

# *Uncertainty Visualization: Beyond Mean and Standard Deviation*

---

## IAMCS-KAUST Workshop on Computational Biomedicine and Geophysics

Kristi Potter

Research Scientist, SCI

[kpotter@sci.utah.edu](mailto:kpotter@sci.utah.edu)

April 5, 2012

# 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

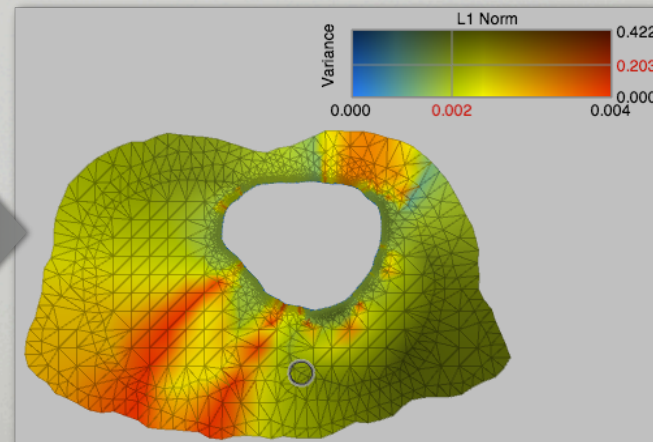
---

- 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

```
Emacs@gaz.sci.utah.edu
-6.4559780651 -7.6607013782 -9.3790748763 -10.3421795812
s-10.8270647789 -11.1665074719 -11.3995670294 -11.51834421
s9 -11.5694824605 -11.4542982469 -11.0780122149 -10.50388
e27453 -9.7671467727 -8.9807218646 -8.1262778940 -7.111238
e3505 -6.152717354 -5.0034190293 -4.0830276112 -3.1020405
e158 -2.1288803169 -1.2288347601 -0.9088405677 0.019594494
s8 0.4311850331 0.7328224082 0.9258162630 1.0707256313 1.1
e607251377 1.2582029828 1.3915989021 1.5247330567 1.691601
e3215 1.9422959373 2.2780582451 2.7043882540 3.1480563185
s3.5368901269 3.8829418530 4.1892839007 4.4347540838 4.6159
e9610941 4.7352624988 4.8001825442 4.8325974480 4.83631515
e47 4.8205537528 4.7725718624 4.6626207660 4.5242767134 4.
s3892340920 4.2660153543 4.1379762636 3.9815550032 3.80886
e09493 3.6296414839 3.4280202288 3.2184502865 3.0053080792
s 2.7935483143 2.6408722698 2.5210164656 2.4268317628 2.35
e05289342 2.2741311583 2.2061806675 2.1512136552 2.1088020
e746 2.0643977512 2.0072980995 1.9357713746 1.8413127627 1.
s.7407278585 1.5892314183 1.3841366106 1.1682484909 0.9578
e510338 0.7774204546 0.5585027396 0.3948303993 0.204794804
s 0.012357464 -0.1743347411 -0.3646198979 -0.5019038342
s-0.7694706022 -0.9993201352 -1.2467333865 -1.5068453860 -3
s1.7878594805 -2.1562552080 -2.5666494765 -3.0320663129 -3
s.7679203565 -4.934811080 -6.5043084330 -7.6718840712 -8.
e0433940683 -10.000599325 -10.9604917705 -11.9053084857 -
s11.9939002098 -11.4774257620 -11.6173115624 -11.703377404
s6 -11.4857172057 -11.0325139210 -10.5436766195 -9.7716268
-u:** *scratch* Top L1 (Lisp Interaction)-----
```



# Challenges to Uncertainty Visualization

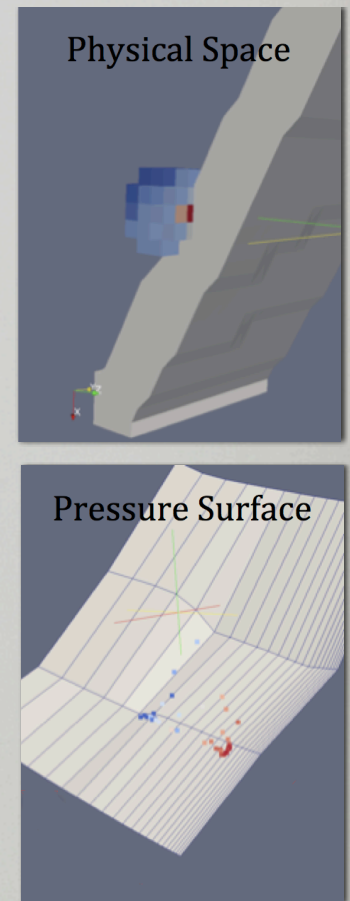
---

- 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

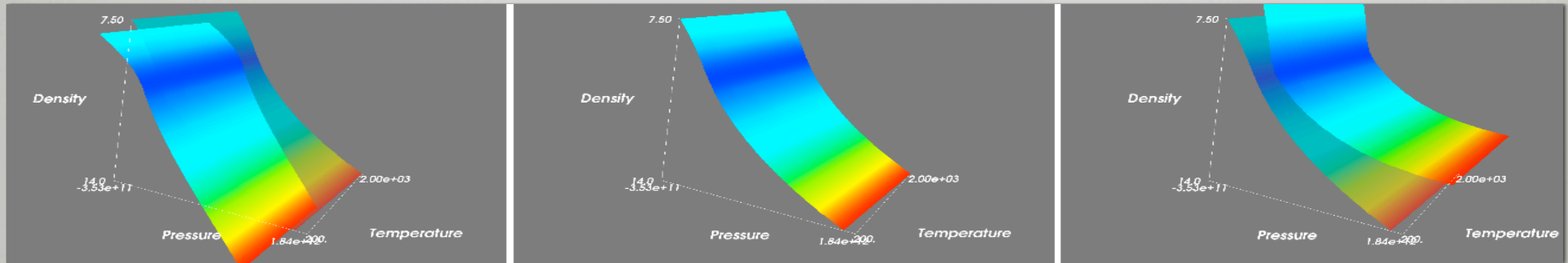
# An Example Application - Material Models

---

- 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?



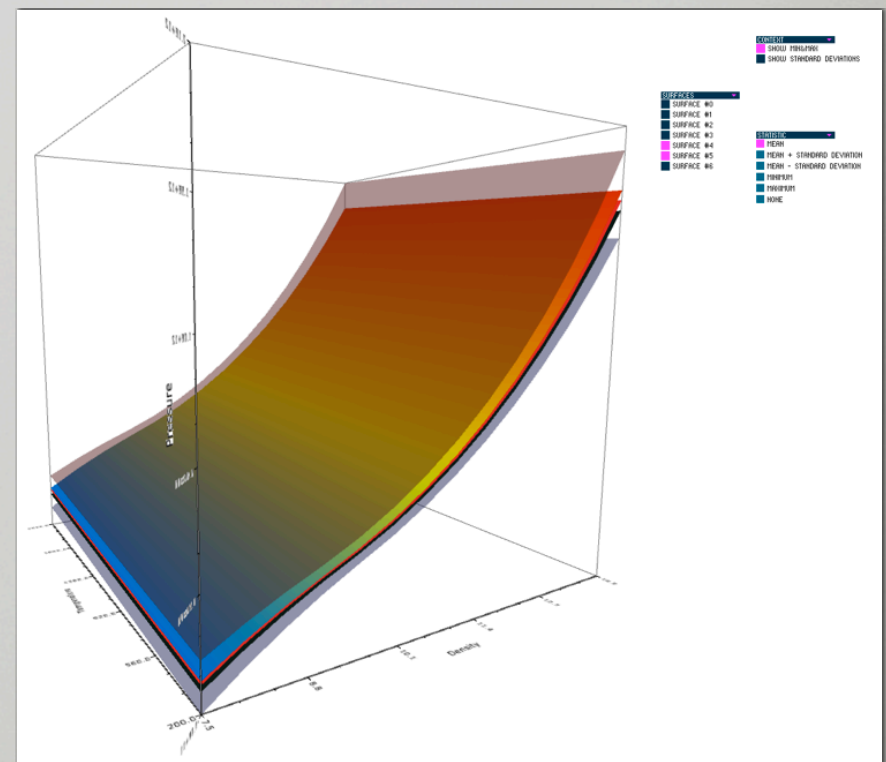
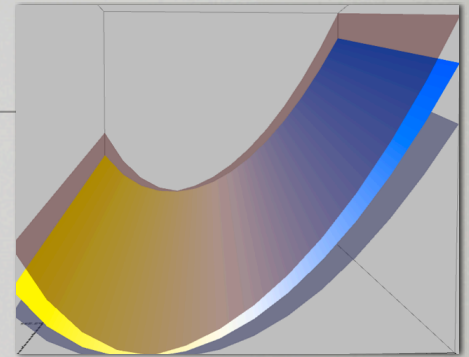
# Visualization of EOS Surfaces



- ▶ 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

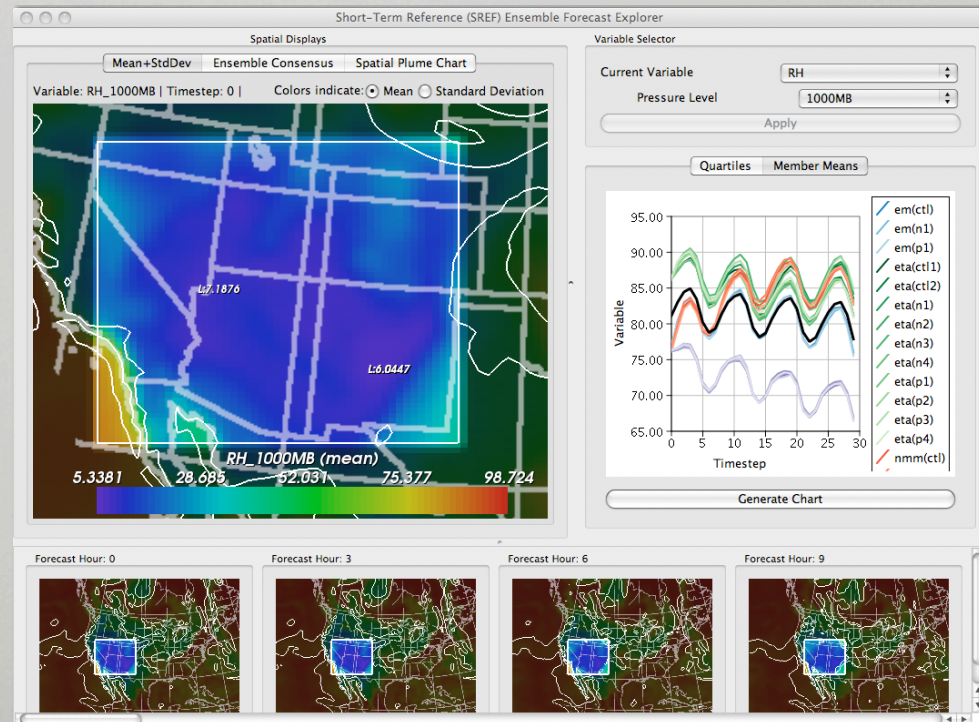
---

Reduce large amounts of data using *summaries*

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

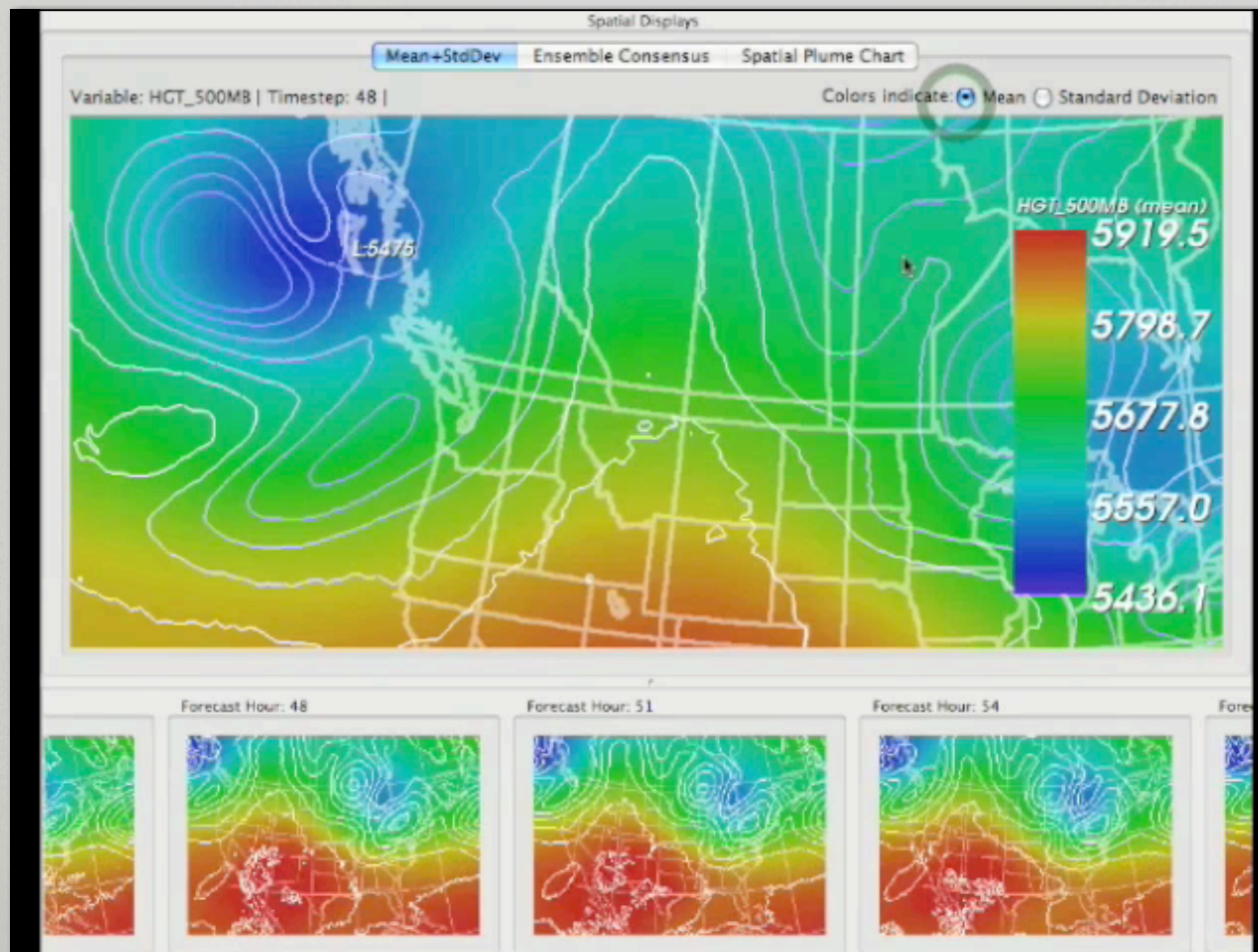
# Example 1 - Ensemble-Vis

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



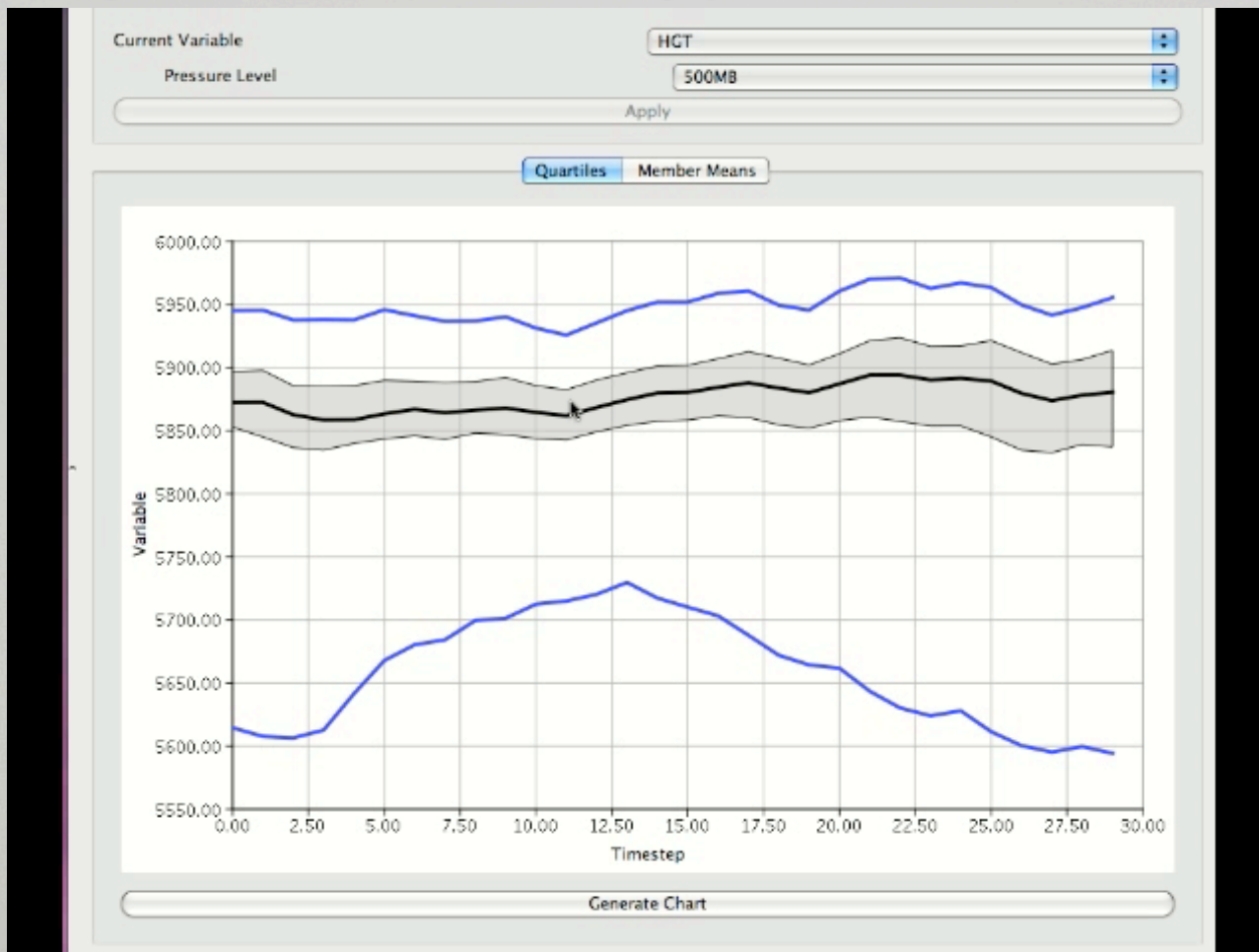
# Grid Point Summaries

- Mean & standard deviation at each grid point



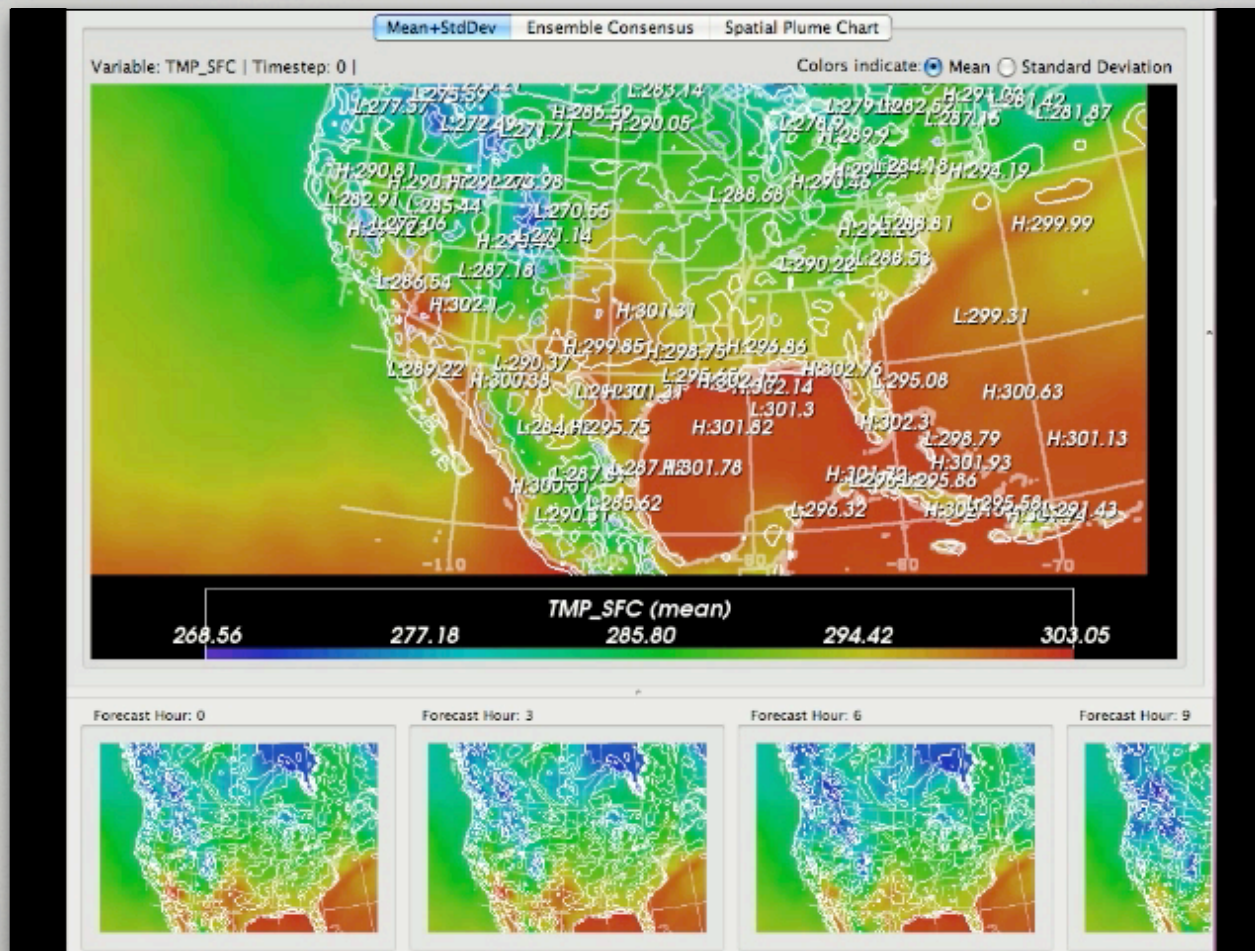
# Statistical Summaries

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



# Time Summary

- ▶ Filmstrip-style small multiples



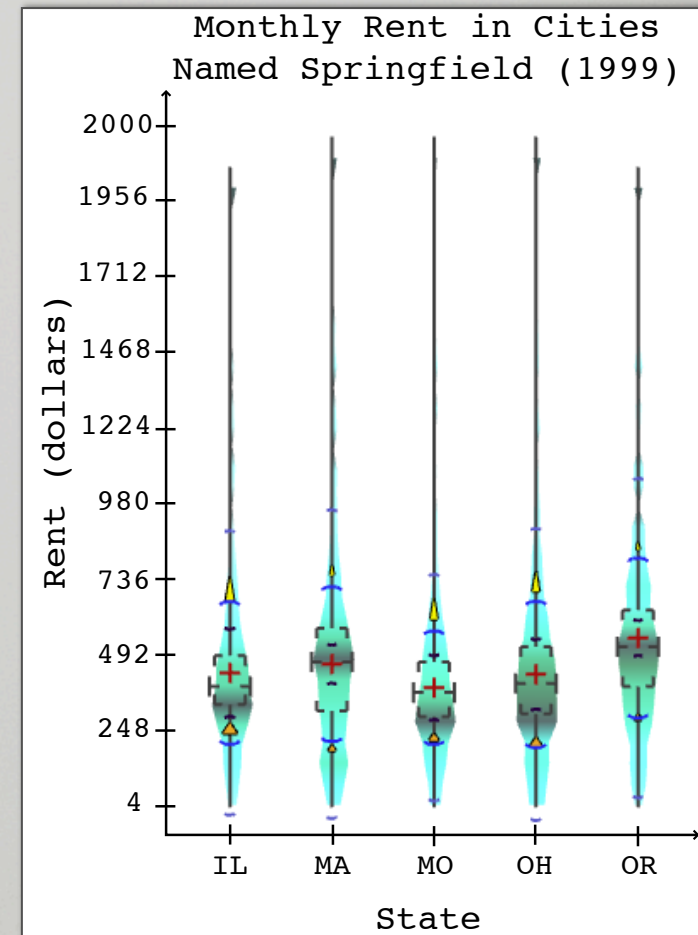
# Ensemble Statistics

---

- 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?

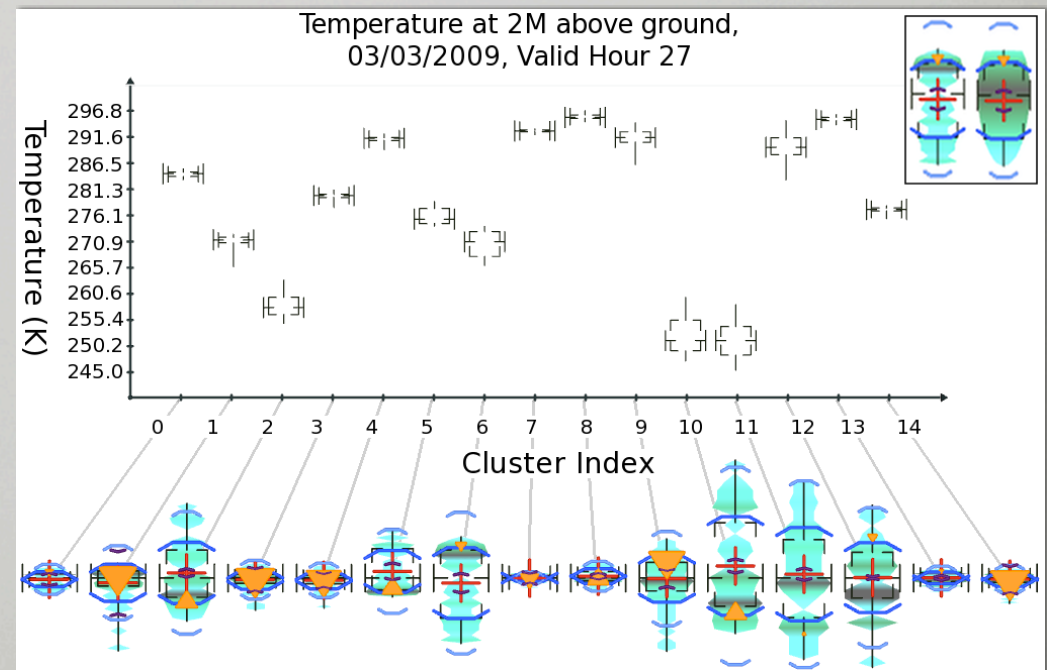
# Example 2 - The Summary Plot

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



# Highlights variance from normal

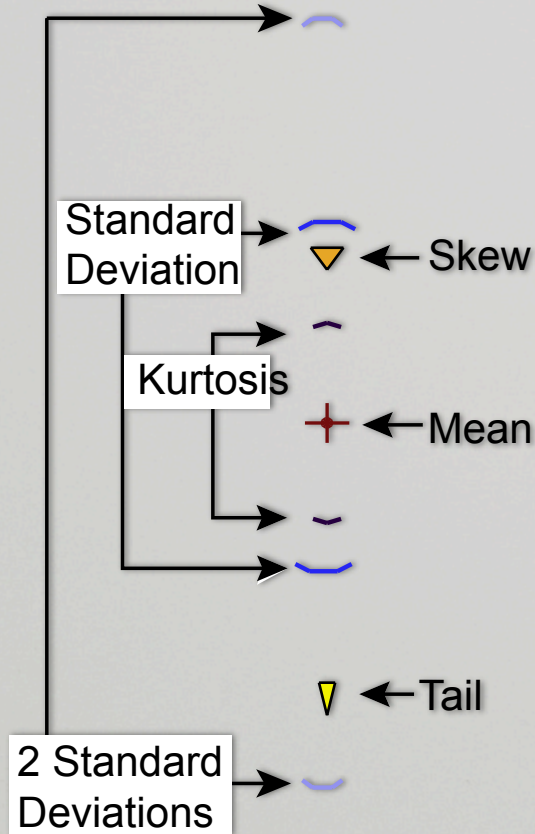
- ▶ Graphical representation of 1D PDF
- ▶ Overlap boxplot with more descriptive statistics
- ▶ Higher order moments





# Anatomy of the Summary Plot

Abbrev.  
Box Plot

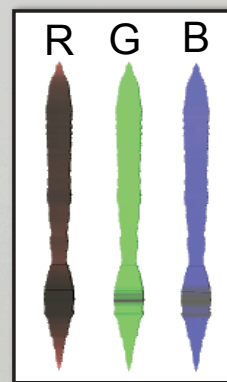


Moment Plot

Histogram



Density



Distribution Fitting

Summary Plot



# Back to higher dimensions

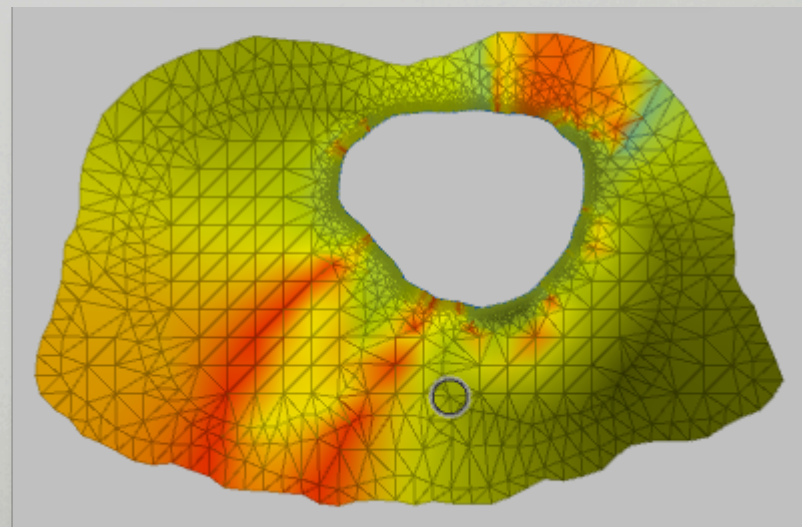
---

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

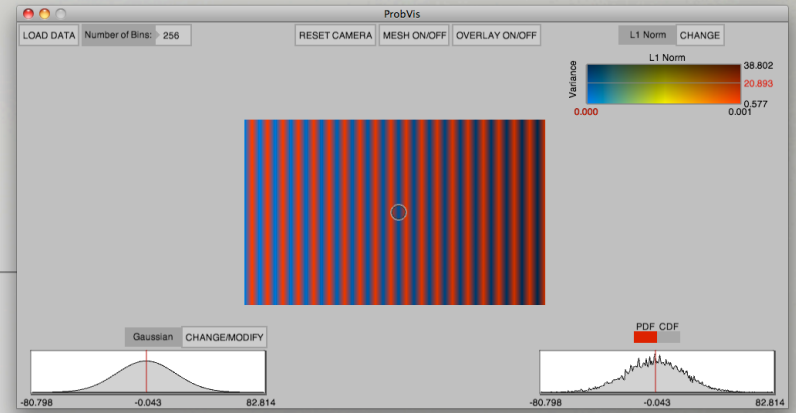
# Example 3 - ProbVis

---

- 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



# 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 canonical PDF
- Use comparison as summarization

# Interactivity for further exploration

LOAD DATA Number of Bins: 256

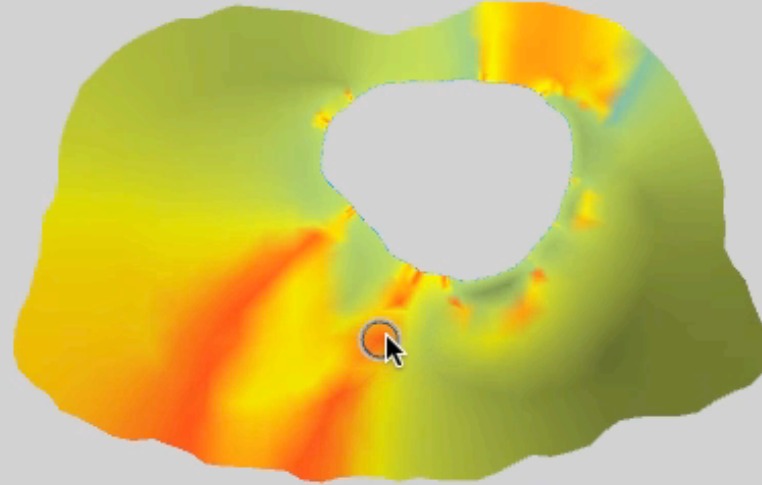
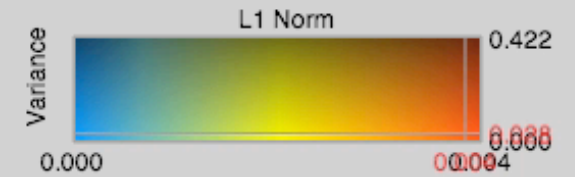
RESET CAMERA

MESH ON/OFF

OVERLAY ON/OFF

L1 Norm

CHANGE



Gaussian

CHANGE/MODIFY



PDF CDF



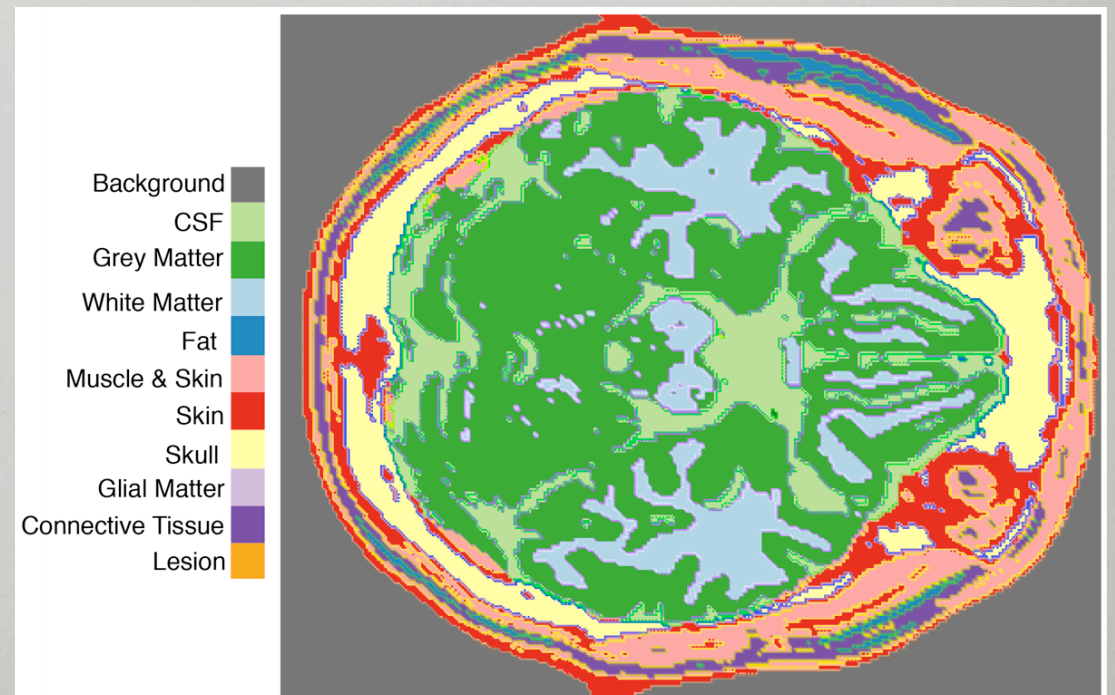
# And 3D?

---

- 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

---

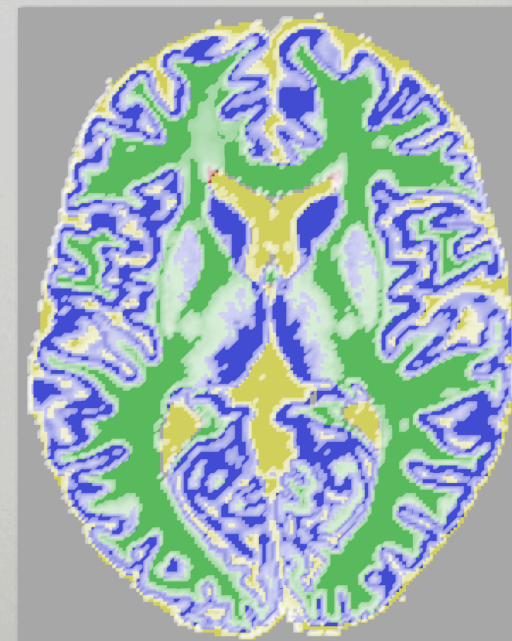
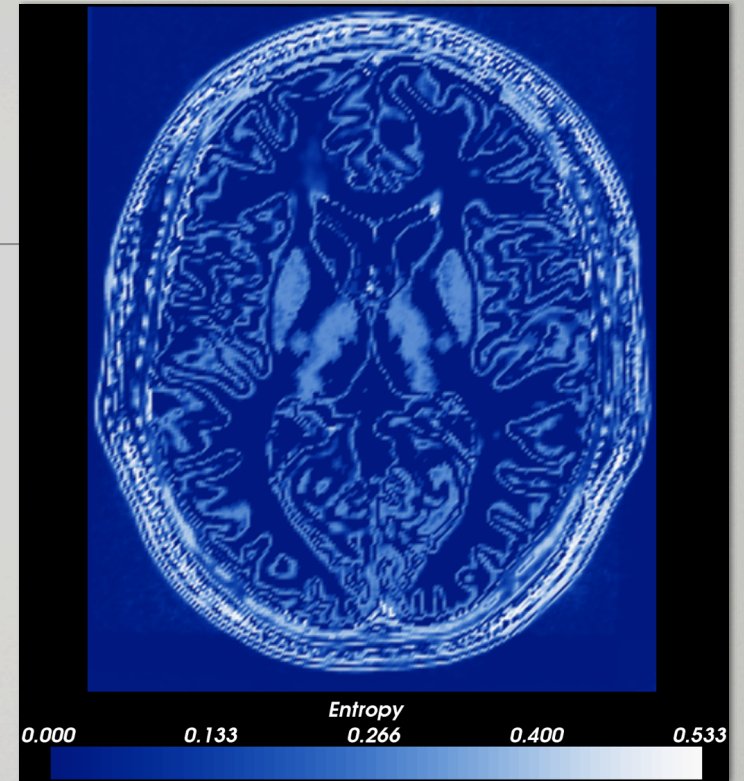
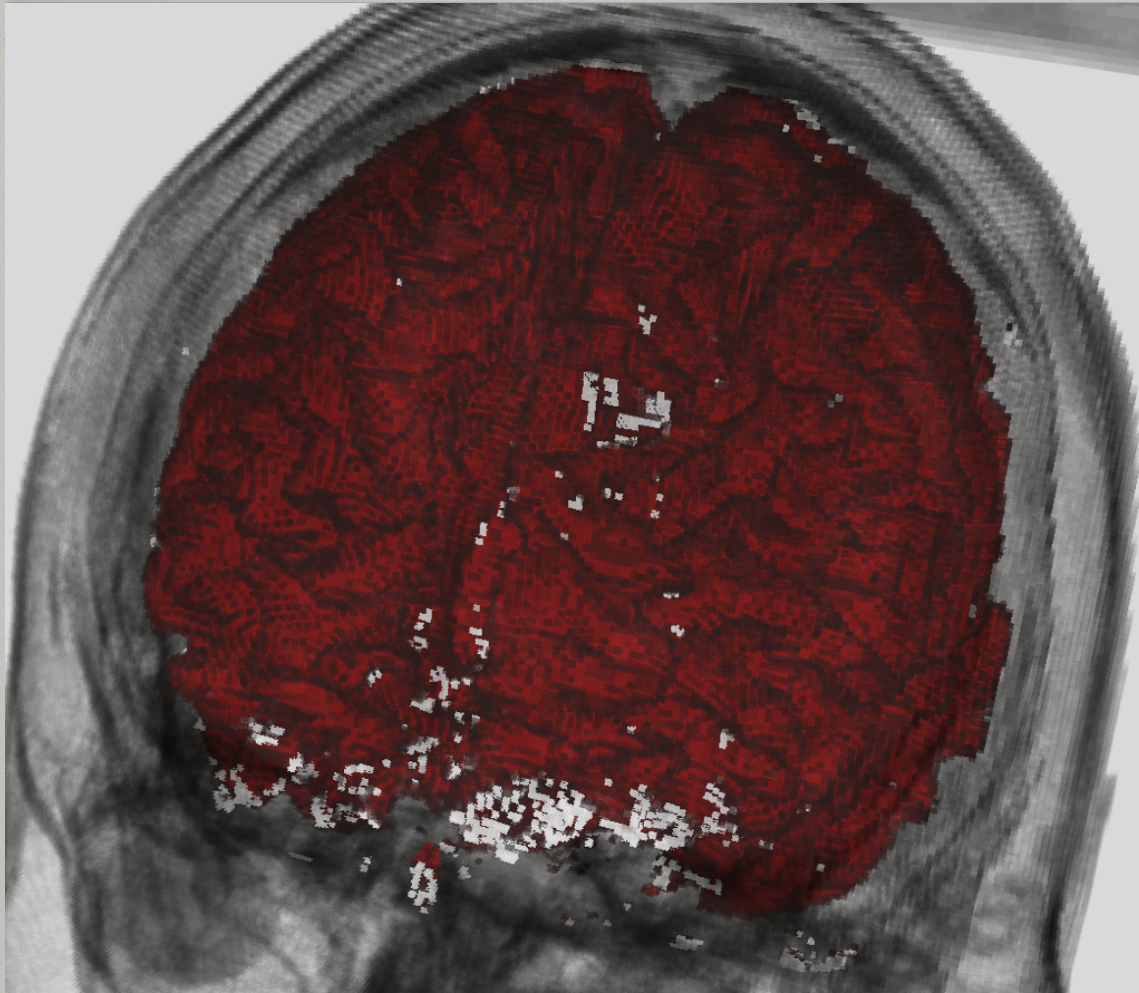
## Shannon Entropy

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

- 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



■ Grey Matter ■ White Matter ■ CSF ■ Lesion

# 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](mailto: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](http://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

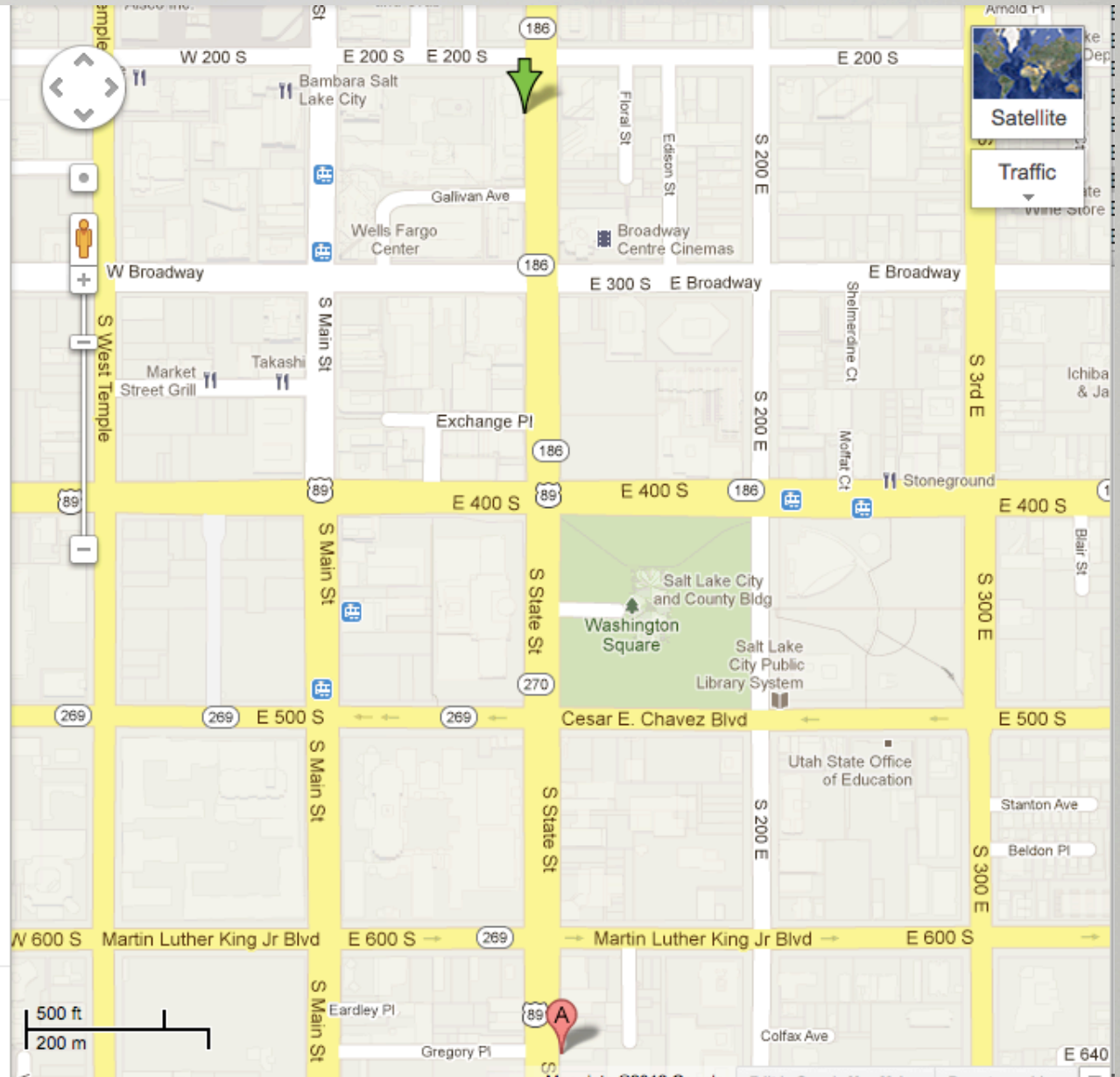


[Directions](#) [Search nearby](#) [Save to map](#) [more](#)

[See all 82 results for bayou](#)

[Report a problem](#) - [Maps Labs](#) - [Help](#)

Google Maps - ©2012 Google - [Terms of Use](#)



Satellite

Traffic



Map data ©2012 Google

Earth data ©2012 Google

Street View Imagery ©2012 Google

Icons ©2012 Google

Labels ©2012 Google

Partners ©2012 Google

Terms of Use

Privacy Policy

Feedback

Map data ©2012 Google

Earth data ©2012 Google

Street View Imagery ©2012 Google

Icons ©2012 Google

Labels ©2012 Google

Partners ©2012 Google

Terms of Use

Privacy Policy

Feedback

Map data ©2012 Google

Earth data ©2012 Google

Street View Imagery ©2012 Google

Icons ©2012 Google

Labels ©2012 Google

Partners ©2012 Google

Terms of Use

Privacy Policy

Feedback

Map data ©2012 Google

Earth data ©2012 Google

Street View Imagery ©2012 Google

Icons ©2012 Google

Labels ©2012 Google

Partners ©2012 Google

Terms of Use

Privacy Policy

Feedback

