CSC 544
Data Visualization

Joshua Levine
josh@arizona.edu

Lecture 25
Uncertainty

Apr. 17, 2023

Today’s Agenda

- Reminders:
  - A06 questions? (due Apr. 24)
  - P03/P04 questions? (due Apr. 26/May 3)
  - Student Course Surveys (SCSs) (due May 3)
    - Currently at 44% class completed (80% threshold unlocks extra credit!)
- Goals for today: Discuss visualizing data in the presence of uncertainty

Presentation Format

- All presentation materials are due Apr. 26th, regardless of when you are scheduled.
- Plan for a 15 minute time slot. This means approximately 10 minute presentation, 3 minutes for questions, 2 for transition
- Q+A during the presentations, will factor into your **class participation** grade.

Presentation Content

- Your presentation must address four points:
    - You must introduce the **research question** that you studied, and **what motivated you** to pursue it.
    - You must briefly describe **related work** in this area.
    - You must describe your **research plan**. Feel free to contextualize this with how your plan evolved, but primarily you should focus on what **techniques** you employed and what **data sets** you may have used.
    - You must describe the **results you achieved**. Where appropriate, feel free to use a live demo any software tools you generated. Mention **how you evaluated** your work as well.

Project Reports

- Reminder: these are due on May 3

Visualizing Uncertainty

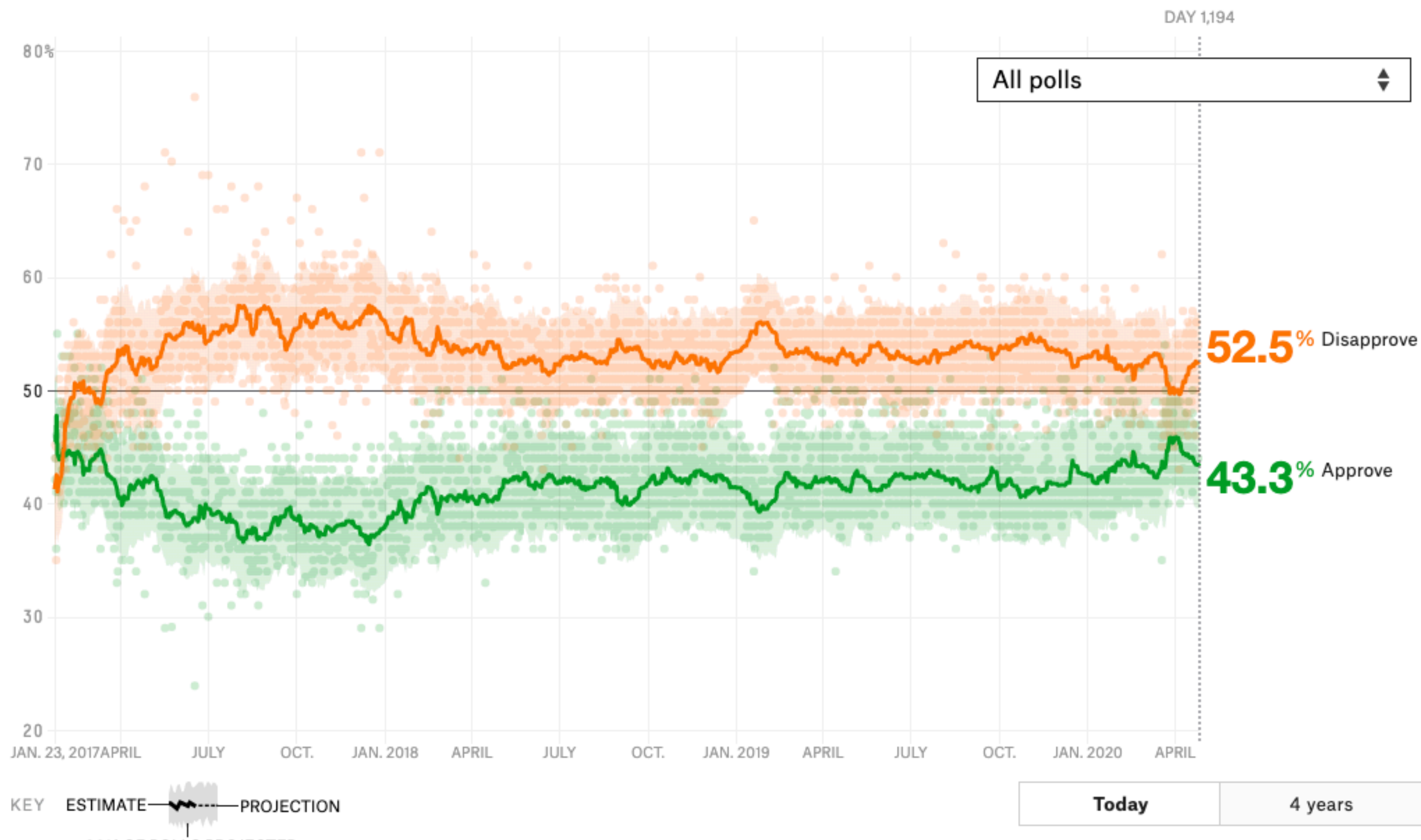
projects.fivethirtyeight.com/trump-approval-ratings/

FiveThirtyEight

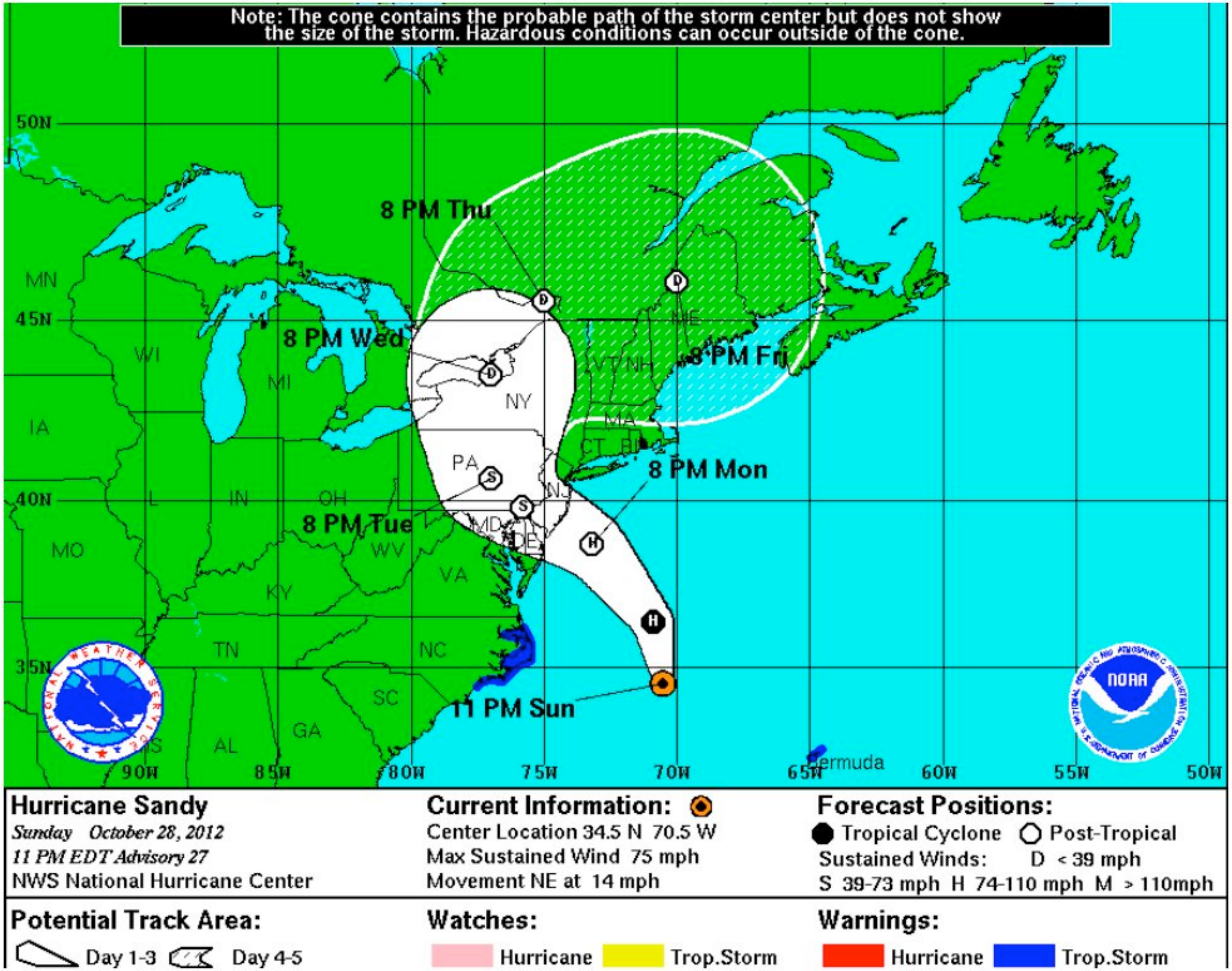
UPDATED APR. 27, 2020 AT 9:36 AM

How popular is Donald Trump?

An updating calculation of the president's approval rating, accounting for each poll's quality, recency, sample size and partisan lean. How this works »



Hurricane Sandy, 2012



https://www.nhc.noaa.gov/archive/2012/SANDY_graphics.shtml

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?

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.

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?

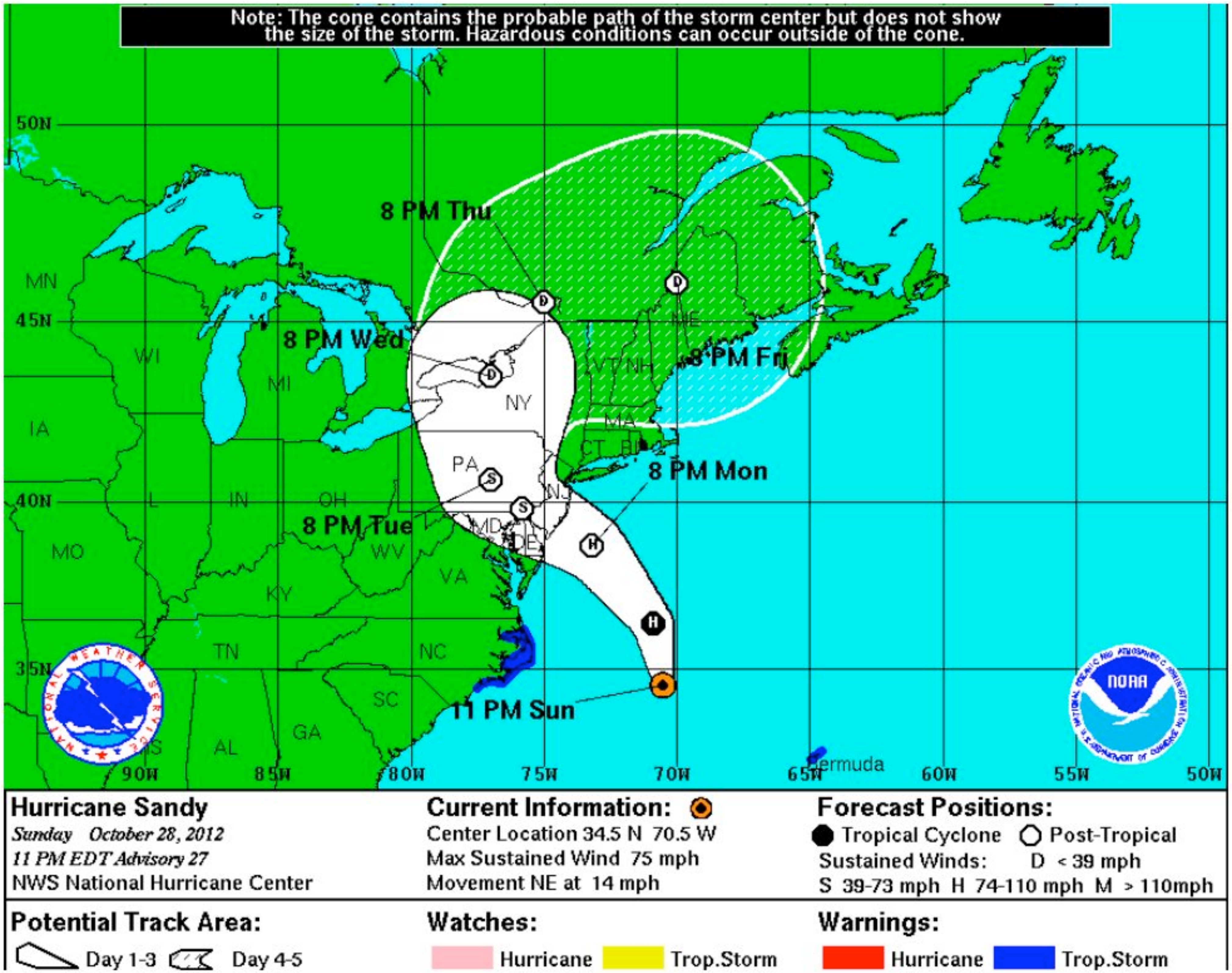
Does it show the areas most likely to be affected by the storm to one degree or another, or just the areas over which the storm’s eye is most likely to pass?

- 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 straying outside the cone are ***about one in three***.

Hurricane Sandy, 2012



https://www.nhc.noaa.gov/archive/2012/SANDY_graphics.shtml

Where the cone grows wider, does that mean the storm is likely to spread out and dissipate?

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.

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?

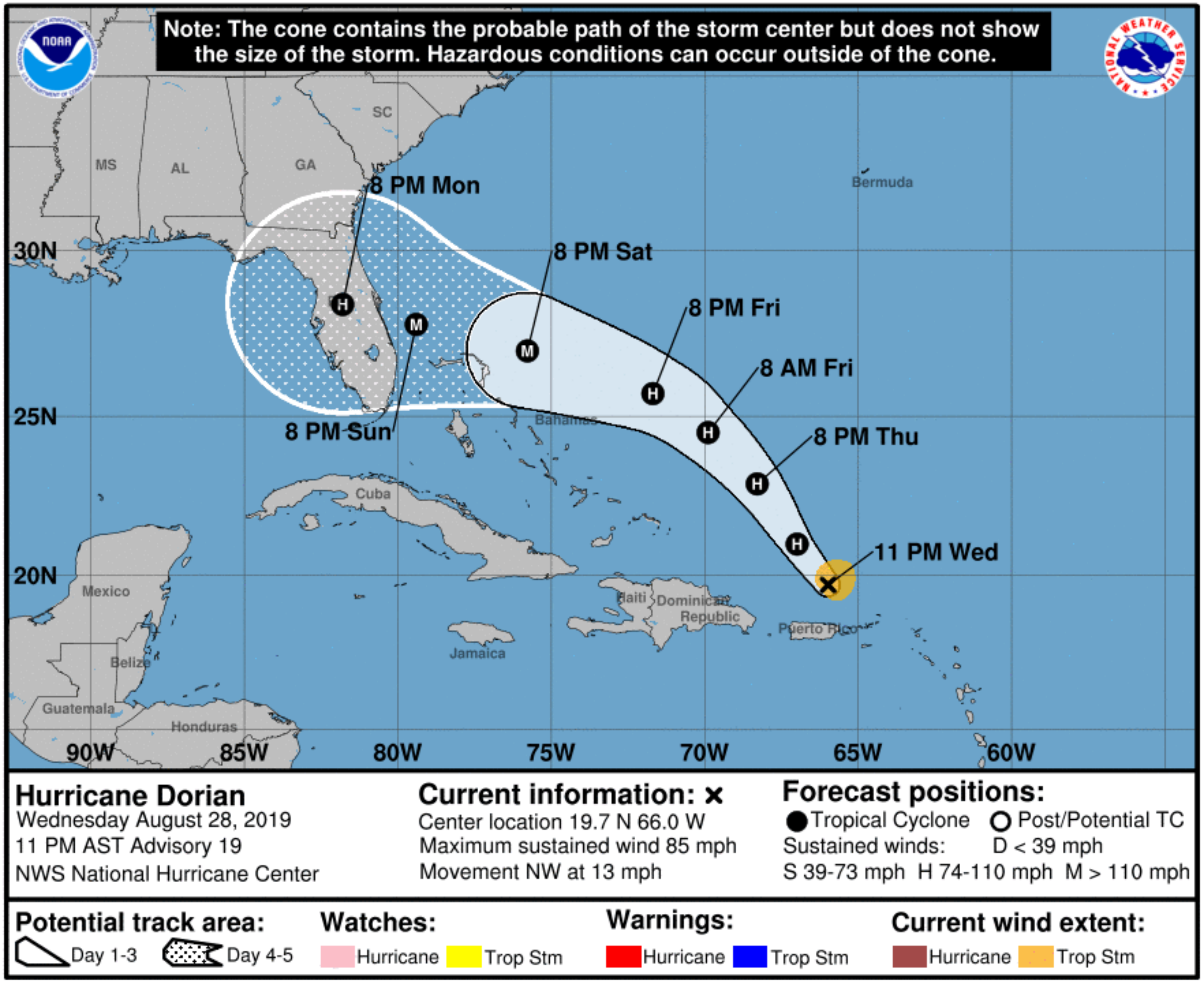
Where the cone grows wider, does that mean the storm is likely to spread out and dissipate?

- 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?

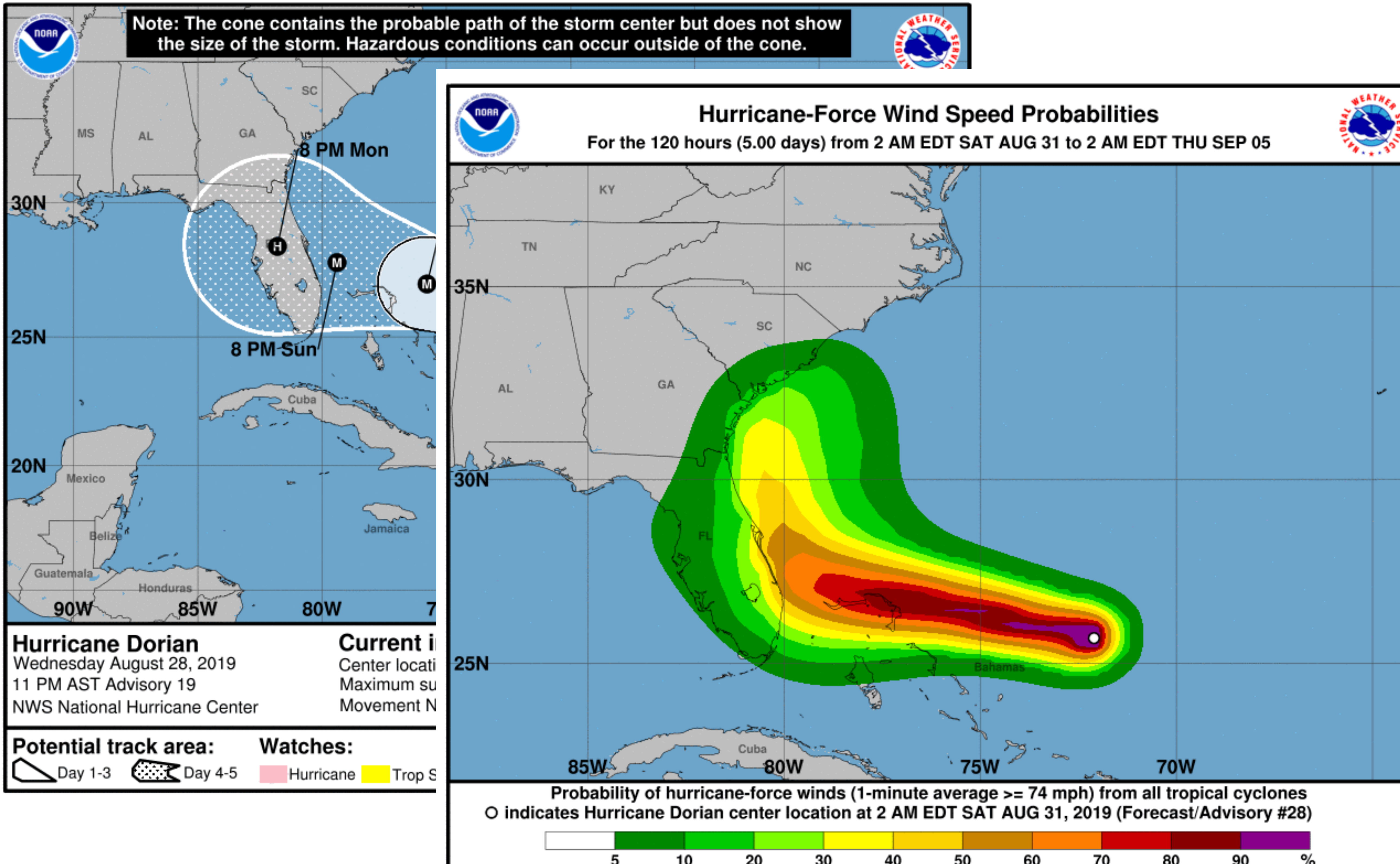
- No, the shape of the cone---that is, how narrow or broad it is at any given point--has ***nothing to do with the individual storm***. It is determined prior to each hurricane season ***based solely on the accuracy of past predictions***, and is the same for every storm that season.

Hurricane Dorian, 2019



https://www.nhc.noaa.gov/archive/2019/DORIAN_graphics.php

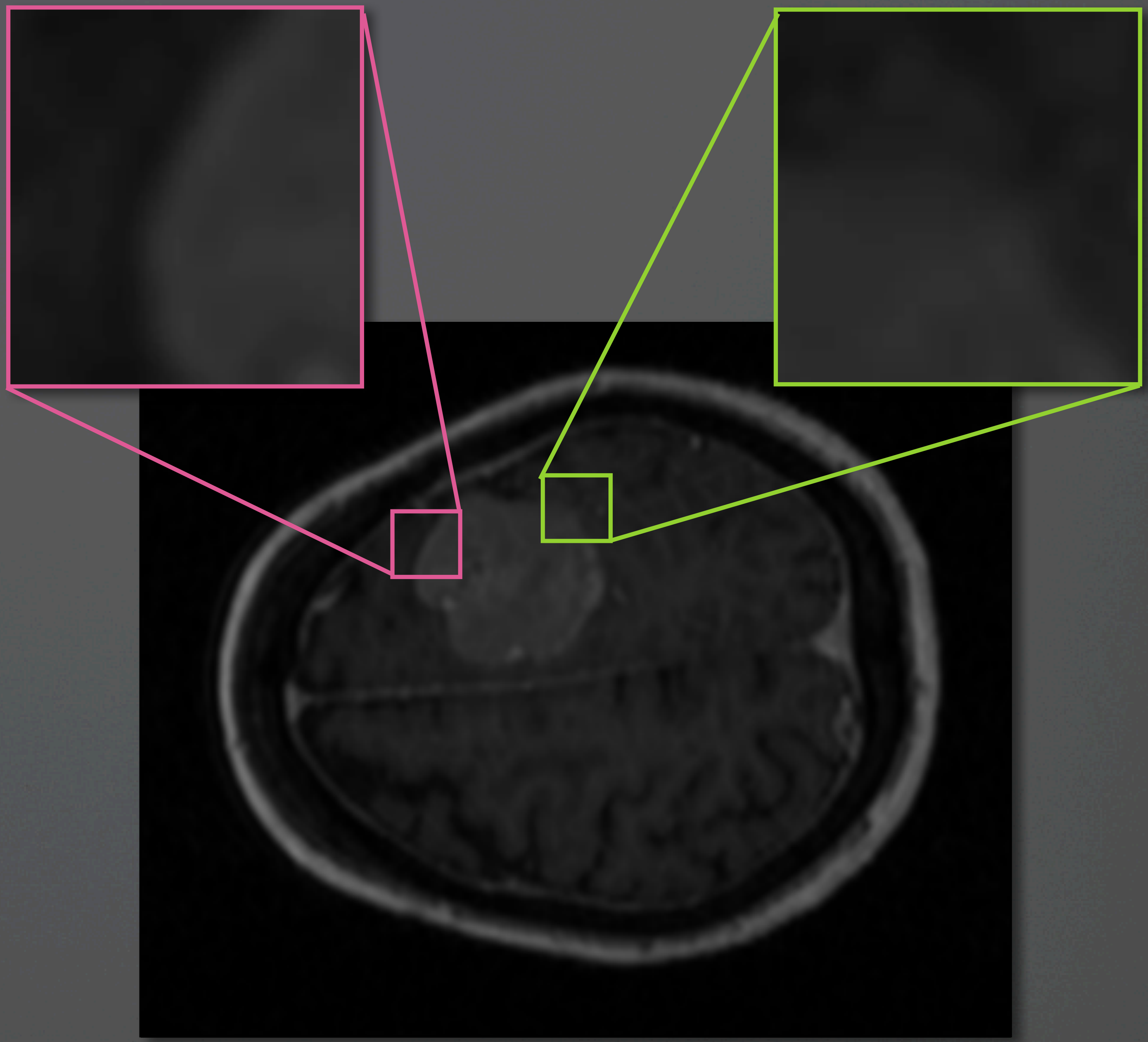
Hurricane Dorian, 2019



https://www.nhc.noaa.gov/archive/2019/DORIAN_graphics.php

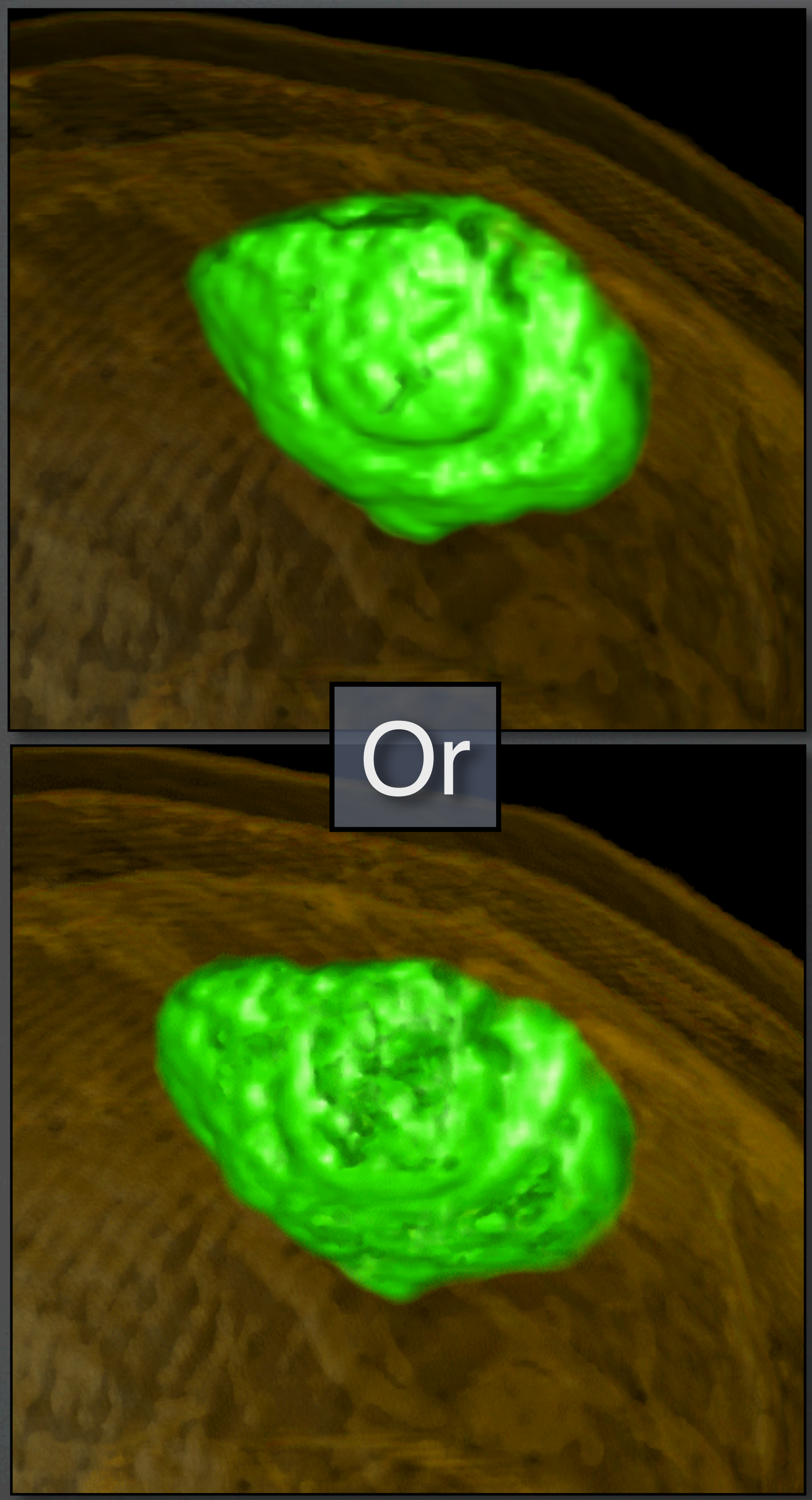
Brain Tumor Example

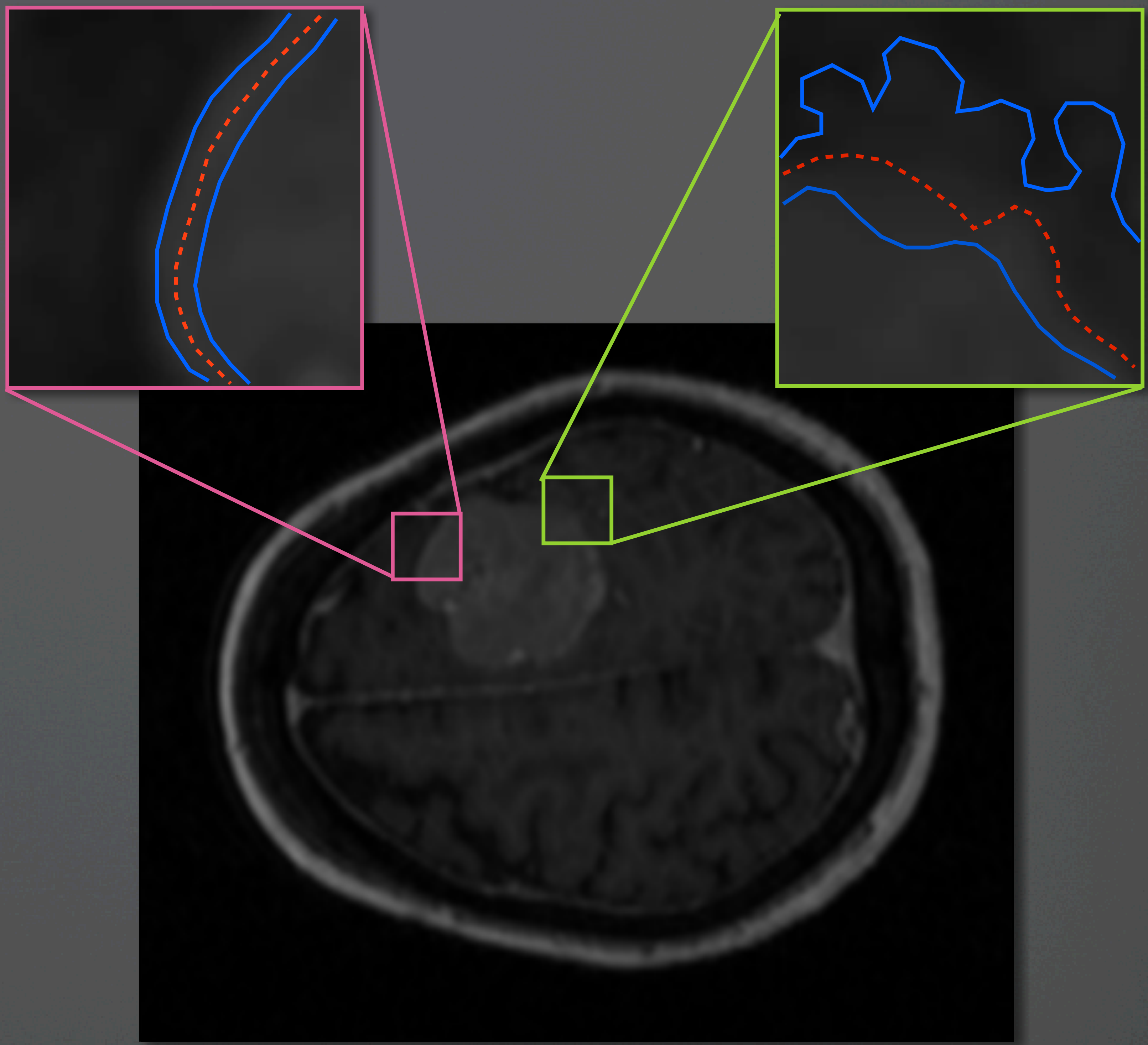
- Fuzzy boundaries exist in the data
- How to distinguish between tumor and gray matter?
- Pre-operative planning: Doctors (and patients!) need to know confidence of the line delinating tissue types



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

Brain Tumor Example - cont





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

Visual Design w/ Uncertainty

Types/Sources of Uncertainty

- Aleatoric (randomness inherent in the process)
- Epistemic (uncertainty for a lack of knowledge)
- Ontological (uncertainty about a model’s veracity)

- When uncertainty can be quantified (often with probability distributions), we can include it in a visualization.

Causes/Sources of Uncertainty in Data

- Data Ensembles:
  - Collections of simulations that explore a space of parameters or realize a collection of instances
- Model Inadequacy:
  - Lack of knowledge of true phenomena that we try to approximate
- Algorithmic Approximations:
  - Caused when we translate models to computational settings
- Uncertainty directly caused by the visualization:
  - e.g. interpolation errors introduced mapping sampled data

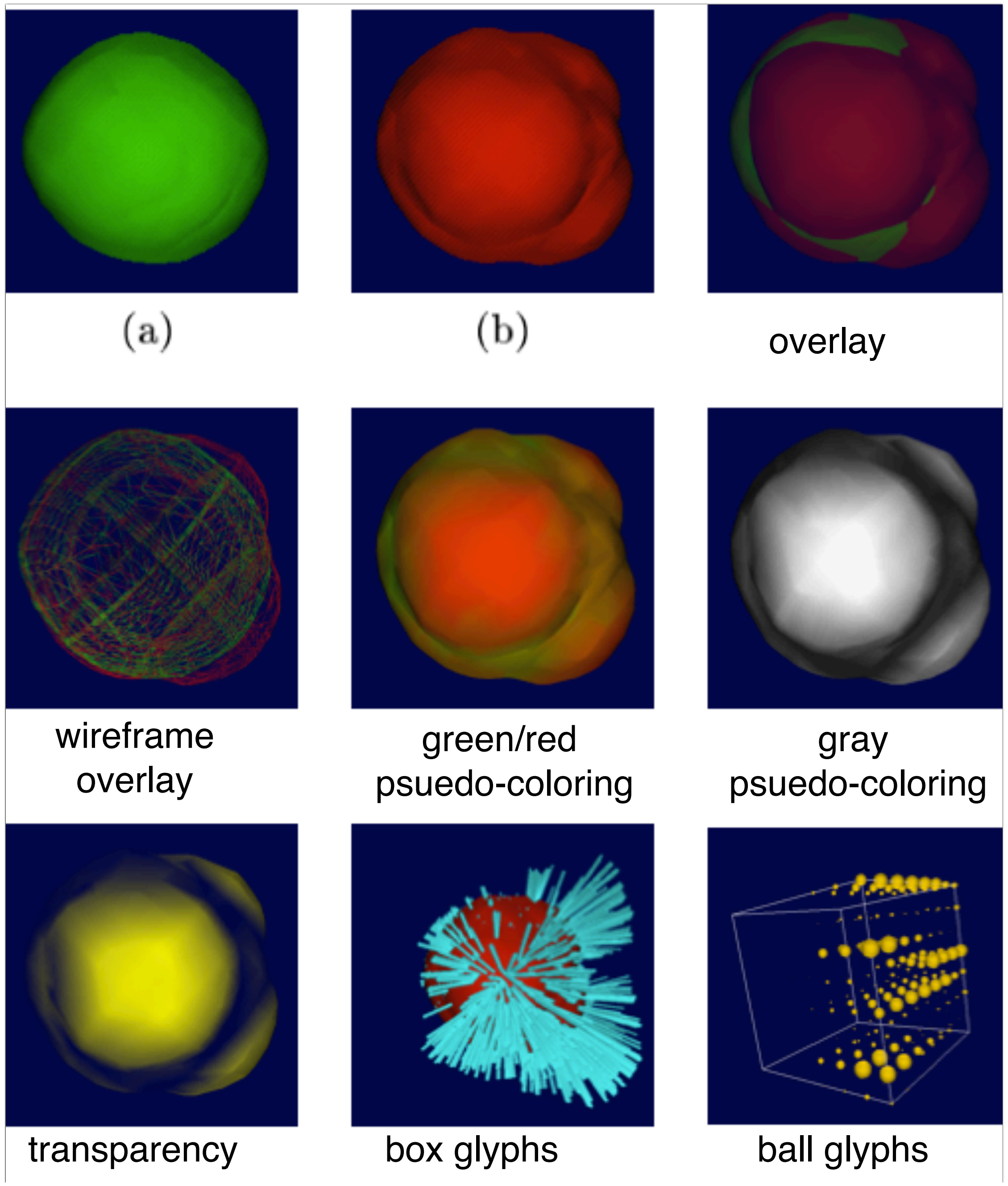
Uncertainty Visualization Hypotheses

- Frequency Framing: Uncertainty is more intuitive as a frequencies (1 out of 10) vs probabilities (10%)
- Attribute Substitution: viewers will substitute difficult mental computations for easy ones: often interpreting uncertainty information as deterministic
- Visual Boundaries = Cognitive Categories: boundaries create false categorical separation
- Visual Semiotics: Certain visual channels will work better for encoding uncertainty

Padilla, Kay, and Hullman. Uncertainty Visualization (Ch. 22). In Computational Statistics in Data Science (pp. 405-421). Wiley, 2022.

Encoding Strategies

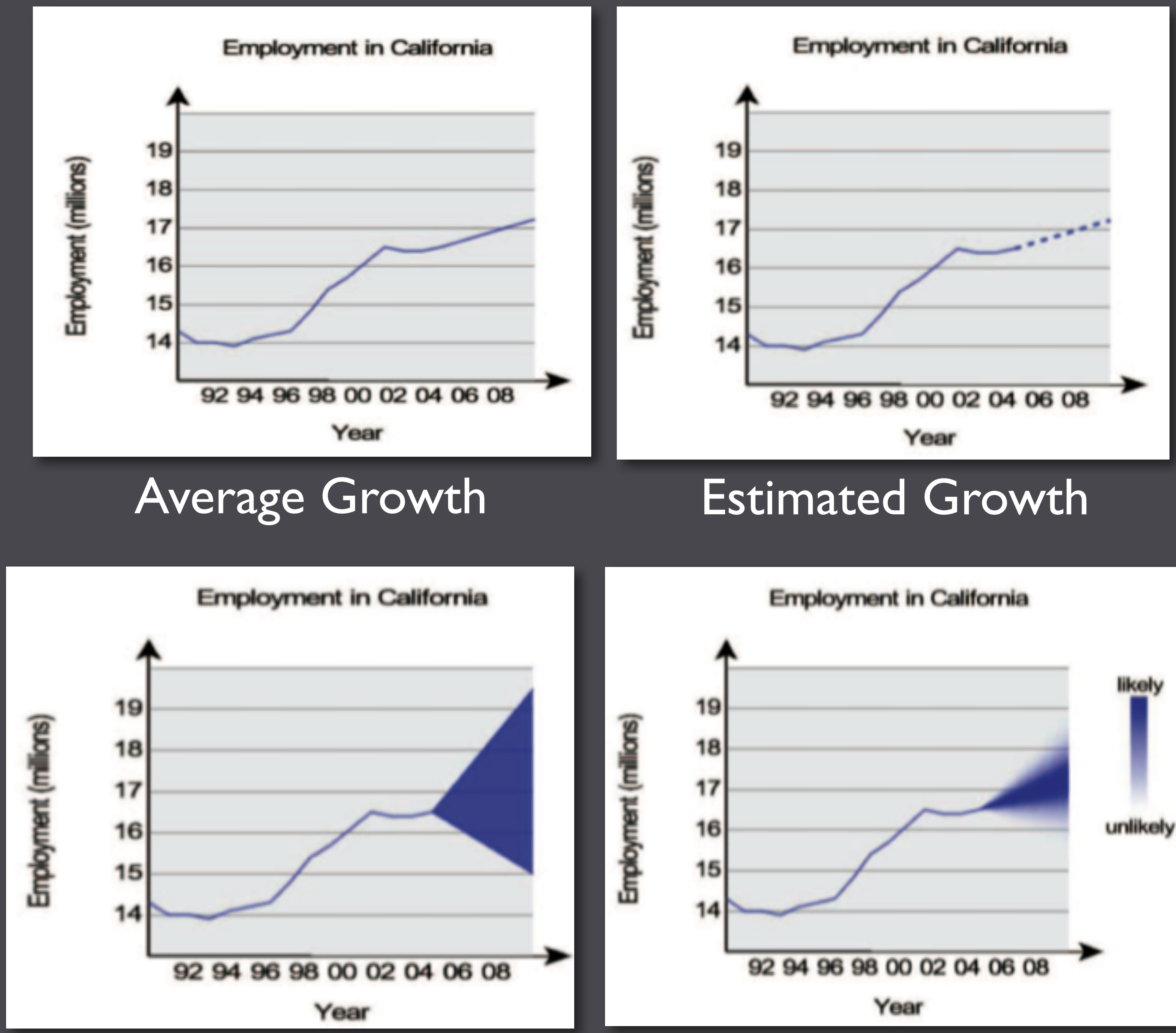
- Often additional visual channels are used to characterize the uncertainty, e.g. hue, texture, opacity
- Sometimes glyphs are used to encode specific annotations
- More complex geometric primitives can be employed, e.g. stream tubes instead of streamlines



Joseph, Lodha, Renteria, and Pang. UISURF: Visualizing Uncertainty in Isosurfaces. In Proceedings of the CGI, 184-191, 1999.

Information Uncertainty

- *indication of known information*
- *qualitative rather than quantitative*
- *spreadsheet interface characterizes the data*



Average Growth

Estimated Growth

Possible Growth

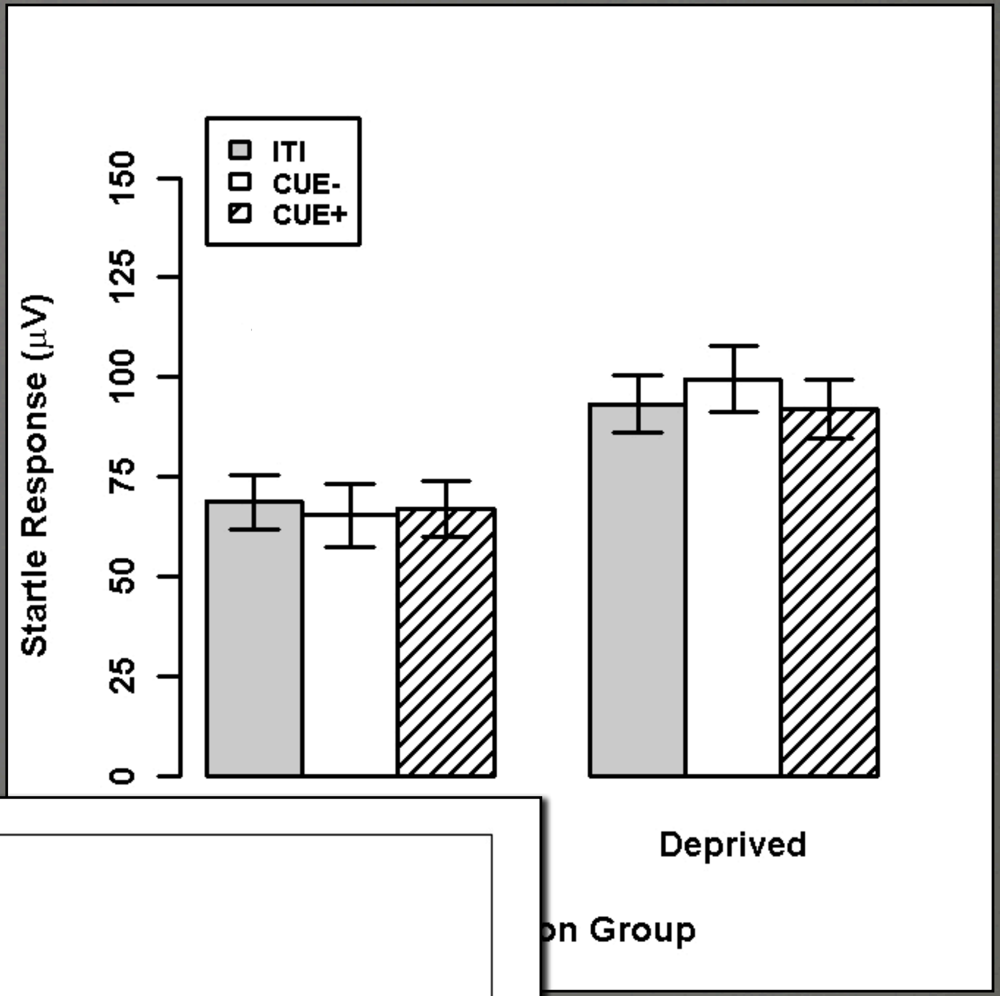
Likely Growth

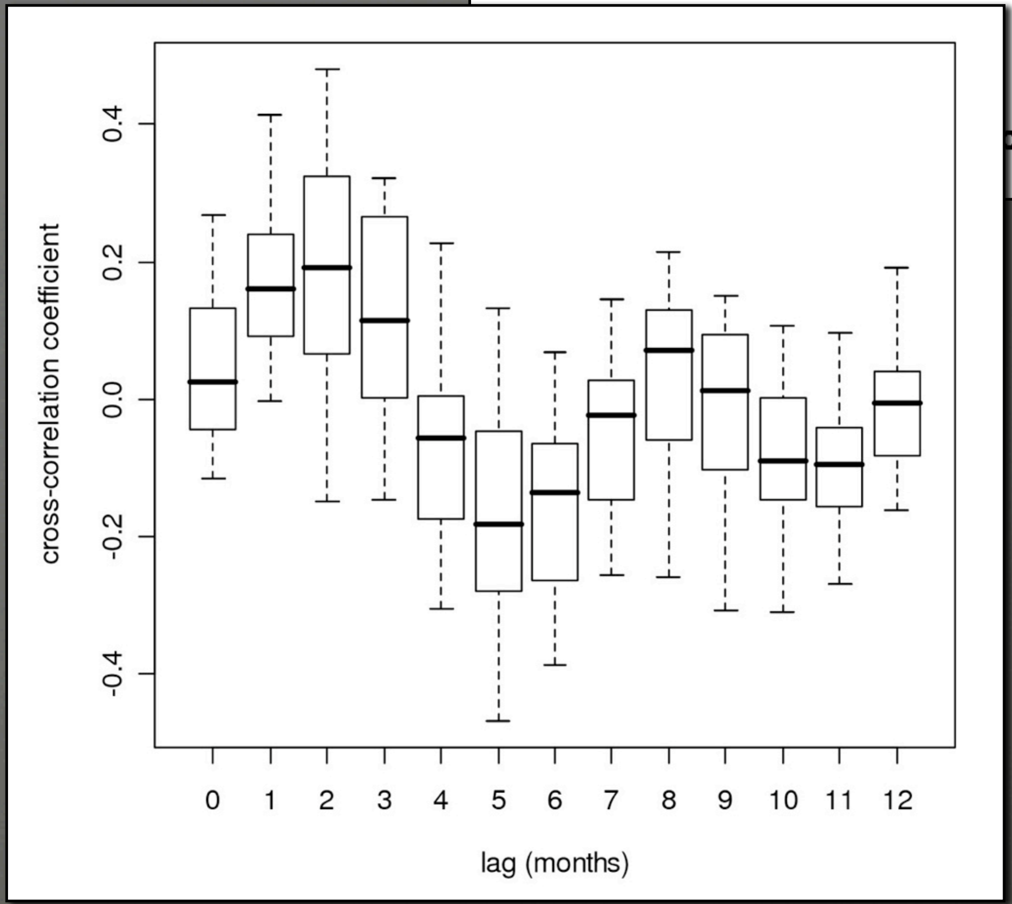
A Spreadsheet Approach to Facilitate Visualization of Uncertainty in Information.
A. Streit, B. Pham, R. Brown. TVCG 14(1), 2008.

SCI INSTITUTE

Traditional Display of Uncertainty

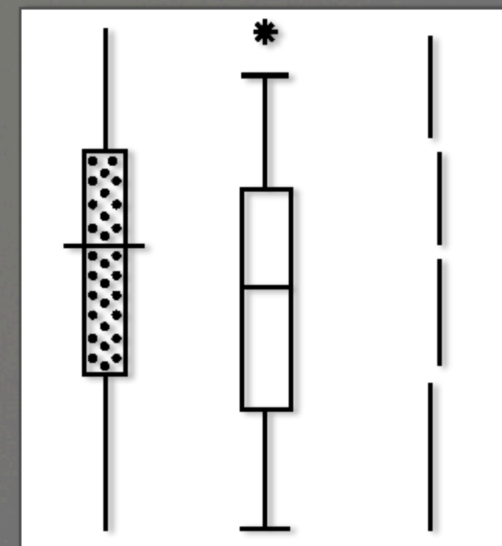
- Error bars
  - convey accuracy by amount of +/- error
  - std dev or std error
- Boxplots
  - Quartile range including median
  - Outliers





Boxplot Modifications

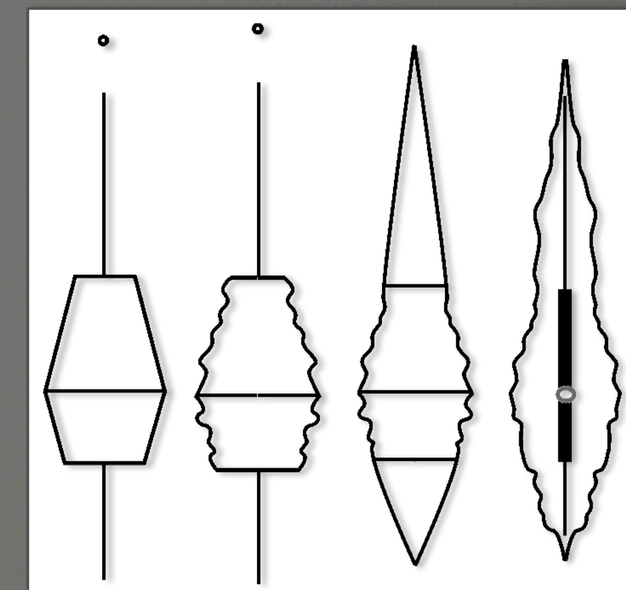
- Visual Modifications
  - Refinement for aesthetic purposes
- Density indications
  - Use the box sides to encode
- Data Characteristics
  - sample size, confidence levels
- Additional Statistics
  - skew, modality



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

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

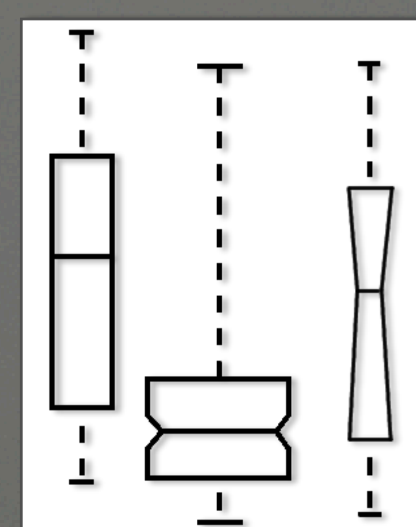
John W. Tukey.
Exploratory Data Analysis.
Addison-Wesley, 1977.



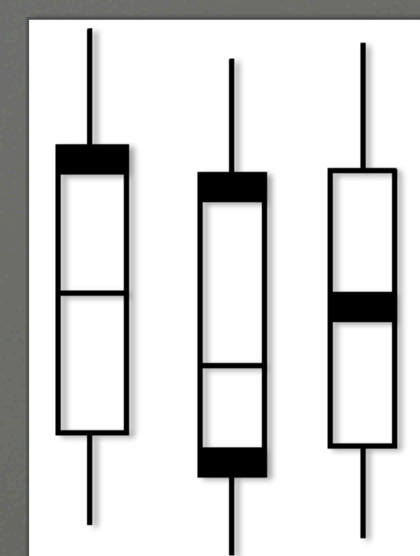
Y. Benjamini.
Opening the box of a boxplot.
TAS, 42(4) ,1988.

W. Esty, J. Banfield.
The box-percentile plot.
JSS, 8(17), 2003.

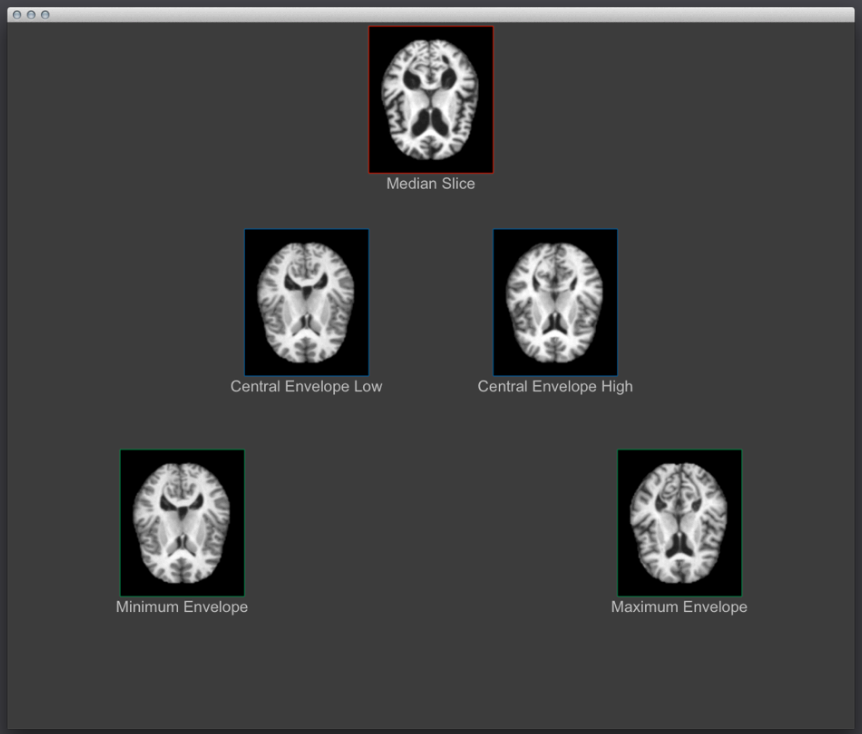
J. Hintze, R. Nelson.
Violin plots.
TAS, 52(2), 1998.



R. McGill, J. W. Tukey, W.A. Larsen,
Variations of box plots.
TAS, 32(1), 1978.



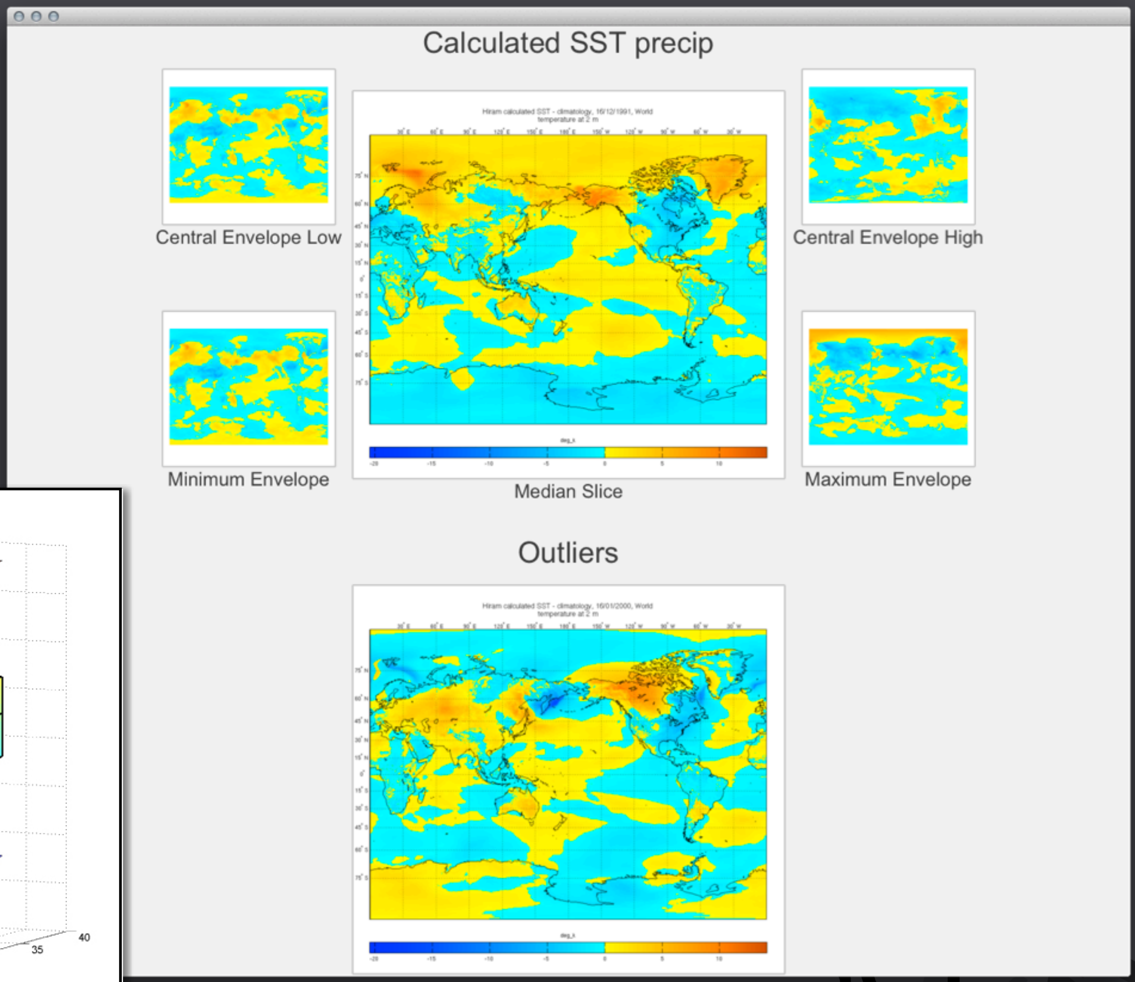
C. Choonpradub, D. McNeil.
Can the box plot be improved?
Songklanakarin J Sci Technol,
27(3), 2005,

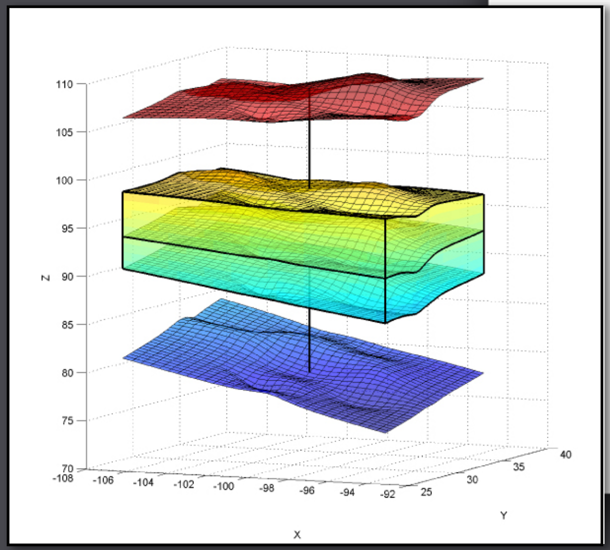


Surface Box Plots

Extension to 3D

- *images rather than curves*
- *volume-based band-depth*





Boxplots for Isocontours

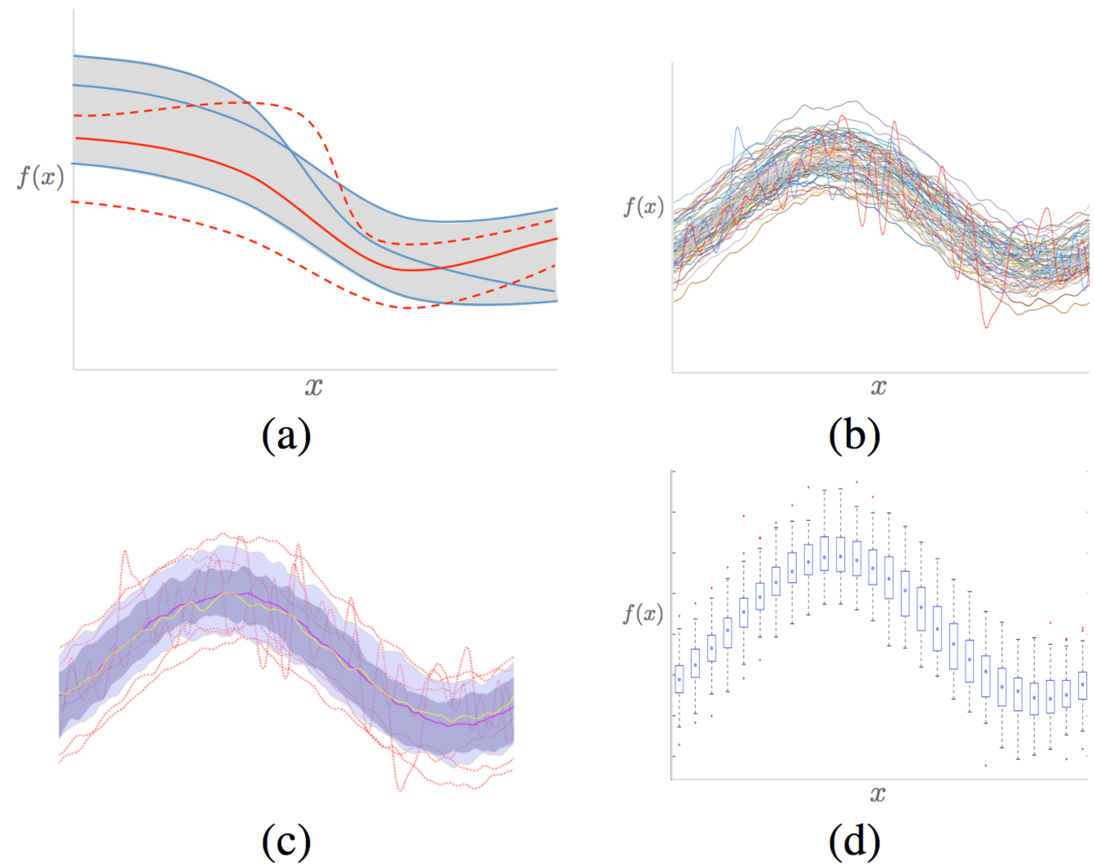


Fig. 2. Examples of function band depth and function boxplots. (a) For band depth, three different curves (in blue) form a band (in grey) against which three other (red) curves are tested, with only the solid curve falling in the band. (b) A set of 80 simulated, noisy curves with some outliers in shape and position. (c) A version of the function boxplot, as proposed by [35, 36]. (d) A conventional, pointwise, boxplot loses information about shapes of curves in the ensemble.

Ross T. Whitaker, Mahsa Mirzargar, Robert M. Kirby. IEEE TVCG. 19(12): 2713-2722 (2013)

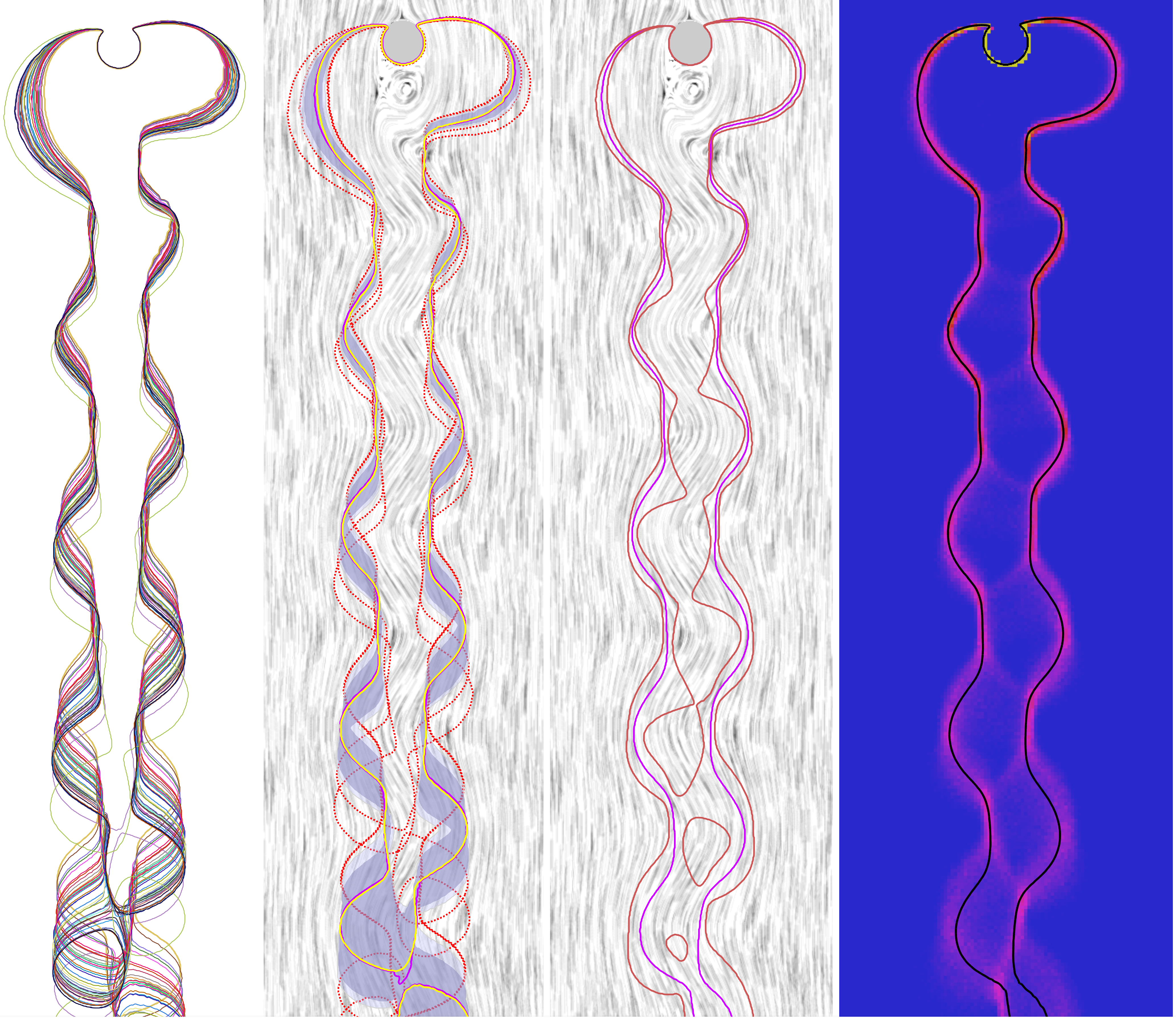
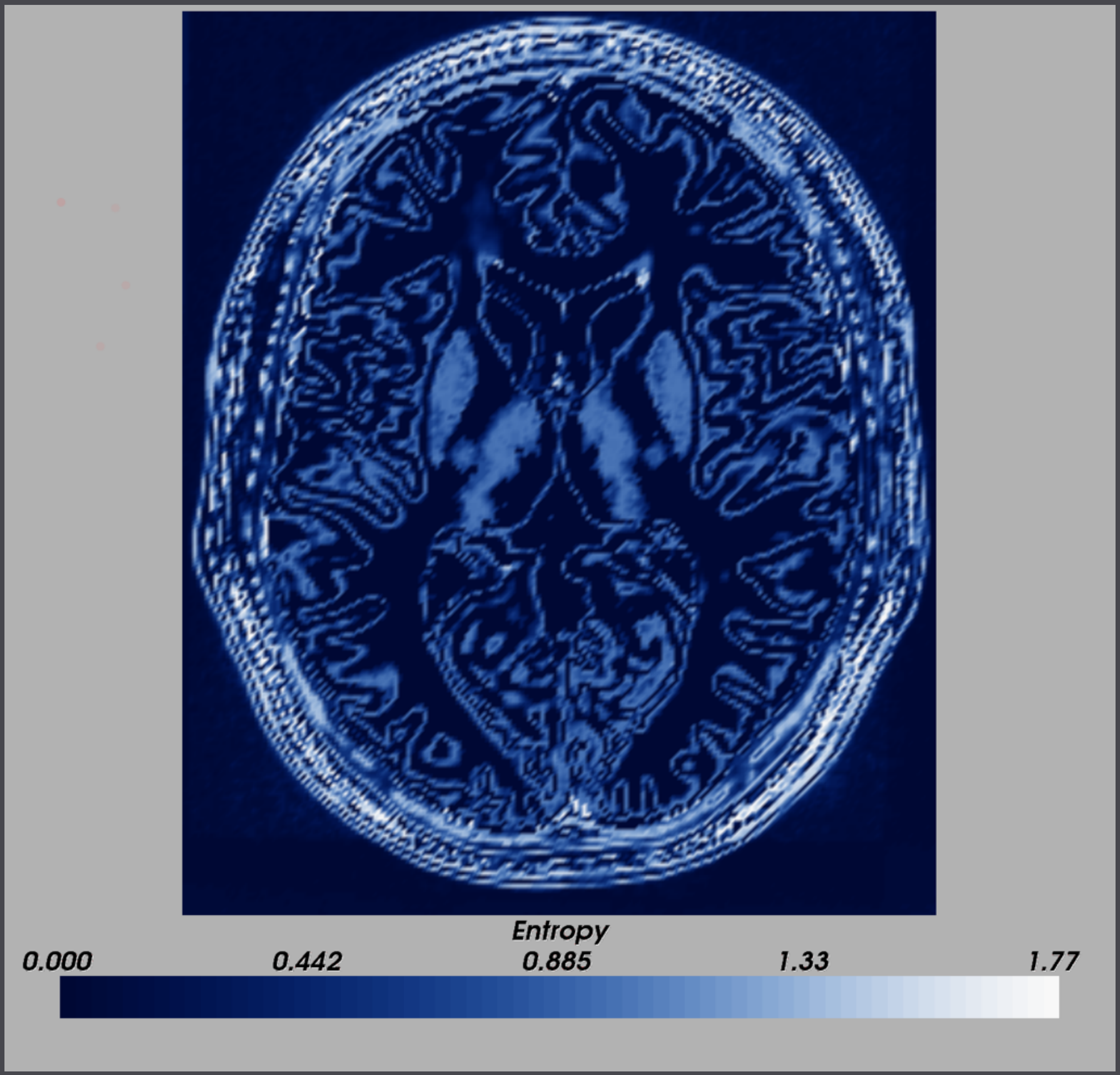


Fig. 7. (a) Ensemble of isocontours of the pressure field (b) Contour boxplot of the fluid flow simulation using 40 ensemble members with different perturbations in the initial conditions. (c) Demonstration of the isocontours of the mean and ± 1 standard deviation of the pressure field. (d) Color mapped level set crossing probabilities of the pressure field calculated using the probabilistic marching cubes algorithm (with the mean pressure shown as a black contour line). The boxplots are rendered over the LIC visualization of the median member of the ensemble.

Measures other than mean



Entropy (information theory)

- *describes the randomness of a voxel*
- *i.e. 0 entropy a tissue's type is defined*
- *higher entropy, more uncertainty*

Visualization of Uncertainty without a Mean

K. Potter, S. Gerber, and E. W. Anderson. CG&A Vis Viewpoints, 2013, to appear.

SCI INSTITUTE

Mean vs. Entropy

- Let X be a random variable probability density p
- Mean: $E[X] = \Sigma_{i=1}^{n} x_i p(x_i)$
  - Only meaningful if addition of random variables is sensible
  - Particularly for categorical data, it might not be!
- Shannon Entropy: $H[X] = -\Sigma_{i=1}^{n} p(x_i) \log(p(x_i))$
  - Only measures randomness rather than values.

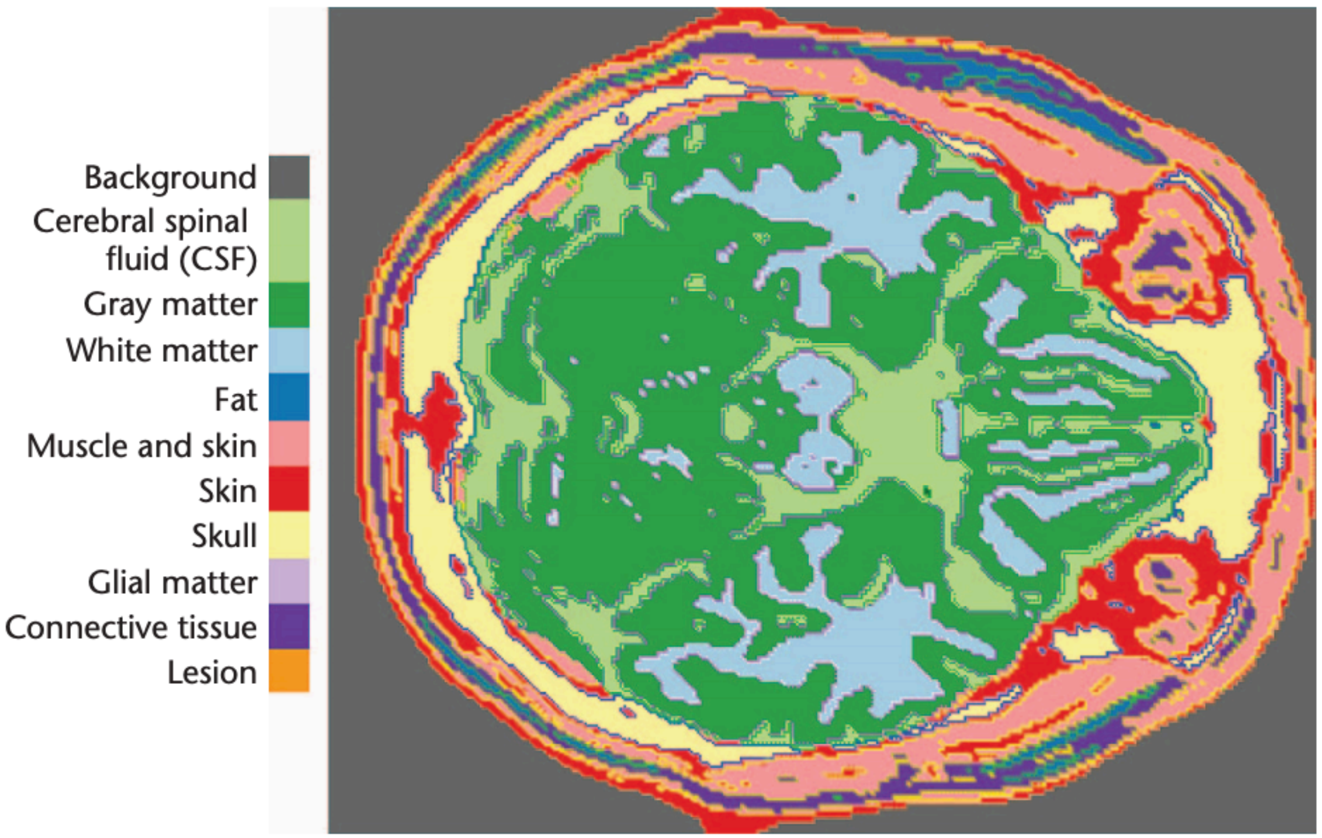


Figure 3. A slice through the tagged data, with each voxel having the color of the corresponding maximum-probability tissue.

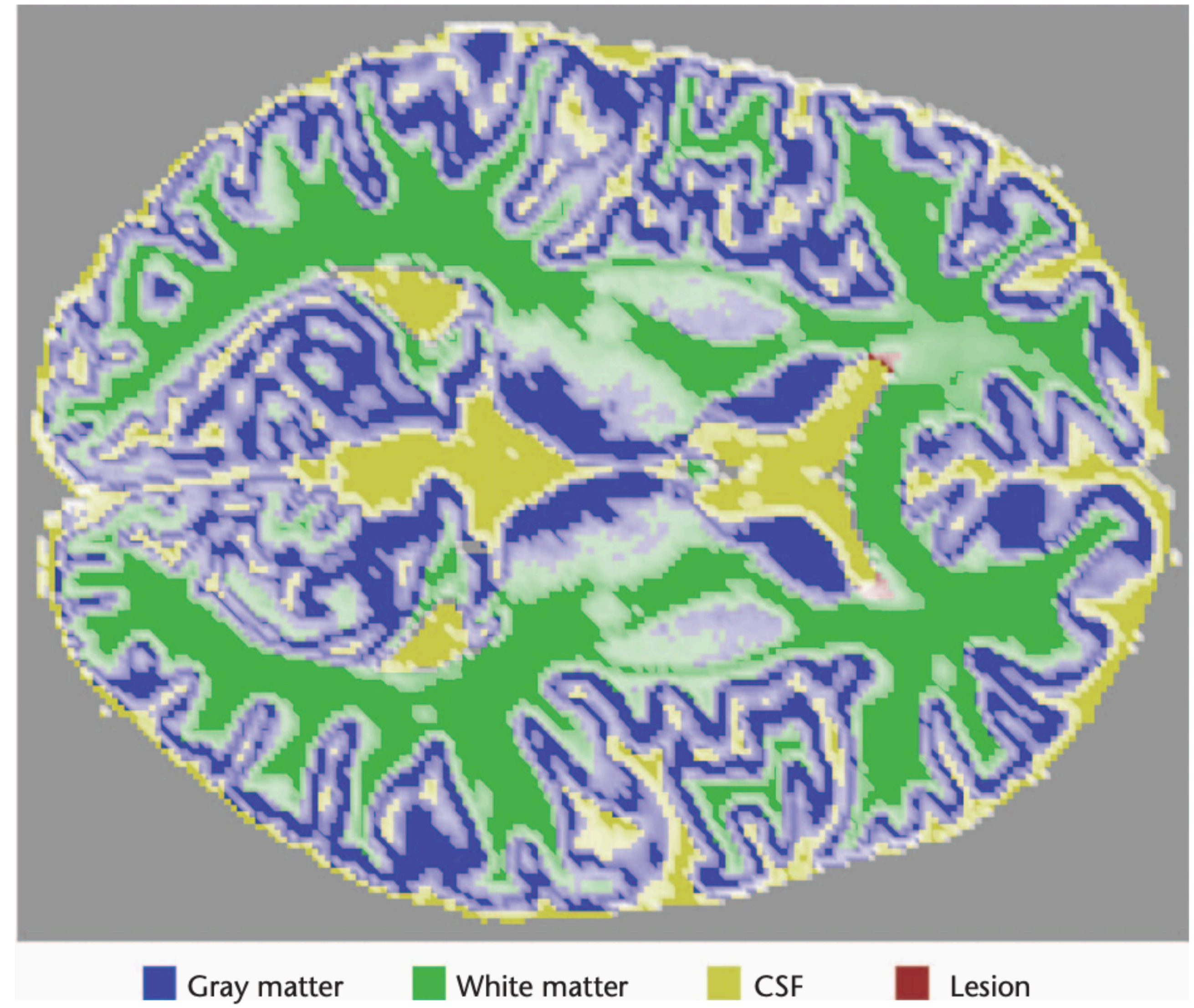


Figure 4. A slice through the combined entropy volume and tagged voxel information. We color-mapped each voxel on the basis of its maximum-probability tissue and the amount of entropy. As a voxel tends toward white, the higher the entropy and the less certainty of a particular tissue type.

Some Case Studies

Not Secure | project-ukko.net

PROJECT

Ukko

Developed as part of the

RESILIENCE PROTOTYPE

in the **EUPORIAS** project

SEASONAL WIND PREDICTIONS FOR THE ENERGY SECTOR

WHY?

Weather forecasts predict future wind conditions only in the range of weeks. Climate predictions look at big changes over years and decades. However, for energy traders, wind farm managers and many others, it would be crucial to understand wind conditions in the next few months.

LEARN MORE

HOW?

Based on sophisticated climate models, we are now able to provide new ways to forecast wind conditions in the next few months.

LEARN MORE

TRY IT OUT

Our interactive browser application allows you to explore the data. Which regions might experience unusual changes in wind activity in the coming months? Find out what our models can tell you.

→ GO

Statistically Quantitative Volume Visualization

Joe M. Kniss*
University of Utah

Robert Van Uitert†
National Institutes of Health

Abraham Stephens‡
University of Utah

Guo-Shi Li§
University of Utah

Tolga Tasdizen
University of Utah

Charles Hansen¶
University of Utah

Abstract

Visualization users are increasingly in need of techniques for assessing quantitative uncertainty and error in the images produced. Statistical segmentation algorithms compute these quantitative results, yet volume rendering tools typically produce only qualitative imagery via transfer function-based classification. This paper presents a visualization technique that allows users to interactively explore the uncertainty, risk, and probabilistic decision of surface boundaries. Our approach makes it possible to directly visualize the combined "fuzzy" classification results from multiple segmentations by combining these data into a unified probabilistic data space. We represent this unified space, the combination of scalar volumes from numerous segmentations, using a novel graph-based dimensionality reduction scheme. The scheme both dramatically reduces the dataset size and is suitable for efficient, high quality, quantitative visualization. Lastly, we show that the statistical risk arising from overlapping segmentations is a robust measure for visualizing features and assigning optical properties.

Keywords: volume visualization, uncertainty, classification, risk analysis

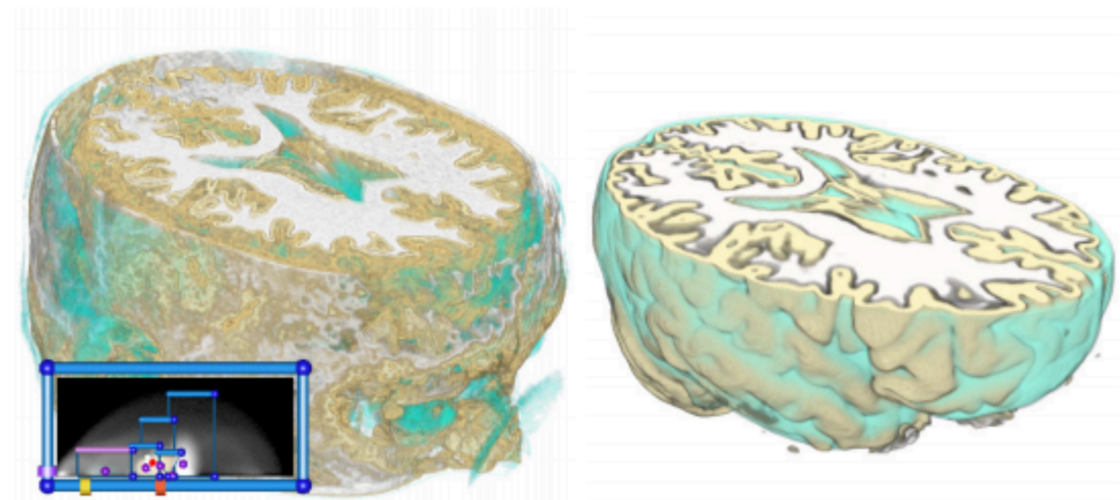
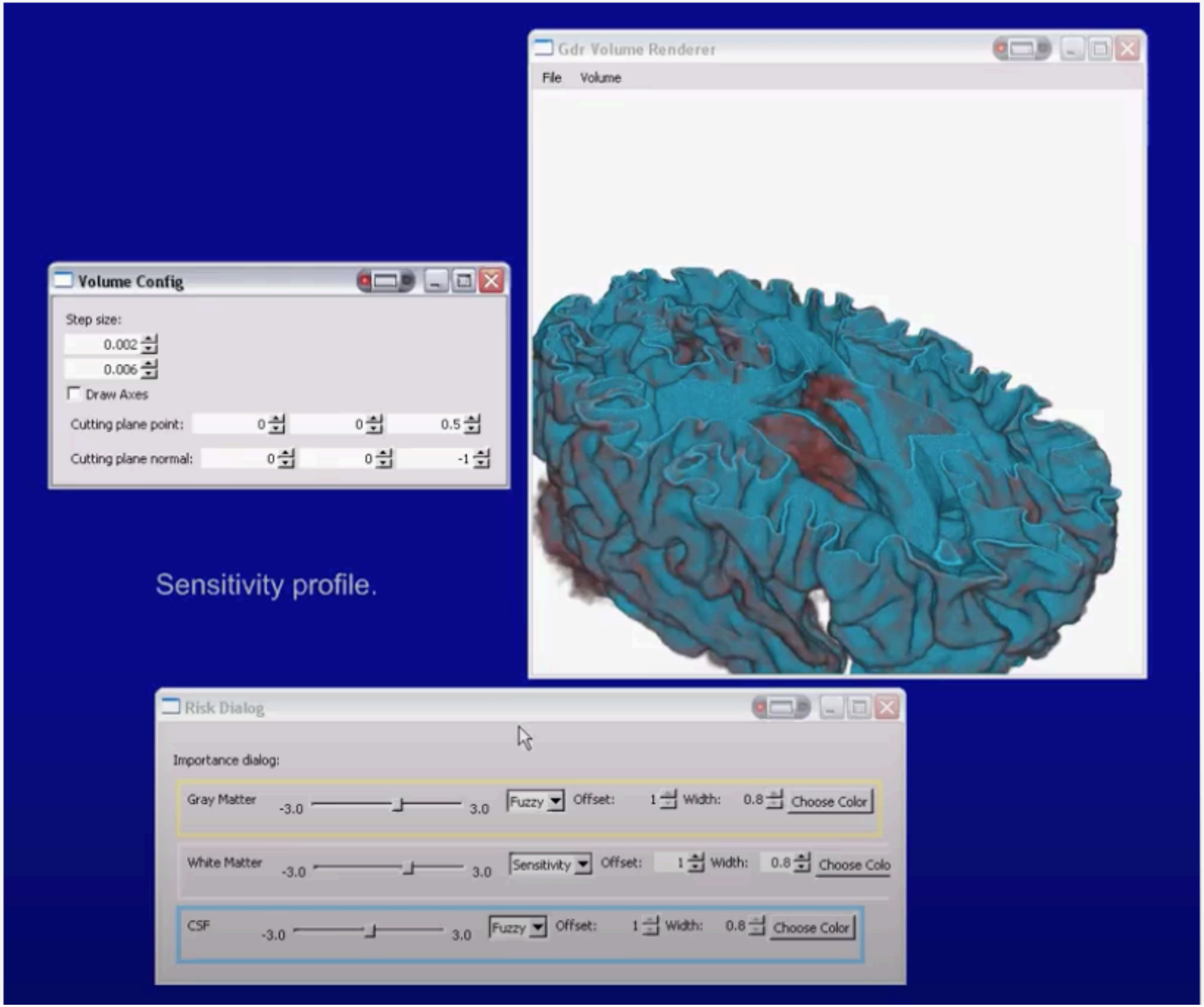


Figure 1: A comparison of transfer function-based classification versus data-specific probabilistic classification. Both images are based on T1 MRI scans of a human head and show fuzzy classified white-matter, gray-matter, and cerebro-spinal fluid. Subfigure A shows the results of classification using a carefully designed 2D transfer function based on data value and gradient magnitude. Subfigure B shows a visualization of the data classified using a fully automatic, atlas-based method that infers class statistics using minimum entropy, non-parametric density estimation [21].

of uncertainty. Transfer functions also tend to be unintuitive



https://www.youtube.com/watch?v=z7AqcZIS-Q4

Noodles: A Tool for Visualization of Numerical Weather Model Ensemble Uncertainty

Jibonananda Sanyal, *Student Member, IEEE*, Song Zhang, *Member, IEEE*, Jamie Dyer, Andrew Mercer, Philip Amburn, *Member, IEEE*, and Robert J. Moorhead, *Senior Member, IEEE*

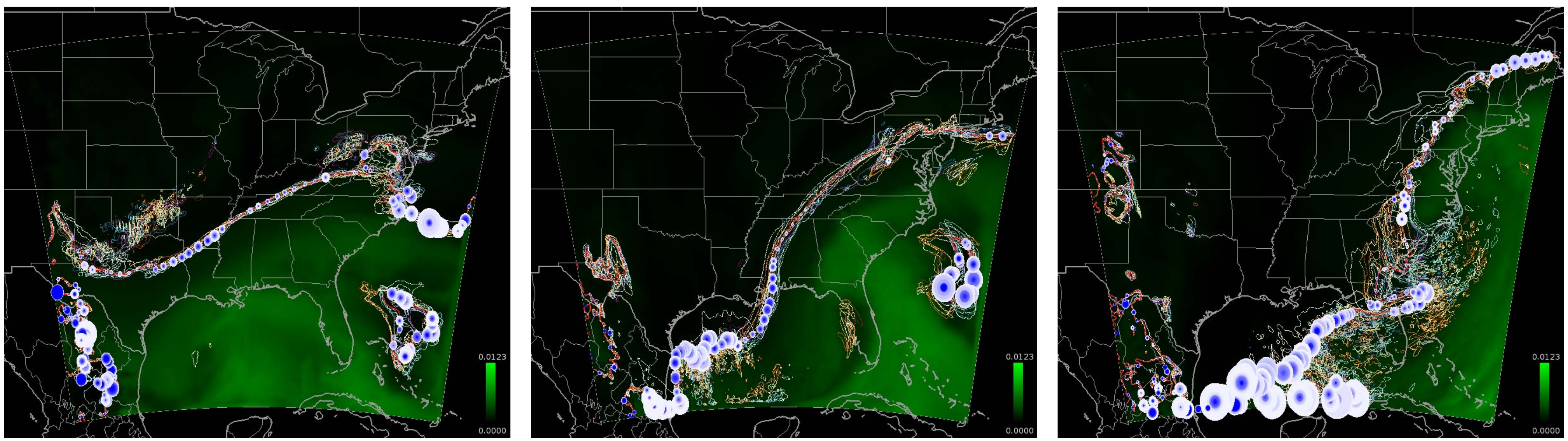


Fig. 1. Visualization of water-vapor mixing ratio at three 12 hour intervals of the simulation starting at 1993-3-13 00:00:00 UTC illustrating the progression of uncertainty by the use of spaghetti plots and uncertainty glyphs.

Abstract—Numerical weather prediction ensembles are routinely used for operational weather forecasting. The members of these ensembles are individual simulations with either slightly perturbed initial conditions or different model parameterizations, or occasionally both. Multi-member ensemble output is usually large, multivariate, and challenging to interpret interactively. Forecast meteorologists are interested in understanding the uncertainties associated with numerical weather prediction; specifically variability between the ensemble members. Currently, visualization of ensemble members is mostly accomplished through spaghetti plots of a single mid-troposphere pressure surface height contour. In order to explore new uncertainty visualization methods, the Weather Research and Forecasting (WRF) model was used to create a 48-hour, 18 member parameterization ensemble of the 13 March 1993 "Superstorm". A tool was designed to interactively explore the ensemble uncertainty of three important weather variables: water-vapor mixing ratio, perturbation potential temperature, and perturbation pressure. Uncertainty was quantified using individual ensemble member standard deviation, inter-quartile range, and the width of the 95% confidence interval. Bootstrapping was employed to overcome the dependence on normality in the uncertainty metrics. A coordinated view of ribbon and glyph-based uncertainty visualization, spaghetti

MISSISSIPPI STATE UNIVERSITY

HPC^2 High Performance Computing COLLABORATORY

MISSISSIPPI STATE
UNIVERSITY

HPC²
High
Performance
Computing
COLLABORATORY

Result-Driven Exploration of Simulation Parameter Spaces for Visual Effects Design

Stefan Bruckner and Torsten Möller, *Member, IEEE*

Abstract—Graphics artists commonly employ physically-based simulation for the generation of effects such as smoke, explosions, and similar phenomena. The task of finding the correct parameters for a desired result, however, is difficult and time-consuming as current tools provide little to no guidance. In this paper, we present a new approach for the visual exploration of such parameter spaces. Given a three-dimensional scene description, we utilize sampling and spatio-temporal clustering techniques to generate a concise overview of the achievable variations and their temporal evolution. Our visualization system then allows the user to explore the simulation space in a goal-oriented manner. Animation sequences with a set of desired characteristics can be composed using a novel search-by-example approach and interactive direct volume rendering is employed to provide instant visual feedback. A user study was performed to evaluate the applicability of our system in production use.

Index Terms—Visual exploration, visual effects, clustering, time-dependent volume data.

1 Introduction

Physically-based simulation is gaining increasing popularity for generating realistic animations of water, smoke, explosions, and related phenomena using computer graphics. Common modeling and animation software packages include built-in fluid dynamics simulators or offer this functionality via add-on modules. These existing tools frequently allow the user to modify the simulation parameters via standard controls such as sliders or numeric input fields. It is difficult, however, to predict the influence of changing one or several of these values. Depending on the exact scene setup, effects may be global or remain rather localized, both in space and time. Even small changes can dramatically affect the appearance of the resulting animation. Graphics artists, who aim to produce a particular visual result, therefore typically have to resort to a cumbersome and time-consuming trial-and-error approach. Moreover, as the simulation process is computationally expensive, interactive visual feedback is frequently not available. While recent advances in real-time fluid simulation help by reducing the simulation time [32, 41], the underlying problem remains: there is virtually no guidance in exploring a vast parameter space.

to make the design process considerable less labor intensive. We also present a novel approach for clustering time-dependent volume data generated by sampling a high-dimensional parameter space. Furthermore, the paper describes new visualization and interaction techniques for volumetric time sequences designed to meet the requirements of graphics artists. Finally, we present a user study performed to evaluate the practical applicability of our approach to visual effects design.

2 Related Work

The visualization of general time-oriented data is an extensive field of research and Aigner et al. [1] as well as Andrienko et al. [5] provide comprehensive surveys. Our work focuses on the visualization of time-varying volume data, a topic which has been intensively studied in the context of science and engineering data [24]. In many cases, the user is interested in tracking certain features over time which can be difficult in animations of complex data. One approach is to consider the time series as a four-dimensional scalar field. Hanson and Heng [17] introduced general techniques for visualizing surfaces and

[redacted]

Lec26 Reading

- Jigsaw: Supporting Investigative Analysis through Interactive Visualization. John T. Stasko, Carsten Görg, Zhicheng Liu. Information Visualization 7(2): 118-132 (2008).
- Bubble Sets: Revealing Set Relations with Isocontours over Existing Visualizations. Christopher Collins, Gerald Penn, M. Sheelagh T. Carpendale. IEEE Trans. Vis. Comput. Graph. 15(6): 1009-1016 (2009)

Reminder
Assignment 06

Assigned: Monday, April 10
Due: Monday, April 24, 4:59:59 pm

Reminder
Project Milestones 03/04

Assigned: Wednesday, March 29

03 (Talk) Due: Wednesday, April 26, 4:59:59 pm

04 (Report) Due: Wednesday, May 3, 4:59:59 pm