Uncertainty and Parameter Space Analysis in Visualization | Oct 15, 2012

STATISTICAL UNCERTAINTY

From Quantification to Visualization

Kristi Potter
Scientific Computing and Imaging Institute
University of Utah

VisWeek 2012 Tutorial

Statistical Uncertainties

Random fluctuations of measurement

Sources

PARAMETER UNCERTAINTY
- *input parameters to models*
- *exact values unknown, cannot be controlled by experiment*

Sources

MODEL INADEQUACY
- *aka model bias, model discrepancy*
- *lack of knowledge of the underlying problem*
- *accuracy of the model to reality*
- *result of model being an approximation*

Sources

ALGORITHMIC UNCERTAINTY
- *aka numerical uncertainty*
- *numerical errors, approximations*
- *translation of mathematical model to the computer*

Sources

EXPERIMENTAL UNCERTAINTY
- *aka observational error*
- *variability of experimental measurements*

Sources

INTERPOLATION UNCERTAINTY

- *lack of available data from simulation/experiment*
- *must interpolate/extrapolate for desired response*

Categories

EPISTEMIC UNCERTAINTY

- *aka* ***systematic*** *uncertainty*
- *things we could in principle know but don't in practice*
- *i.e. insufficient measurement or modeling, missing data*

REDUCIBLE: *can be alleviated by better models, more accurate measure*

Categories

ALEATORIC UNCERTAINTY

- *aka* ***statistical*** *uncertainty*
- *unknowns that differ on each run*
- *i.e. throwing dice*

Irreducible: *cannot removed/suppressed through improvements in models or measurements*

Categories

Intention of uncertainty quantification

reduce epistemic uncertainties to aleatoric

relatively straightforward quantification

generally how we think of uncertainty in vis

Statistical Uncertainties

Representations for visualization

Representations

RANGE

- *outcome known to lie within some interval*

Interval Arithmetic

set bound on errors

define operations on intervals

Representations

Probability Distribution Functions (PDFs)

• *approximate outcome through a probability function*

Probability Density
continuous random variables
repetition of outcome values

Representations

Probability Distribution Functions (PDFs)

• *approximate outcome through a probability function*

Categorical Distribution
discrete random variables
finite set of values

Representations

Probability Distribution Functions (PDFs)

measures on PDFs

Mean
expected value, arithmetic mean

Representations

Probability Distribution Functions (PDFs)

measures on PDFs

Standard Deviation
spread of values

Representations

Probability Distribution Functions (PDFs)

measures on PDFs

Mode
most frequently occurring value

Representations

Probability Distribution Functions (PDFs)

measures on PDFs

Tail
region of least frequently occurring values

Representations

Probability Distribution Functions (PDFs)

measures on PDFs

Support

interval where value probability is not zero

Data

Ensembles

Multi-Run Simulations

- *explore space of parameters*
- *mitigate model error*
- *cover range of initial conditions / outcomes*
- *combine multiple models*

Ensembles

Collection of Datasets

- *members, realizations*
- *full simulation run for each parameter set/input condition*

Ensembles

Multi-

- *dimensional*

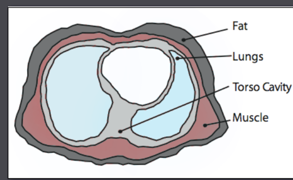
  spatial domain & time
- *variate*

  simulate over many variables
- *valued*

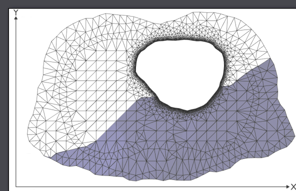
  many vales for each variable/location

Two Datasets

Electrical Conductivity of the Heart

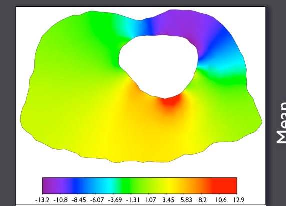


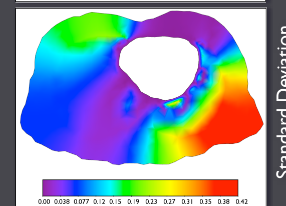
Electrocardiogram
- *Simulate how signals from the heart propagate across the torso*
- *Distinguish normal changes (breathing, movement) from abnormal heart function*



Potentials Data

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





Weather Forecasting

Short-Range Ensemble Forecasts (SREF)

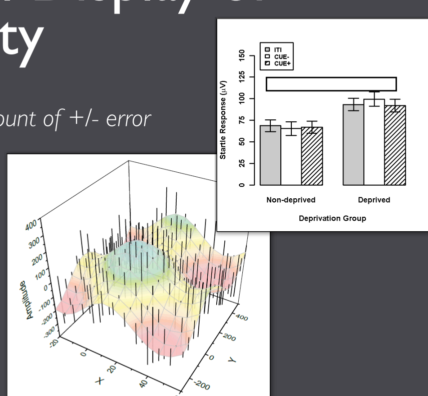
NOAA / NCEP
- *Domain across North America*
- *Forecast weather variables out to ~3.5 days*

Goal
- *Public notification, warnings, aviation*

Graphical Data Analysis

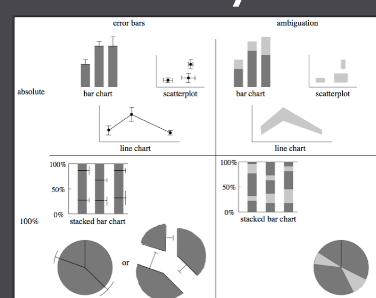
Traditional Display of Uncertainty

Error bars
- *convey accuracy by amount of +/- error*
- *std dev or std error*



Bounded Uncertainty

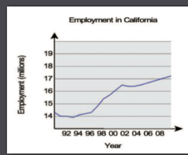
- *numeric interval guaranteed to contain data value*
- *no assumptions about the pdf within the interval*



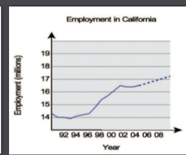
Statistical Uncertainty   Bounded Uncertainty

Visualizing Data with Bounded Uncertainty.
C. Olston, J.D. Mackinlay. InfoVis, 2002.

Information Uncertainty

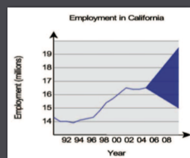


Average Growth

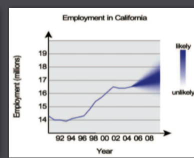


Estimated Growth

- *indication of how much confidence*
- *qualitative rather than quantitative*
- *spreadsheet interface characterizes the data*



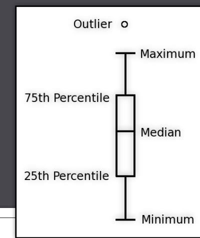
Possible Growth



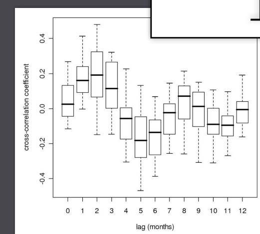
Likely Growth

A Spreadsheet Approach to Facilitate Visualization of Uncertainty in Information.
A. Streit, B. Pham, R. Brown. TVCG 14(1), 2008.

Traditional Display of Uncertainty

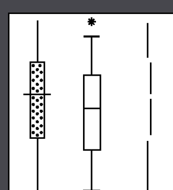


Boxplots
- *quartile range including median,outliers*
- *assume Gaussian*



Boxplot Modifications

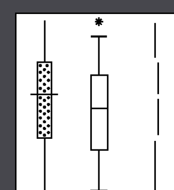
VISUAL MODIFICATIONS
refinement for aesthetic purposes


range bar

Charting Statistics.
Mary Eleanor Spear. McGraw-Hill, 1952.

Boxplot Modifications

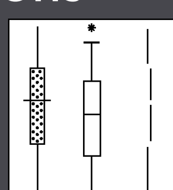
VISUAL MODIFICATIONS
refinement for aesthetic purposes


box plot

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

Boxplot Modifications

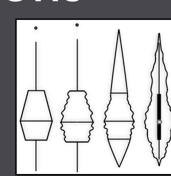
VISUAL MODIFICATIONS
refinement for aesthetic purposes


inner-quartile plot

The Visual Display of Quantitative Information.
Edward Tufte. Graphics Press, 1983.

Boxplot Modifications

DENSITY MODIFICATIONS
add indication of value prevalence

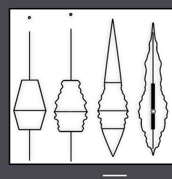


Opening the Box of a Boxplot.
Y. Benjamini. The American Statistician, 42(4), 1988.

Boxplot Modifications

Density Modifications

add indication of value prevalence

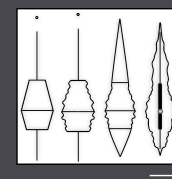


The Box-Percentile Plot.
W. Esty, J. Banfield. Journal of Statistics Software, 8(17), 2003.

Boxplot Modifications

Density Modifications

add indication of value prevalence

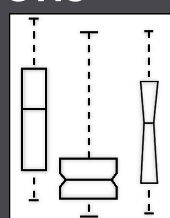


Violin Plots.
J. Hintze, R. Nelson. The American Statistician, 52(2), 1998.

Boxplot Modifications

Data Characteristics

sample size, confidence levels

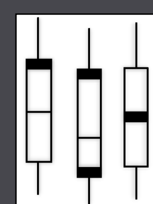


Variations of Box Plots.
R. McGill, J.W. Tukey, W.A Larsen. The American Statistician, 32(1), 1978.

Boxplot Modifications

Additional Statistics

· *moments, modality*



Can the Box Plot Be Improved?
C. Choonpradub, D. McNeil. Songklanakarin J Sci Technol, 27(3), 2005,

Boxplot Modifications

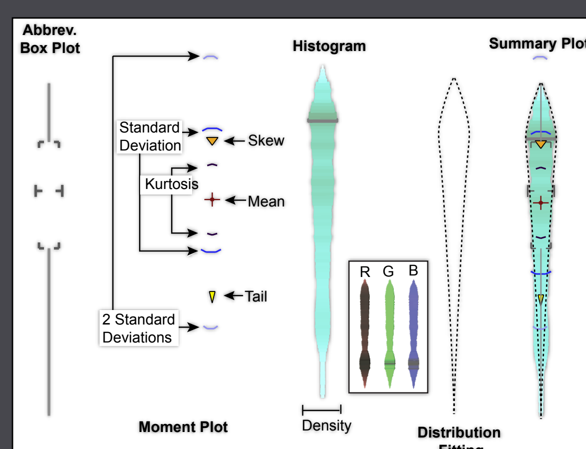
Summary Plot

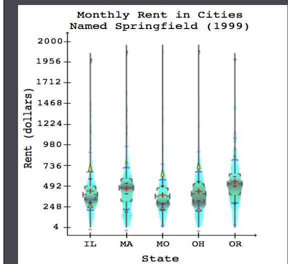
combine 4 plots into one

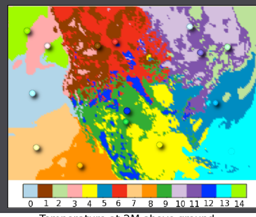
augment with more descriptive statistics

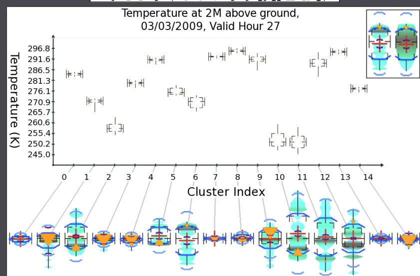
indicate quantity & location of uncertainty

Visualizing Summary Statistics and Uncertainty.
K. Potter, J. Kniss, R. Riesenfeld, C.R. Johnson. CGF 29(3), 2010.





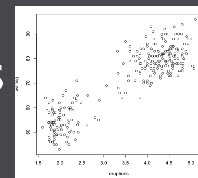




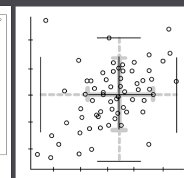
S. Becketti, W. Gould, Rangefinder box plots. TAS 41(2), 1987.

2D Box Plots

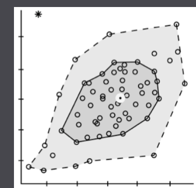
Scatterplots
•2D position of samples
RangeFinder Plot
•1D boxplot per axis
Two-Dimensional Boxplot
•Robust line partition
Bagplot
•Halfspace depth (spatial quartiles)



Scatterplot

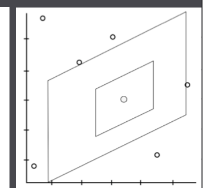


Rangefinder Plot



Two-Dimensional Boxplot

P. Tongkumchum, Two-dimensional box plot. Songklanakarin J Sci Technol, 27(4),2005.



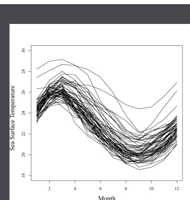
Bagplot

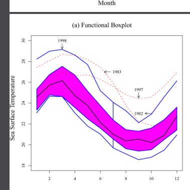
P.J. Rousseeuw, I. Ruts, J. Tukey. The bagplot: A bivariate boxplot. TAS, 53(4), 1999.

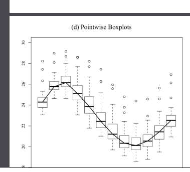
Functional Box Plot

Box Plot percentiles on 2D functions
• compute the amount of time a function lies within the full set of functions
• band-depth used to order data
• highest band-depth is median
• indicate envelopes for central region, maximum, non-outlying region
• mark outliers

Ying Sun, Marc G. Genton.
Functional Boxplots.
Journal of Computational and Graphical Statistics 20:2, 316-334.

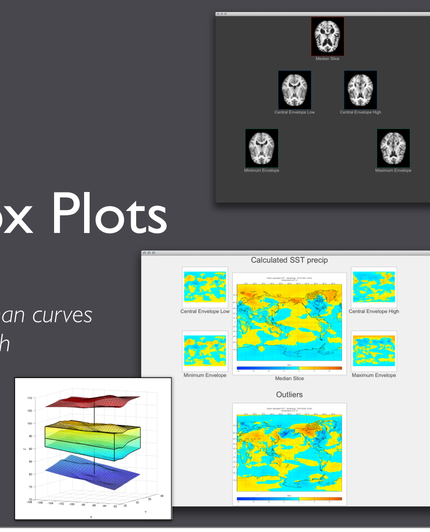






Surface Box Plots

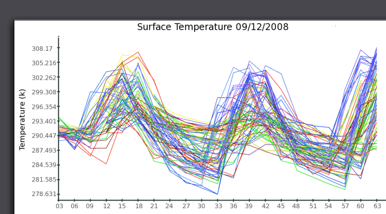
Extension to 3D
• used for images rather than curves
• volume-based band-depth



Visualization

Challenges

Increased complexity
• Visual clutter
• Data concealment
• Confusion



Information-Seeking Mantra

"Overview first, then zoom and filter, and finally, details on demand."

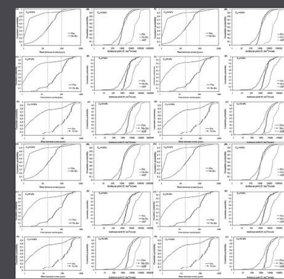
-Ben Shneiderman

Overview → Summary → Details

Overview vs Summary

Overview

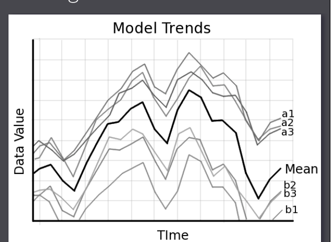
- *Show all data at once through many charts*
- *Manual search for patterns*
- *Finite screen resolution*



Overview vs Summary

Summary

- *Aggregate data along some dimension*
- *Automation through statistics/machine learning*
- *Have an idea of the questions*



Aggregation for Visualization

Reduce Data Using Summaries

- *mimic human visual system*
- *done implicitly*

phenomena modeling
floating pt quantization
limited # of pixels

Aggregation for Visualization

But How?

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

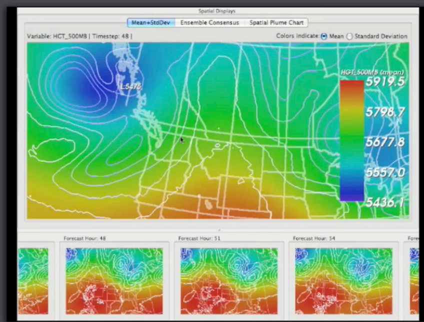
Global Summaries

Questions:

- *What is the average model temperature at a given time step?*
- *Where do the models vary?*

Global Summaries

- Aggregate over models at each grid point
- Colormap & contour
- Show one timestep



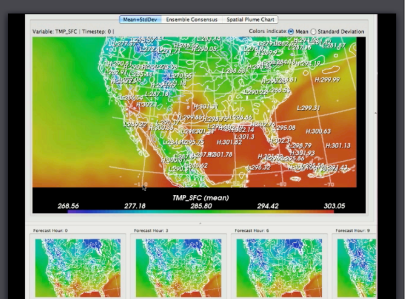
Time Summary

Questions:

- How does the temperature change across time?
- What time step am I most interested in?

Time Summary

- Small multiples of overview summary (low resolution)
- User interaction to scroll through time
- Select step of interest, reflect to overview



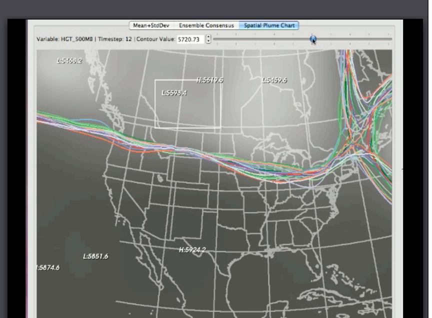
Contour Summaries

Questions:

- Where does a particular value exist for each model?
- Where does that value move across time?

Contour Summaries

- Isocontour of value across spatial domain for each model
- Model bundle shows variation, outliers, divergence

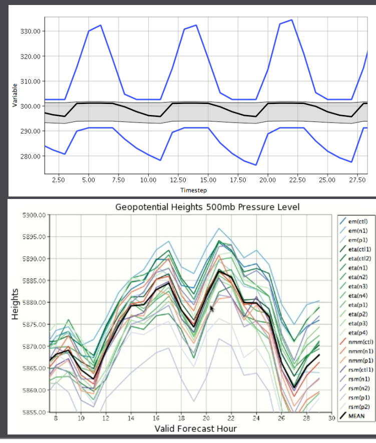


Spatial Summaries

Questions:

- What is the trend of the models over a region?
- What is the average trend?

Spatial Summaries



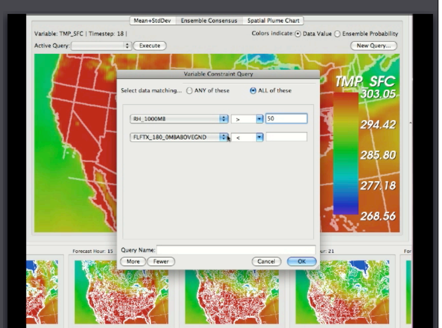
- *Summarize over user selected region*
- *average per model*
- *overall average/ boxplot statistics*

Query Summaries

Questions:

- *Where does the data express particular characteristics?*
- *What fraction of the data expresses it?*

Query Summaries



- *SQL type queries to filter the data*
- *Contours of ensemble fraction that predicts the condition*

Using the 3rd Dimension

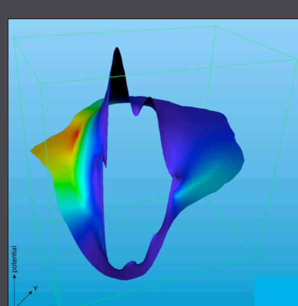
Summarizing reduces to 2D

use the free 3rd dimension for encoding

Displacement Mapping

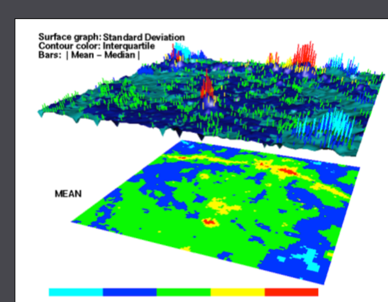
Change on Z axis

- *height encodes mean*
- *color encodes standard deviation*



Towards the Visualization of Multi-Dimensional Stochastic Distribution Data
Kristin Potter, Jens Krueger, Chris R. Johnson. (IADIS) 2008.

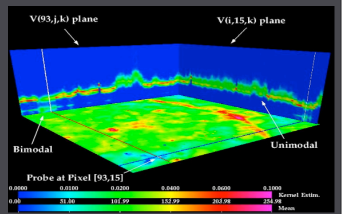
Layering



Four Statistics

- *(bottom) mean*
- *(top) deform by standard deviation*
- *color by interquartile range*
- *bar glyphs equal difference between mean/median*

Visualizing Spatially Varying Distribution Data
David Kao and Alison Luo and Jennifer L. Dungan and Alex Pang.
Information Visualisation, pp. 219--225, 2002.

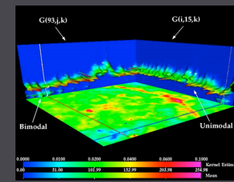


Cutting Planes

SLICE SPACE INTERACTIVELY

- *mean plane as summary*
- *reduce space by choosing cross-sections via probe*
- *project onto back walls to see PDFs for points under lines*

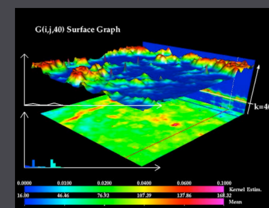
Visualizing Spatially Varying Distribution Data
David Kao and Alison Luo and Jennifer L. Dungan and Alex Pang.
Information Visualisation, pp. 219--225, 2002.



Local Surface Graphs

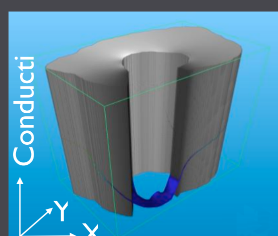
DISPLACE BY DENSITY

- *height + color easier to see*
- *combine with cutting planes*

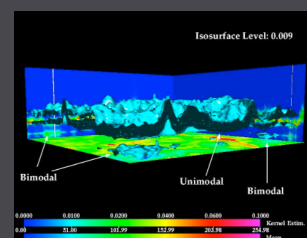


Visualizing Spatially Varying Distribution Data
David Kao and Alison Luo and Jennifer L. Dungan and Alex Pang.
Information Visualisation, pp. 219--225, 2002.

Direct Volume Rendering



- *z-axis represents input value (low to high)*
- *stack realizations/slices*
- *2D slices occlude each other*
- *not clear how to use transparency*
- *overall not effective*

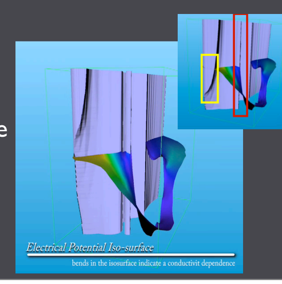


Isosurfaces

BASED ON DENSITY ESTIMATE

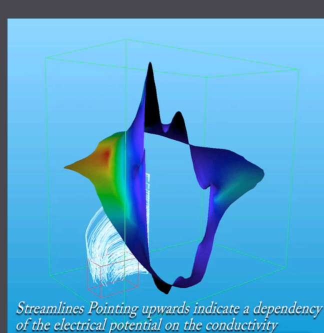
- *show structure of distribution*
- *global attributes*

curves in isosurface indicate dependence on input



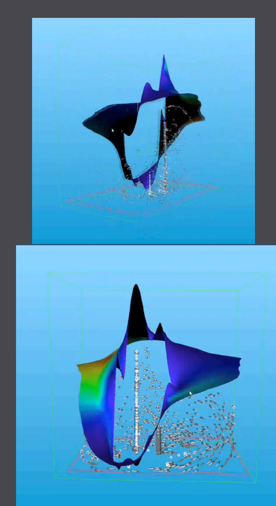
Streamlines

- *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*



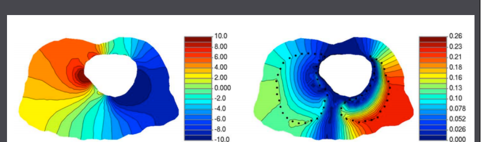
Are we asking the right questions?

Torso Data

Questions
- What is the average potential across the domain?
- What is the variation of potential across the domain?

Global Summary

- mean/standard deviation
- show each separately



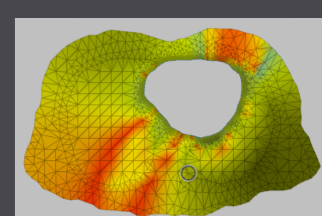
Mean Standard Deviation

But is it meaningful?

- know pdf of parameter input is roughly uniform
- are mean & standard deviation appropriate statistics?
- can we look at the data without individually inspecting each grid point?

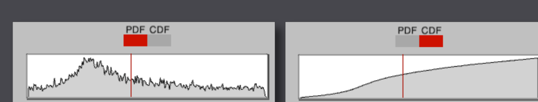
Torso Data

New Questions
- what do the PDFs look like across the domain?
- where are the PDFs similar or different?



ProbVis

Visualization & Exploration for Distributions
- show differences between PDFS
- summarize all data in a single view



Method

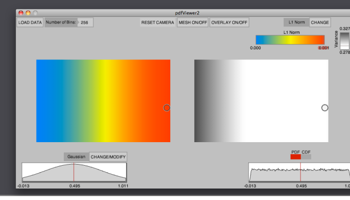
Comparison Between Two PDFS

- compare all data points to a single PDF
- get a single metric for each data point
- compare each pts metric

Method

Define a Difference Measure

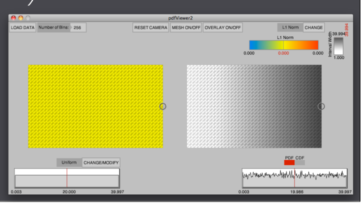
- shape: L1 Norm or Hellinger Distance



Method

Define a Difference Measure

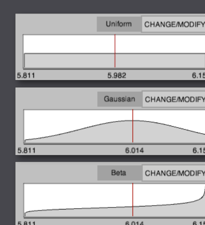
- interval: (range of sample values)



Method

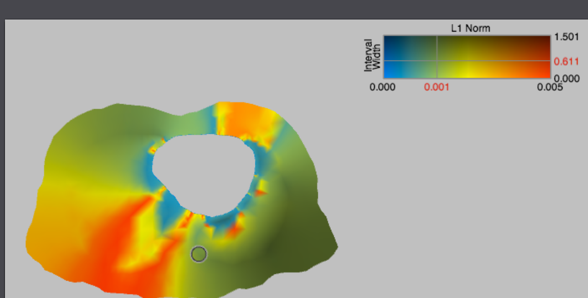
Library of Canonical Distributions

uniform, gaussian, beta



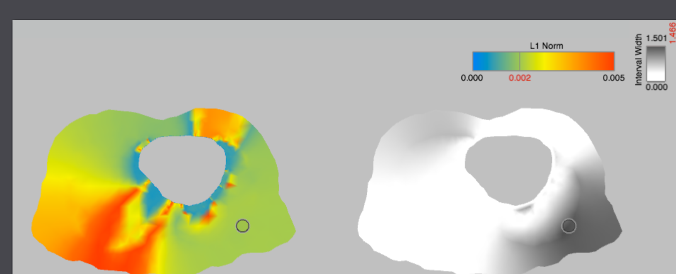
Visualization

Colormap distance measure

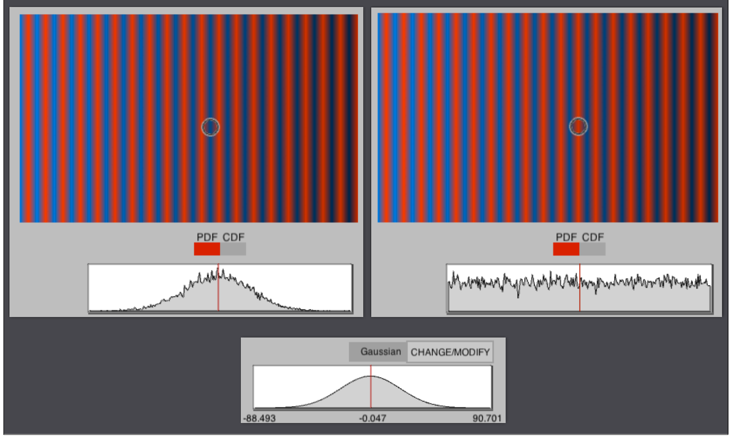


Visualization

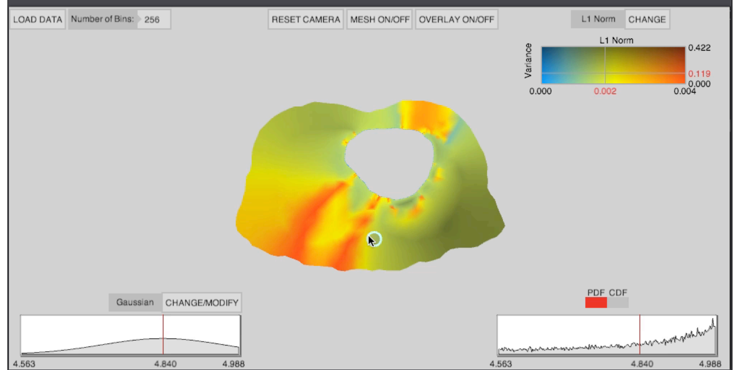
Colormap distance measure



Alternating Example



Interactivity



Summary

Aggregation takes place everywhere

- *required for visualization*
- *controllable for certain questions*

Future Directions

More work is needed!

- *increase our vocabulary of uncertainty measures*
- *further the visual metaphors used for uncertainty*
- *interaction will be required for most applications*

Thanks!

Questions

- *KAUST No. KUS-C1-016-04*
- *DOE NETL DE-EE0004449*