Uncertainty Visualization

Alex Pang
University of California, Santa Cruz

Bring umbrella?

Go sailing?

Hurricane Katrina

From NOAA

Ambiguity in fiber tracks


Policy & news uncertainty index
(Baker and Bloom, 2012)
Uncertainty is Certain

Really?

Motivation

- Important to know about uncertainty when analyzing and understanding data
- Even more important to know about uncertainty when making decisions

Where does uncertainty come from?

- Variability in nature
- Deficiency in instrumentations e.g. insufficient resolution, calibration drifts, ...
- Deficiency in modeling e.g. fidelity in physics, complexity, numerical imprecision, ...
- Insufficient or conflicting information
- Others e.g. introduced during visualization

Uncertainty in visualization pipeline
More places uncertainty is added
- Different methods of processing data
- Different rendering algorithms e.g. DVRs
- Filling in missing data
- Smoothing out high frequency data
- Filtering out outliers
- Improper use from what the visualization was originally designed for

Requirements for End-to-End Data Understanding with Uncertainty
- Uncertainty representation
- Uncertainty quantification
- Uncertainty propagation
- Uncertainty visualization

Uncertainty representation
- Scalar quantity e.g. confidence level, standard deviation, data quality, RMS, SNR, ...
- Pair e.g. min-max range values, mean and spread, ...
- 1D form e.g. probability density function
- 2D form e.g. covariance matrix
- Others?

Uncertainty quantification
- Statistics and probability
- Fuzzy set theory
- Possibility theory
- Evidence theory
- Information theory
- ...

Uncertainty propagation and evolution
- Belief propagation
- Interval arithmetic
- Stochastic PDE
- Ito calculus
- Fokker Planck
- ...

Uncertainty visualization
- Histograms
- Box plots
- Uncertainty glyphs
- Embellishments
  - Pseudo-coloring, transparency, texture
  - Fuzziness, dashed lines, dust clouds
- ...

IEEE Visualization 2012, Uncertainty and Parameter Space Analysis
Map with traffic & weather

Visualizing Scalar Uncertainty

Disclaimer
- Non-exhaustive list
- Can often swap what to highlight or to hide e.g. to emphasize or to de-emphasize uncertainty

Pseudo-Color: Seismology

Pseudo-Color: Point probe
(2d array of distributions)

HSV: Monte Carlo
Brightness: Bathymetry from sidescan sonar

Attributes: Fat, faint, noisy lines

Dashed contours

Example: Soundspeed (42 levels, 80 realizations)

Transparency

Bluriness, fuzziness
Localized fuzziness

Exploiting stereo-vision

Glyphs for positional uncertainty

Dust cloud

Grigoryan and Rheingans, 2002

1D Transfer Function

Transfer function: Increasing opacity with uncertainty

2D Transfer Function

Mean salinity

Salinity data
Animation

Fuzziness: Infovis

Glyphs: Graph Comparison

Color Coding: WikiTrust

Parallel Coordinates

Visualizing Vector Uncertainty
Uncertainty glyphs

Textures
(cross advection, multi-frequency noise)

Botchen et al., 2005

FTLE and Predictability (2012)
(Hlawatsch et al.)

Low predictability when started in high FTLE (reddish regions)

Can be used to place seeds in finely folded ridges

Visualizing Spatial Distributions

Motivation

Deficiency of parametric statistics

Same mean and std dev
Motivation
- Limitation of traditional scalar viz

Histograms

Histograms
(cannot scalable in space)

Motivation
- Increasing use of Monte Carlo simulations to capture uncertainty in models and parameters
- Increasing use of multiple models to form ensembles
- Results in multi-dimensional data with multiple values for the same variable at each location and time.

Possible Strategies
- Use an existing method e.g. histograms
- Reduce dimensionality:
  - Parametric statistics
  - Shape descriptors
- Deal with multi-value directly:
  - Operator algebra
- ...

IEEE Visualization 2012, Uncertainty and Parameter Space Analysis
Histograms and box plots

- simplicity, familiarity
- difficult to scale to higher spatial dimensions

Multivariate-Parametric approach

- Collect statistics at each point
  - e.g. mean, standard deviation, skewness
  - Convert 2D array of distributions into 2D array of n-tuples
- Map n-tuples at each point to visuals
  - E.g. map statistics to color, surface height, contours, glyphs
  - Add additional layers as desired

- Image plane deformed by standard deviation.

- Surface graph colored by inter-quartile range.

- Line glyphs indicate |mean-median| with a threshold.
Advantage/disadvantage:

+ familiarity
- potential clutter
- shows summaries, not distributions
e.g. 2 distributions with same mean and standard deviations, but different shapes

Non-parametric approach

Show shape of the distributions:

- Show tiny histograms at each point
- Does not scale with resolution and dimension
- Need other shape descriptors aside from parametric statistics

Distribution Profiles

3D histogram slices

Pseudo-Color : Point probe

(2d array of distributions)

Advantage/disadvantage

+ more general, does not assume well-behaved distributions
- hard to extend to higher dimensions
**Operator algebra**

- Provide a set of operators that specify how multi-value data can be combined with other multi-value data or data types.
- Operations can be mathematically or procedurally defined.

**Simple Operators**

\[ s = \text{ToScalar}(M) \]

\[ v = \text{ToVector}(M) \]

**Pseudocoloring**

\[ \text{Mean} = \text{ToScalar}(M) \]

**HSV Mapping**

\[ (\text{mean, stddev, skewness}) = \text{ToVector}(M) \]

- Mean = hue
- Skew = saturation
- Stddev = value

**Arithmetic**

- Combining scalars with multi-values:
  
  \[ M' = s + M \]
  
  \[ M' = sM \]

- Combining multi-values:
  
  \[ M' = M1 + M2 \]
  
  \[ M' = M1 \times M2 \]

**Interpolation**

- Exercise: think about how to interpolate 2 distributions – both are Gaussian with the same standard deviation, but different means.
- How does the interpolated distribution look like halfway through?
Streamlines

- Euler integration
  \[ P_{i+1} = P_i + v \Delta t \]
- Addition of multi-values
  - Convolution addition (Gerasimov et al.)
  - Binwise addition (Gupta & Santini)
  - ...

Spaghetti Plots

Streamlines with Convolution +

Streamlines with Binwise +

Comparing multi-values

- Euclidean distance:
  \[ ED(P, Q) = \left( \int (P(x) - Q(x))^2 \, dx \right)^{\frac{1}{2}} \]
- Kullback-Leibler distance:
  \[ KL(P, Q) = \int P(x) \log \frac{P(x)}{Q(x)} \, dx \]
- Kolmogorov-Smirnov distance:
  \[ KS(P, Q) = \max |cdf(P(x)) - cdf(Q(x))| \]
Isolines / Contour Lines

Pseudo-colored by similarity distance.

Isosurfaces

Pseudo-colored by standard deviation.

Advantage/disadvantage

+ more general approach
+ can be used for feature extraction
+/- results vary with operators
- learning curve for interpretation

Uncertain Isocontours (2012)

(Pothkow et al.)

Uncertainty in InfoVis

Extend what we learned in SciVis to visual attributes of graph elements: node shade, color, size, etc.; edge thickness, sharpness, etc.

Adapt techniques for UncVis e.g. Parallel coordinates

New metaphors and layout e.g. bulls-eye that utilize layout to convey uncertainty

Layout: Bulls-Eye Metaphor
Challenges

- End-to-end treatment of uncertainty
- Tools for data analysis and visualization of spatially dense pdfs e.g. ensembles
- Uncertainty in infovis / visual analytics
- How can we make better visualizations for decision makers

Hazards of Communicating Uncertainty

- Cone of Uncertainty Confusing? (w. oremus)
  - Does it show the areas most likely to be affected by the storm to one degree or another, or just the areas over which the storm's eye is most likely to pass?
  - The cone does not show the extent of the storm or how widespread its damages might be—just the areas over which its very center is most likely to pass.
  - If you live outside of the cone entirely, what are the chances that you'll suffer a direct hit anyway?
  - If you live outside of the cone entirely, you could still get hit by the eye of the storm. The chances of the storm's center leaving outside the cone are about one in three.*
  - Where the cone grows wider, does that mean the storm is likely to spread out and dissipate?
  - Not necessarily. Where the cone grows wider, it simply means that there is more uncertainty about where exactly the storm will strike.
  - Does the shape of the cone vary depending on the nature of the storm?
  - No, the shape of the cone—that is, how narrow or broad it is at any given point—has nothing to do with the individual storm. It is determined prior to each hurricane season based solely on the accuracy of past predictions, and is the same for every storm that season.

Future Research Directions

- How does one do feature extraction with multi-value data set: uncertainty in features e.g. critical point vs fuzzy critical region
- How can we retain the spatial correlation information in multi-value data sets?
What’s a good representation for sparse or missing data?

How can we encode the uncertainty in graphics/visualization algorithms e.g. LODs, clustering, features, ...

Need more research in uncertainty visualization for infovis/visual-analytics.

Questions?