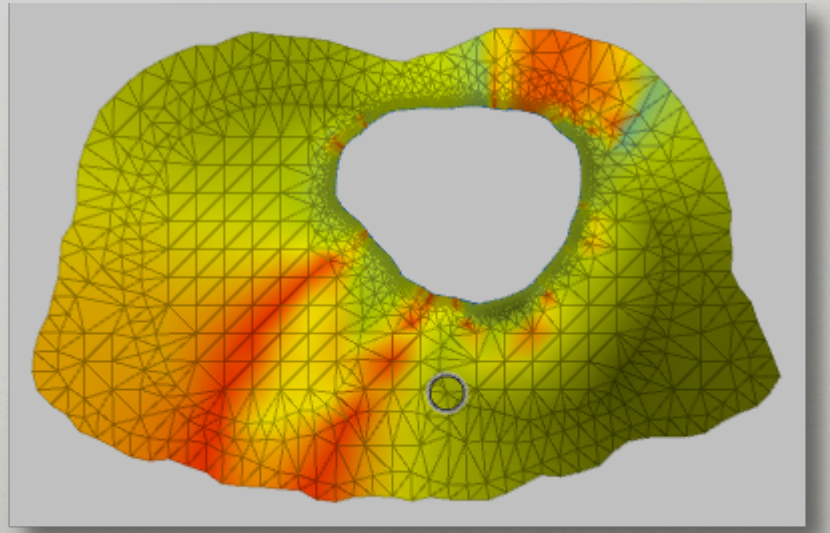


Back to higher dimensions

- Assumptions of normality not realistic
- Glyph-based approach not feasible globally in higher dimensions (clutter, obfuscation, etc)

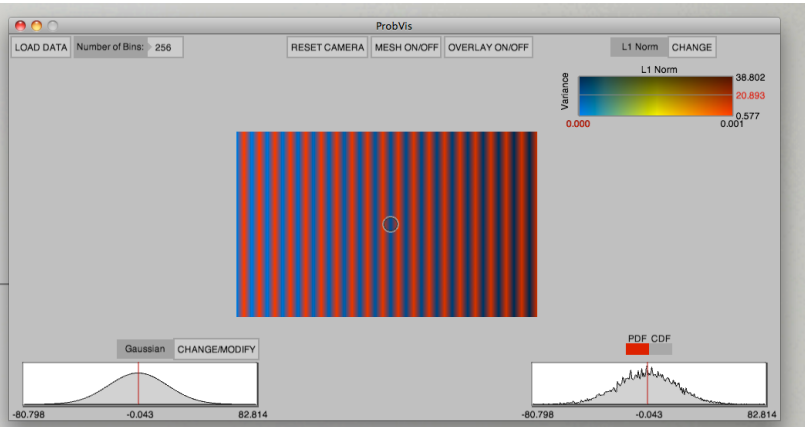
Example 3 - ProbVis

- Visualization and exploration tool for distributions defined across 2D domain
- Goal 1: to show differences *between* PDFS across domain
- Goal 2: summarize all data in a single view



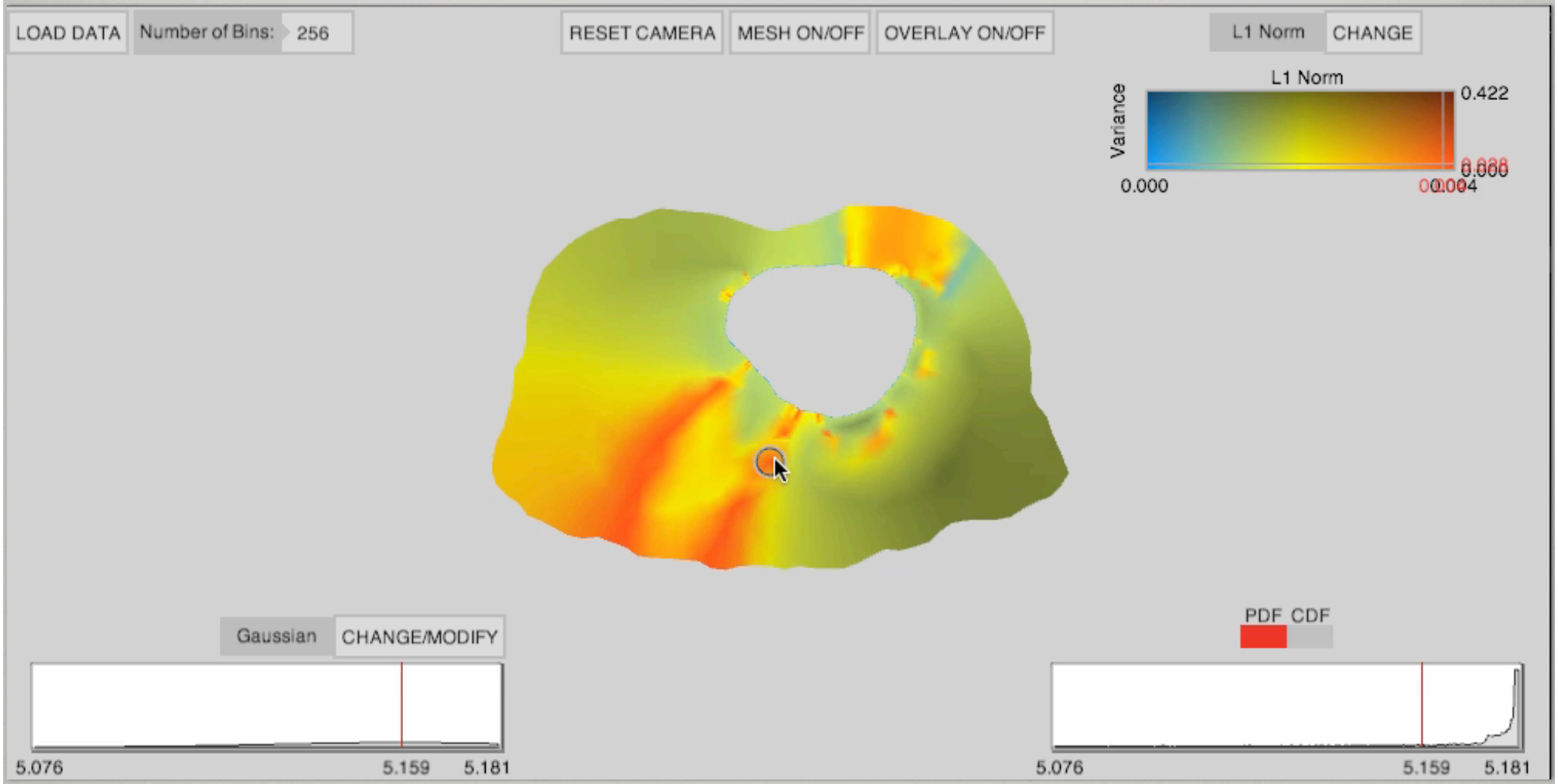
SCI INSTITUTE

Technique



- Define a difference measure between 2 PDFs
  - shape: L1 norm or Hellinger distance
  - interval: compare range of distribution
- Compare every data distribution to a defined cannonical PDF
- Use comparison as summarization

Interactivity for further exploration

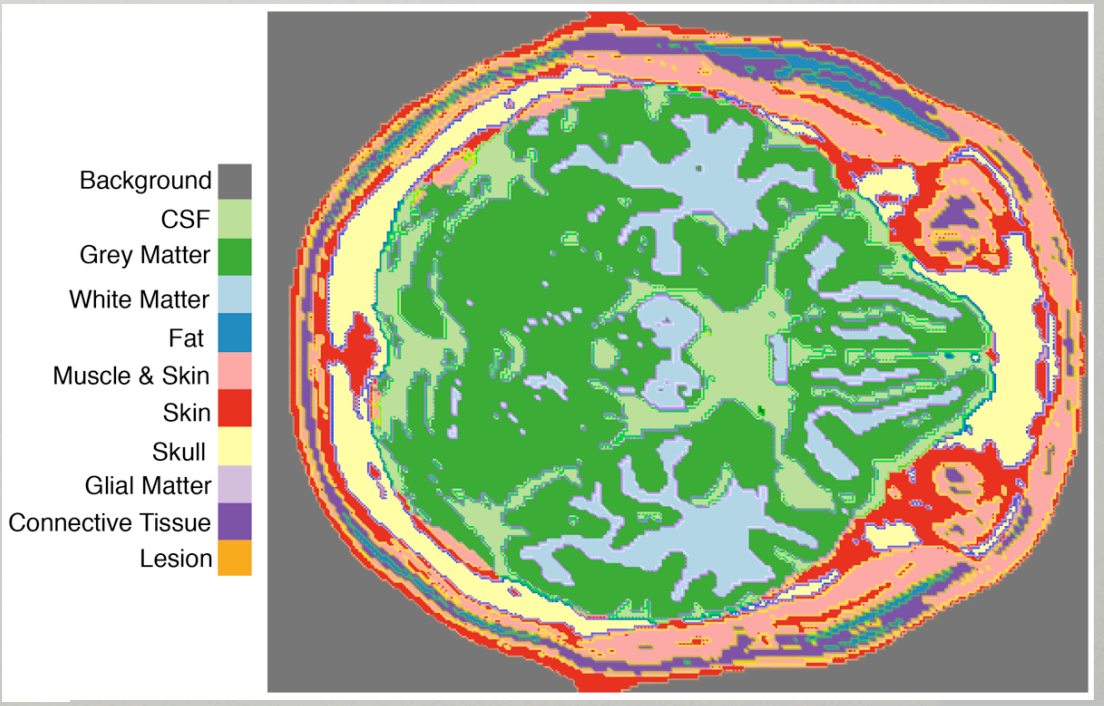


And 3D?

- Future work: methods to show PDFs in 3D
  - color mapping or glyphs probably not going to do it
- What about data that is not defined in this way?

Example 3 - Entropy for Categorical Data

- Probabilistic volumetric data
  - 3D volumes: 1 for each tissue type
  - Every voxel has probability of tissue membership
  - Fuzzy membership
  - mean is not defined
- How do we even look at this data?



Maximum Probabilities

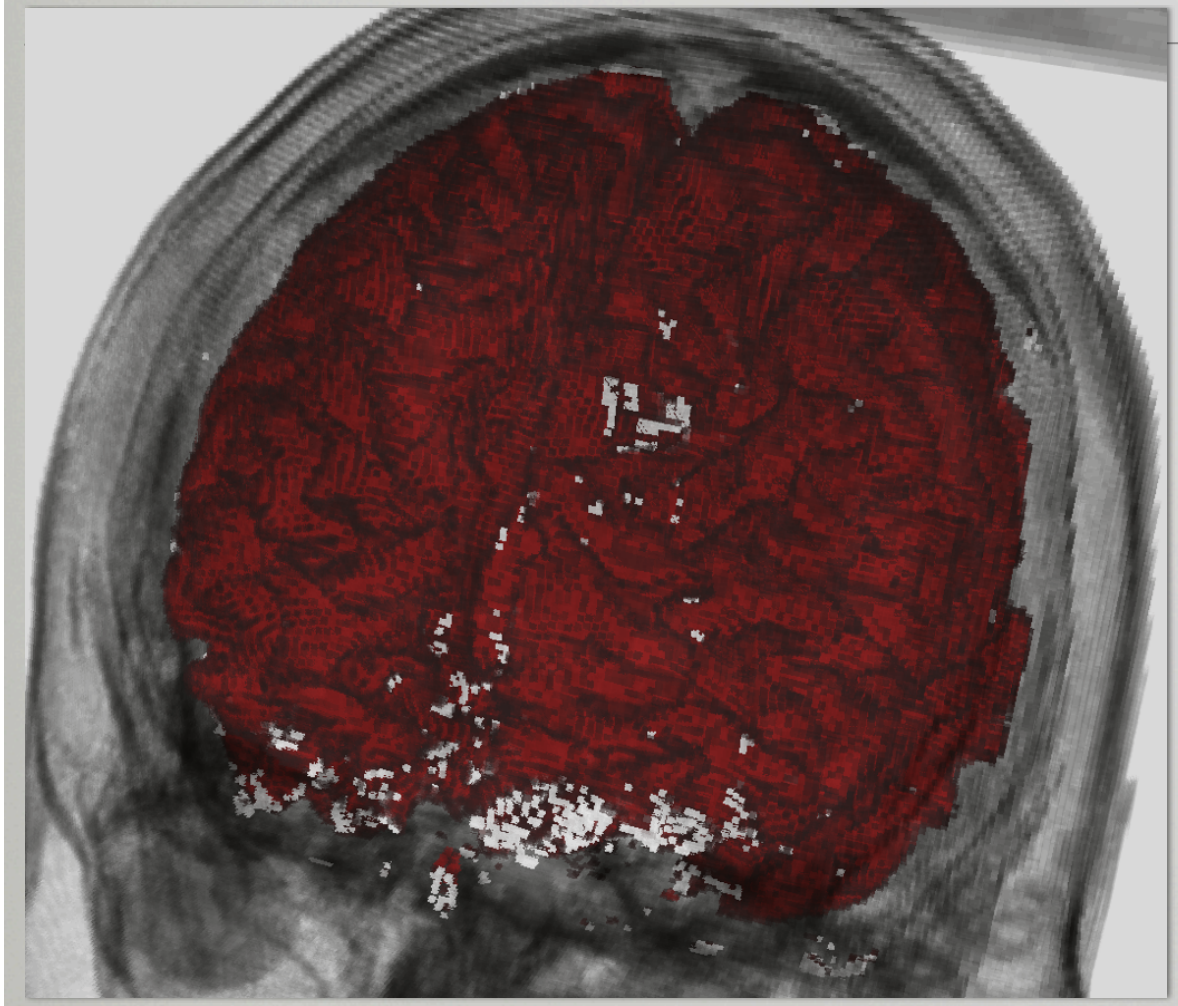
Draw from Information Theory

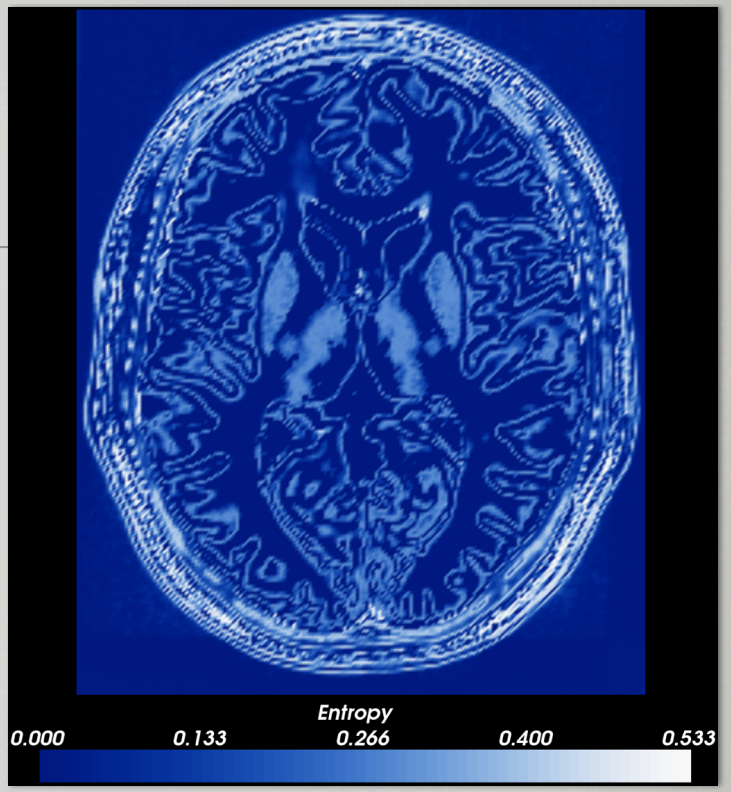
Shannon Entropy

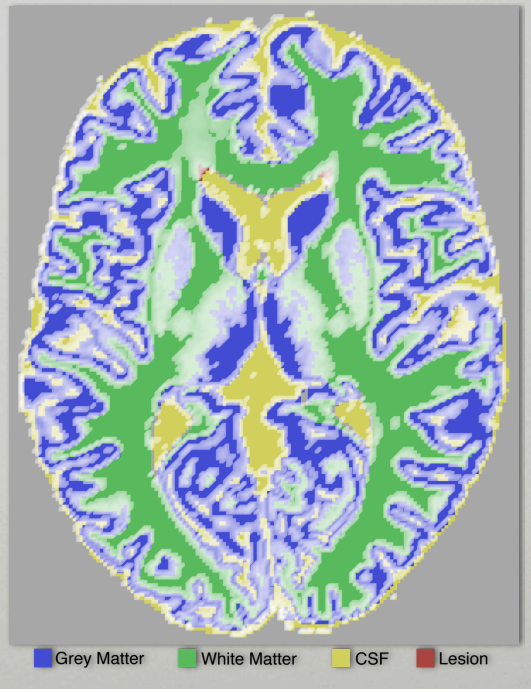
$$H[X] = -\sum_{i=1}^{n} p(x_i) \log(p(x_i))$$

- Measures how spread out a distribution is
- Higher entropy for voxels that contain lots of tissues or close probabilities
  - Example: distribution with
  - single outcome, p(i) = 1; entropy = 0
  - discrete uniform; p(i) = 1/n; entropy = -log(1/n)

Results







SCI INSTITUTE

Conclusion

- Uncertainty not a singular definition
- Multiple ways to summarize the data
  - really depends on application/questions
- Main focus of ongoing work:
  - are the current summary techniques we are using appropriate?
  - what are there other/better ways to summarize?
  - how do we ensure the chosen method is best?

Thanks!

kpotter@sci.utah.edu

Get directions | My places

bayou near Salt Lake City Marriott City Center

The Bayou
645 South State Street, Salt Lake City, UT 0.6 mi S
(801) 961-8400 · utahbayou.com
★★★★☆ 135 reviews $$

beer selection · sweet potato fries · cajun · live music · live jazz

""Bustling with young and old locals" this "big" "boisterous" Downtown" - zagat.com

Directions | Search nearby | Save to map | more

See all 82 results for **bayou**

Report a problem - Maps Labs - Help

Google Maps - ©2012 Google - Terms of Use