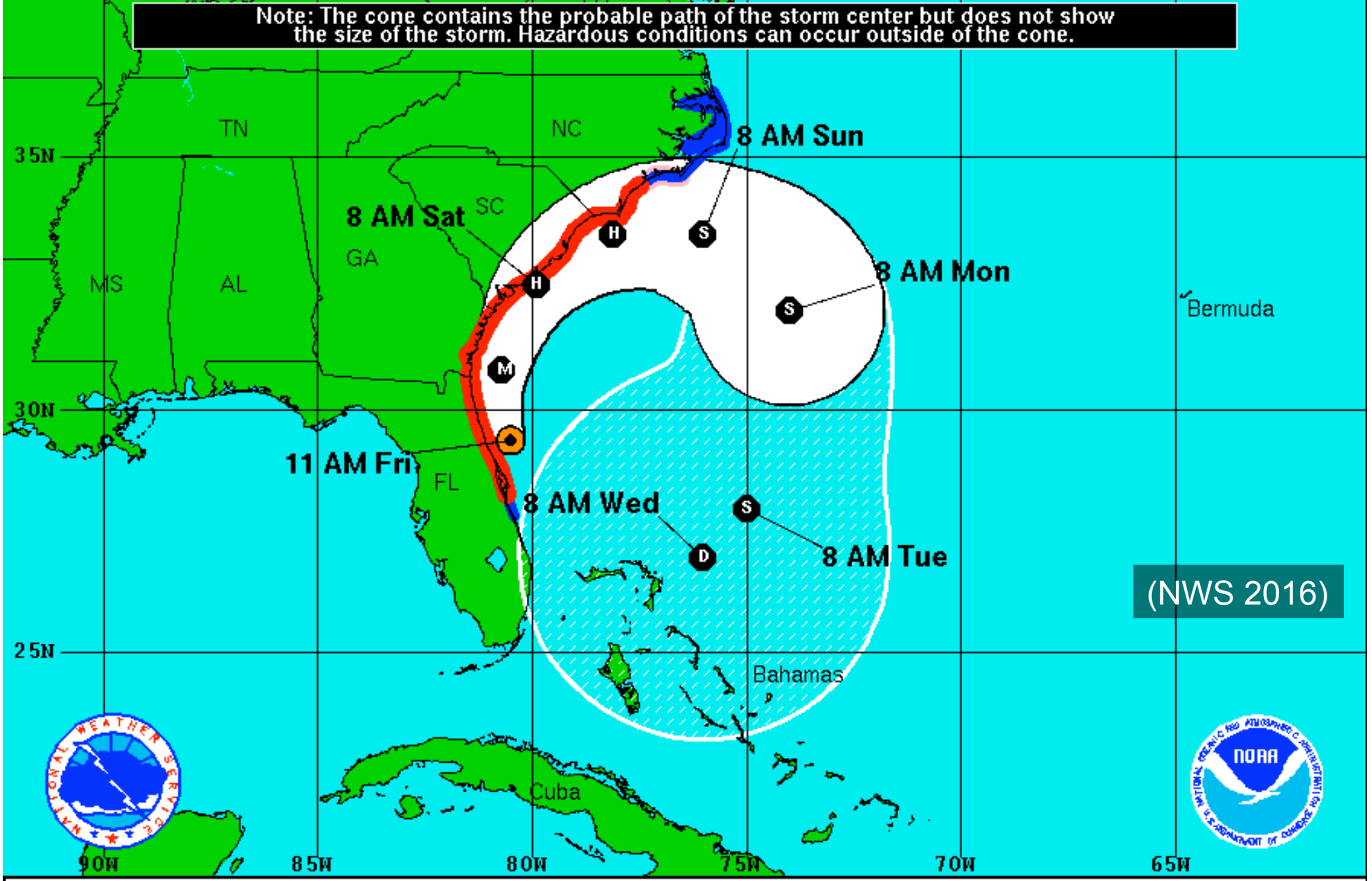
# DATA ANALYSIS UNDER UNCERTAINTY

Slides by Christoph Kinkeldey

Inria




**Hurricane Matthew**
*Friday October 7, 2016*
*11 AM EDT Advisory 38*
NWS National Hurricane Center

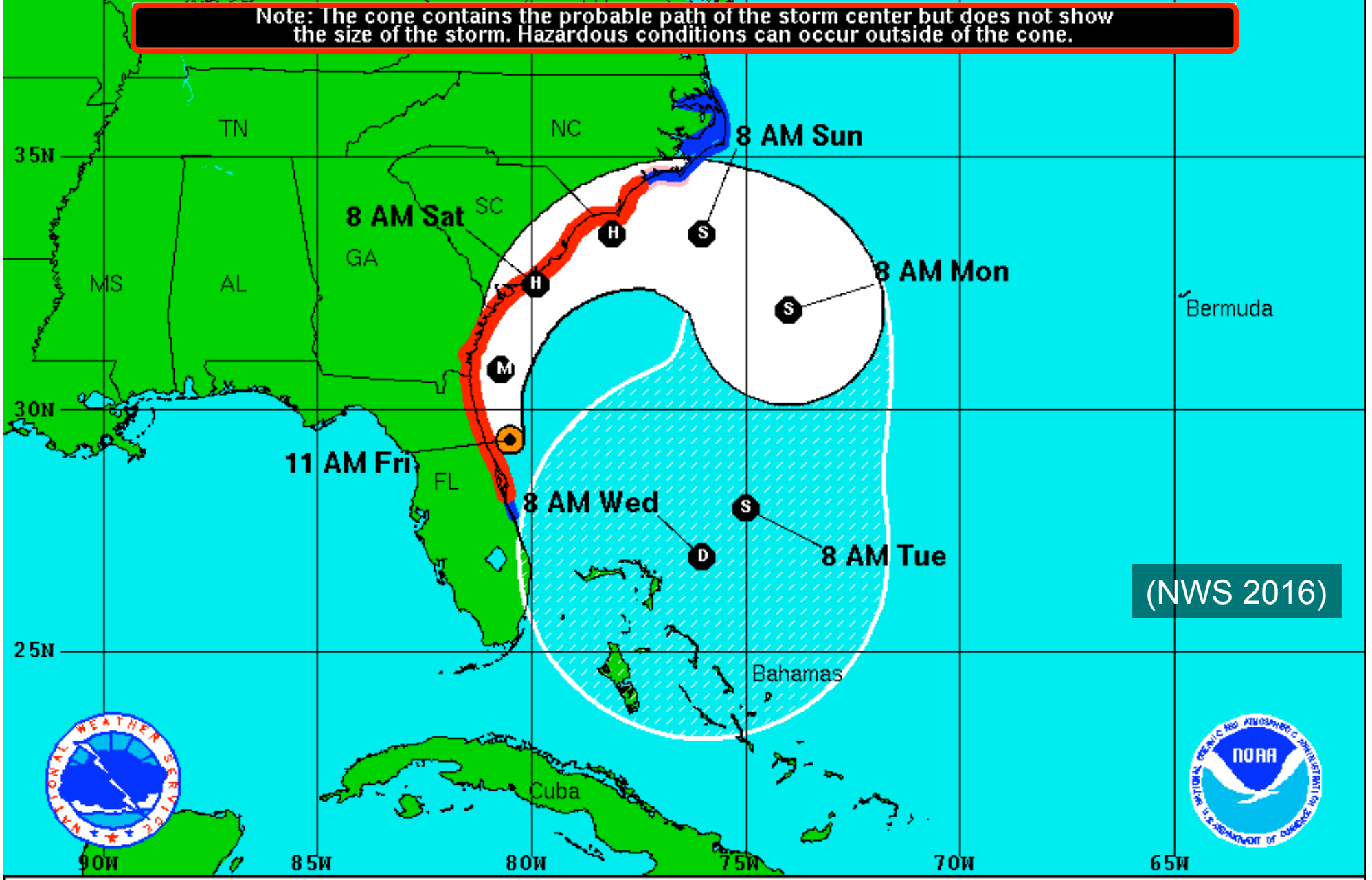
**Current Information:**
Center Location 29.4 N 80.5 W
Max Sustained Wind 120 mph
Movement NNW at 12 mph

**Forecast Positions:**
Tropical Cyclone  Post-Tropical
Sustained Winds: D < 39 mph
S 39-73 mph H 74-110 mph M > 110mph

**Potential Track Area:**
Day 1-3  Day 4-5

**Watches:**
Hurricane  Trop.Storm

**Warnings:**
Hurricane  Trop.Storm




**Hurricane Matthew**
*Friday October 7, 2016*
*11 AM EDT Advisory 38*
NWS National Hurricane Center

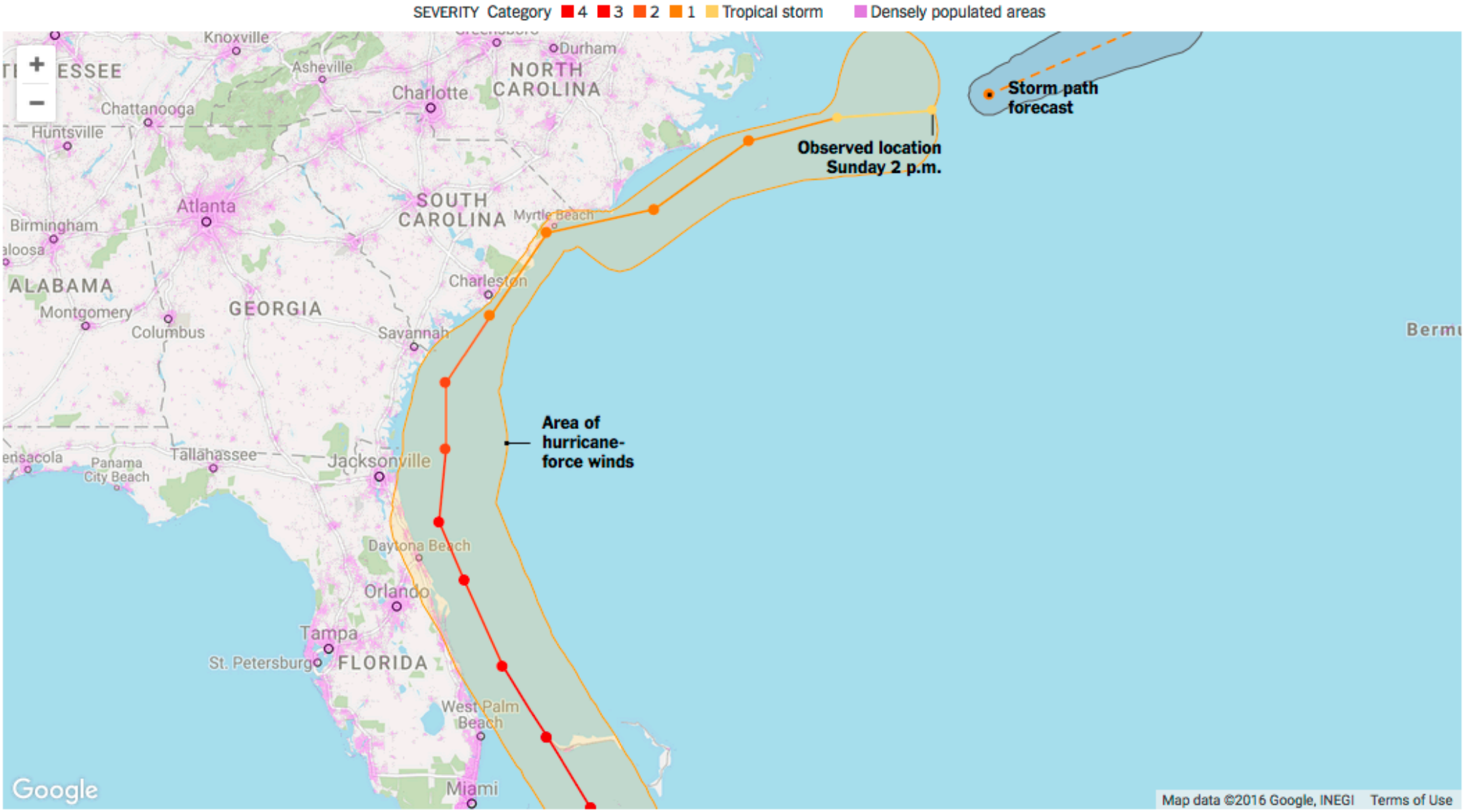
**Current Information:**
Center Location 29.4 N 80.5 W
Max Sustained Wind 120 mph
Movement NNW at 12 mph

**Forecast Positions:**
Tropical Cyclone  Post-Tropical
Sustained Winds: D < 39 mph
S 39-73 mph H 74-110 mph M > 110mph

**Potential Track Area:**
Day 1-3  Day 4-5

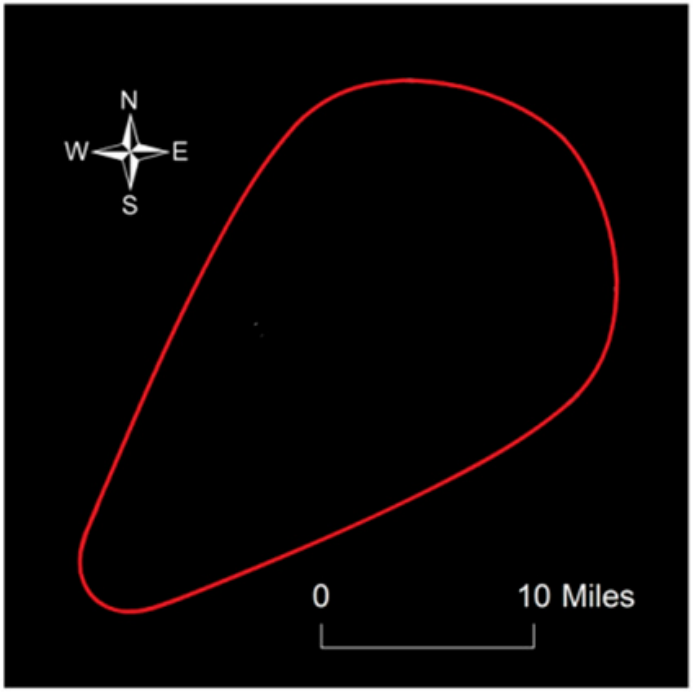
**Watches:**
Hurricane  Trop.Storm

**Warnings:**
Hurricane  Trop.Storm

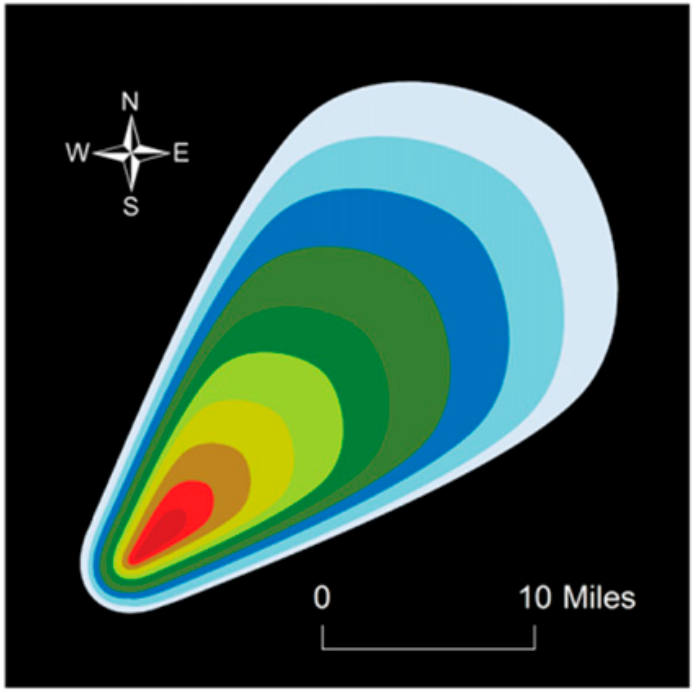



Sources: National Weather Service; LandScan population database

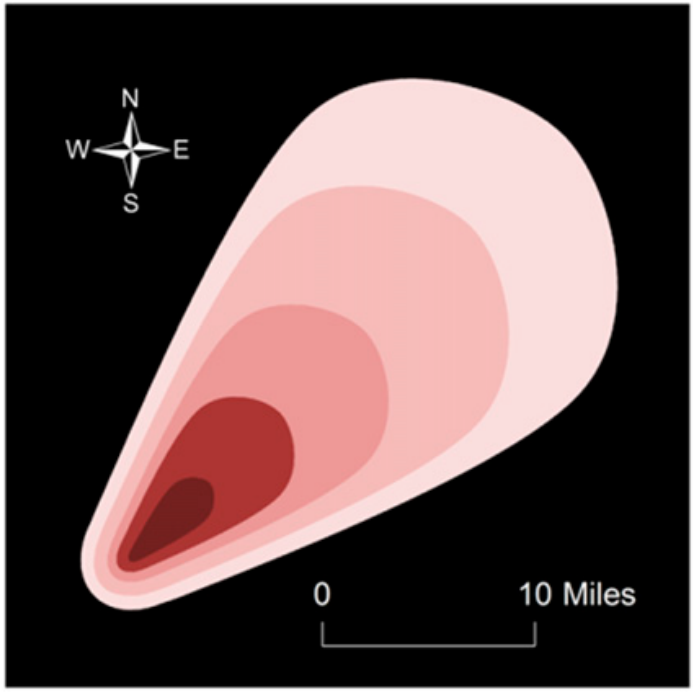
(NYTimes 2016)




1a.) Original

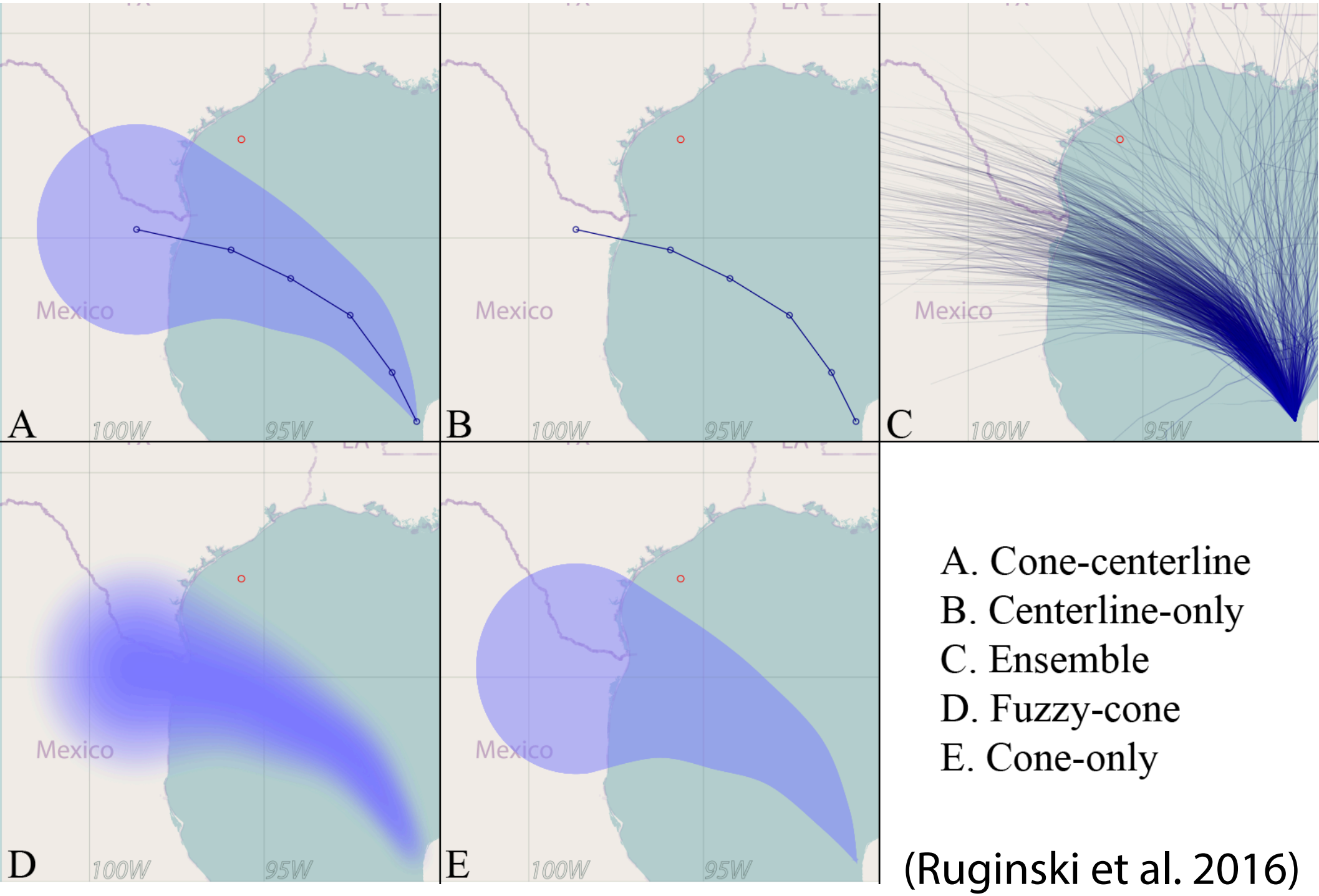



1b.) Spectral




1c.) Red Gradient

(Ash et al. 2014)




A. Cone-centerline
B. Centerline-only
C. Ensemble
D. Fuzzy-cone
E. Cone-only

(Ruginski et al. 2016)

# Motivation

Important to know about uncertainty when

- analyzing and understanding data
- making decisions based on data

# Geodata Uncertainty // Definition

Lack of knowledge about:

- objects of the real world due to
  - erroneous measurement,
  - vague definitions and concepts or
  - unknown and ambiguous meaning
- effects of transformations performed on the data
- the data's suitability for the intended application

*(Leyk 2005)*

# Uncertainty // Definition

Uncertainty and error are different concepts!

Error: True value is known

Uncertainty: True value is not known

Uncertainty always describes a lack of knowledge

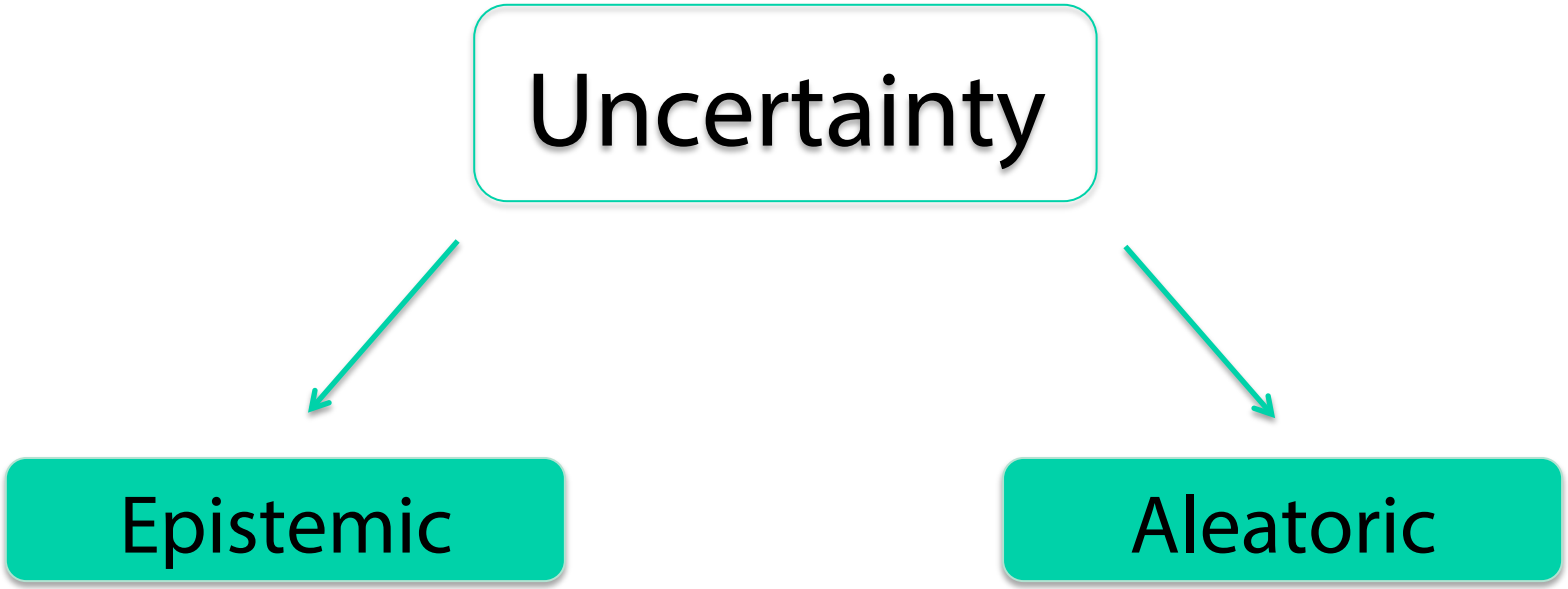
➔ no ‘correct’ values of uncertainty

# Uncertainty // Sources

- Variability in nature
- Deficiency in measurement methods and equipment (resolution, accuracy…)
- Deficiency in modeling (imprecision, lack of complexity…)
- Insufficient of conflicting information
- Others, e.g. uncertainty introduced when visualizing

# Uncertainty // Categories

New challenges regarding....



# Uncertainty // Categories

Epistemic uncertainty:

- “systematic uncertainty“
- Things we could in principle know but in practice we do not know
- e.g. insufficient measurement or modeling, missing data

Reducible: can be minimized by more accurate measurement, better models, more data...

# Uncertainty // Categories

Aleatoric uncertainty:

- “statistical“ uncertainty
- Unknowns that can differ each run
- e.g. outcome from throwing a dice

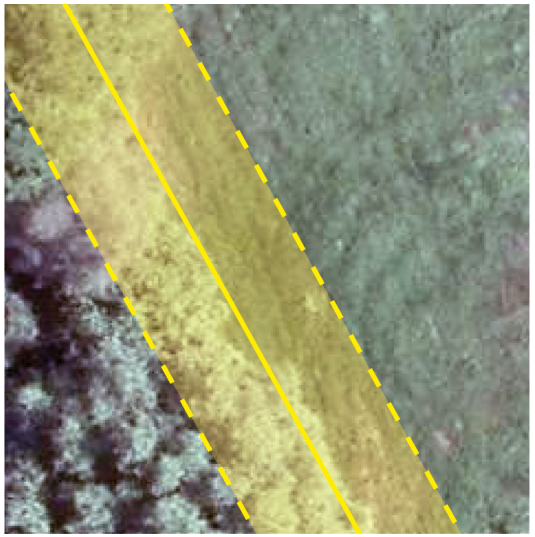
Irreducible: can NOT be minimized through improvements of measurements or models

# Geodata Uncertainty // Categories

## Attribute uncertainty

Decidious forest?
Mixed forest?

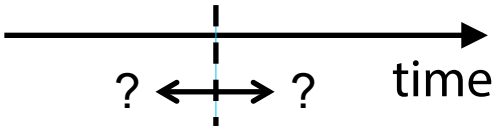
## Geometric uncertainty



Position of boundary?

## Temporal uncertainty

Valid for what time interval?

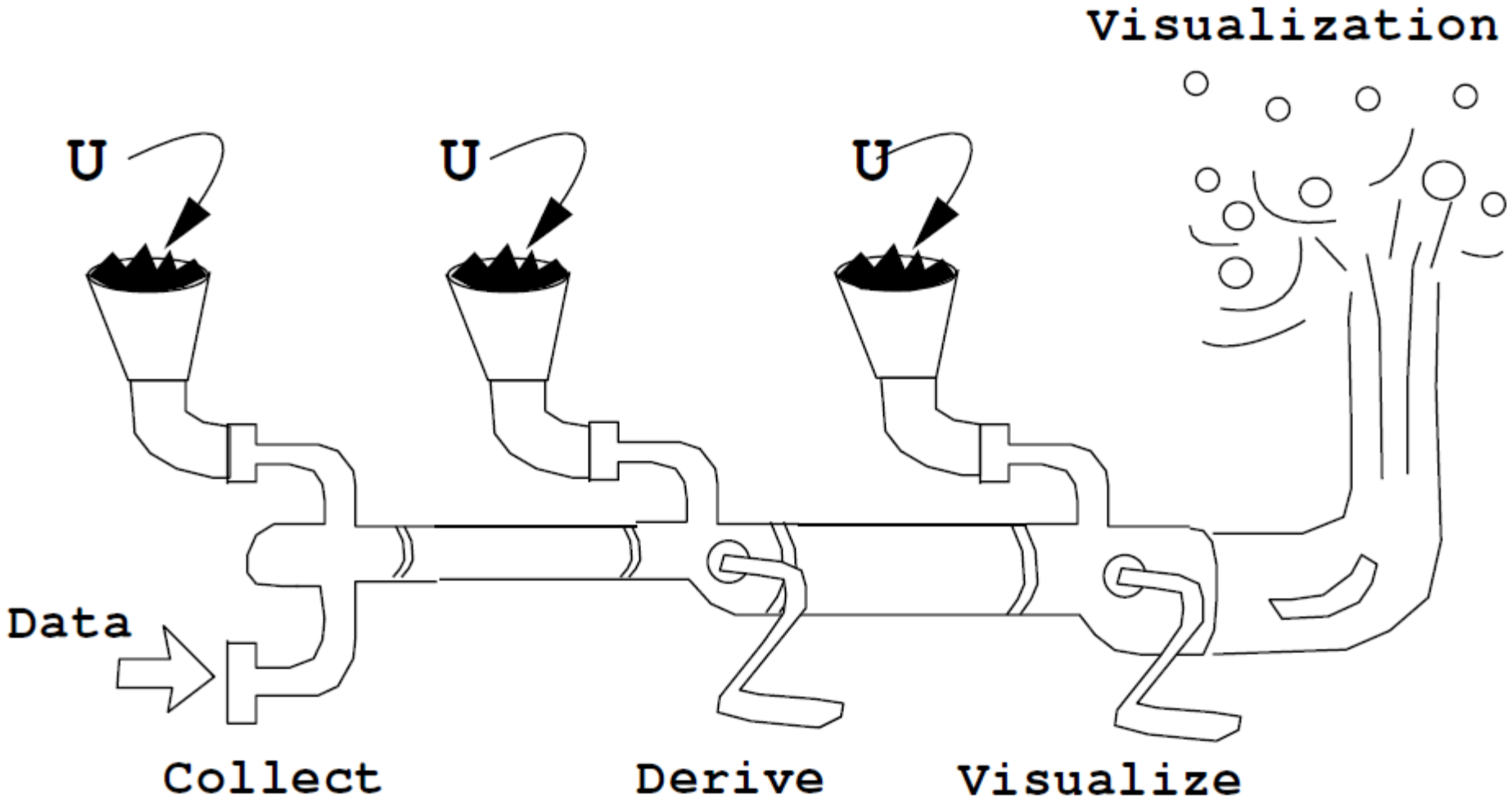


# Geodata Uncertainty

| Category | Attribute Examples | Location Examples | Time Examples |
|---|---|---|---|
| **Accuracy/error** | counts, magnitudes | coordinates, buildings | +/- 1 day |
| **Precision** | nearest 1000 | 1 degree | once per day |
| **Completeness** | 75% of people reporting | 20% of photos flown | 2004 daily/12 missing |
| **Consistency** | multiple classifiers | from / for a place | 5 say Mon; 2 say Tues |
| **Lineage** | transformations | #/quality of input sources | # of steps |
| **Currency** | census data | age of maps | $C = T_{present} - T_{info}$ |
| **Credibility** | U.S. analyst interpretation of financial records <…> informant report of financial transaction | direct observation of training camp <…> e-mail intercep-tion with reference to training camp | time series air photos indicating event time <… > anonymous call predicting event time |
| **Subjectivity** | fact <…> guess | local <…> outsider | expert <…> trainee |
| **Interrelatedness** | all info from same author | source proximity | time proximity |

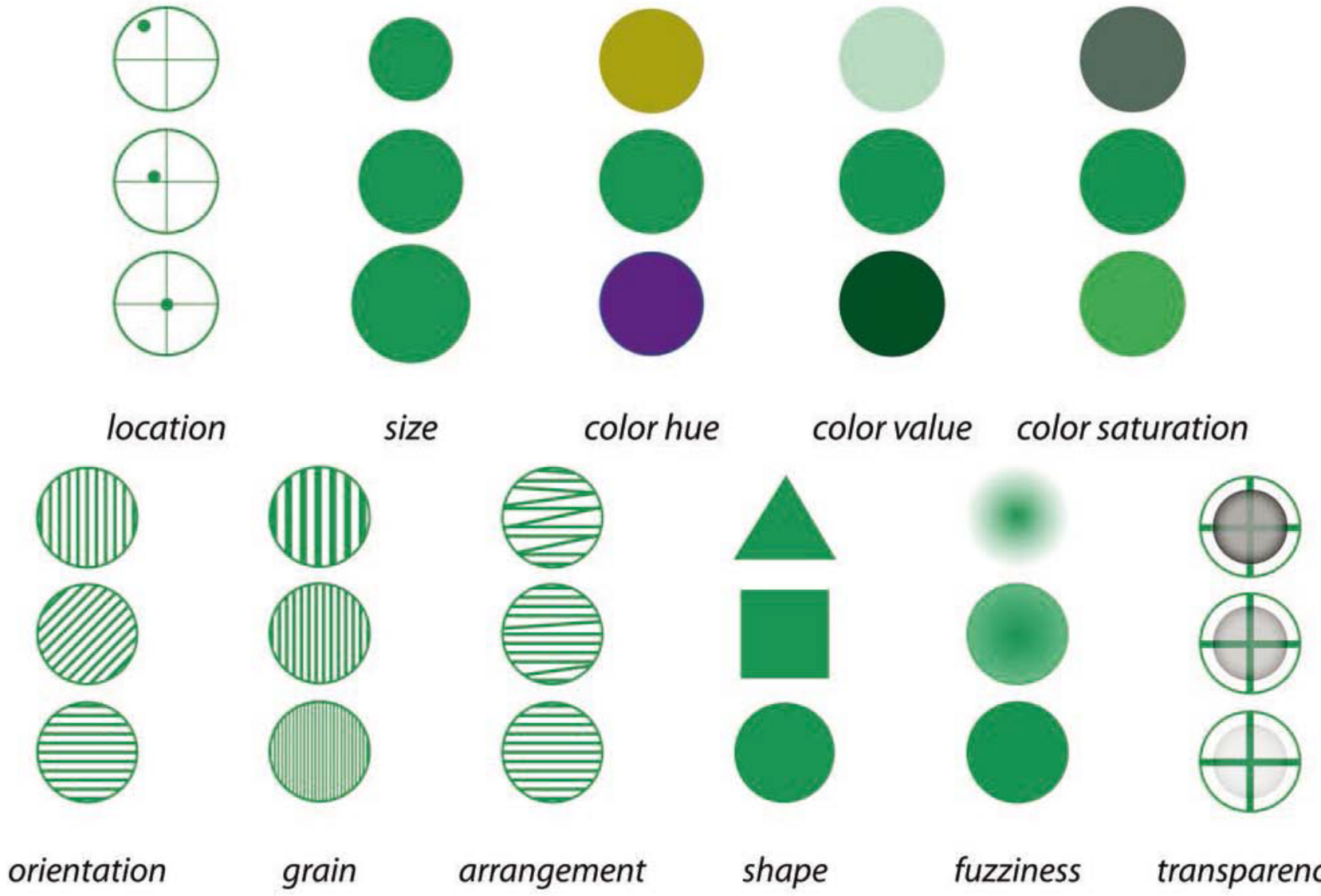
*(MacEachren et al. 2005)*

# Uncertainty Visualization Pipeline



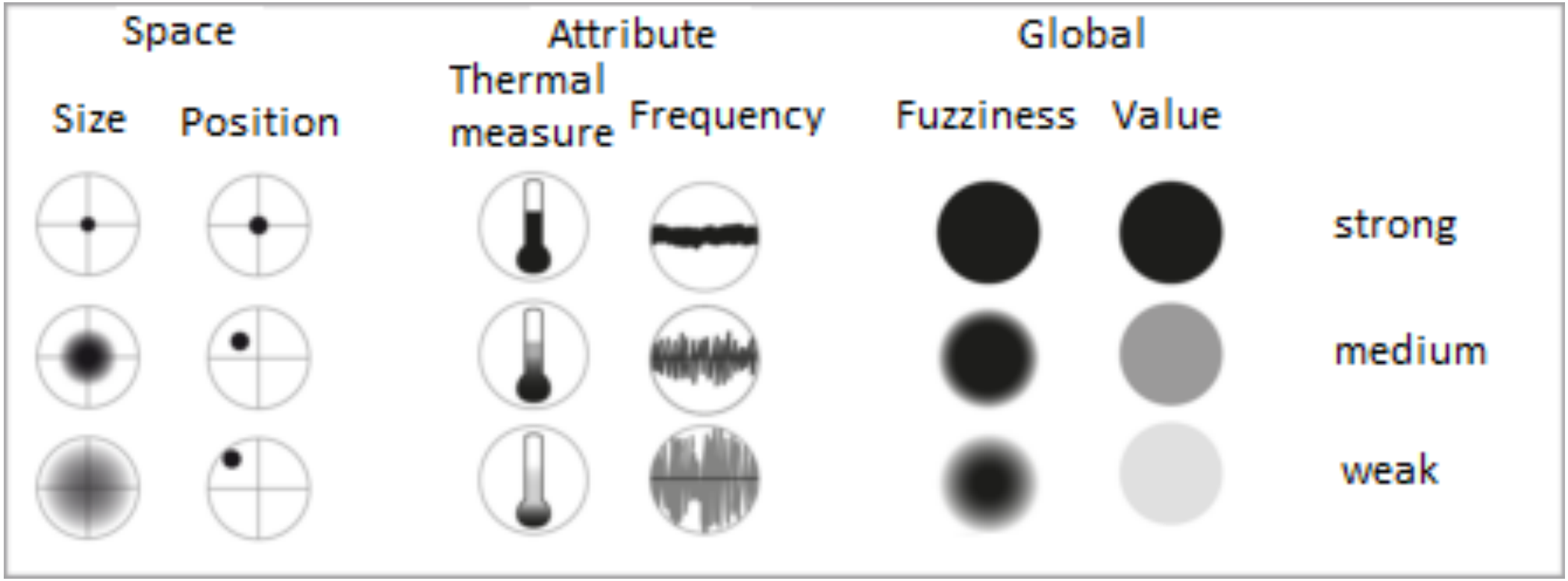

*(Pang et al. 1997)*

# Visual Variables



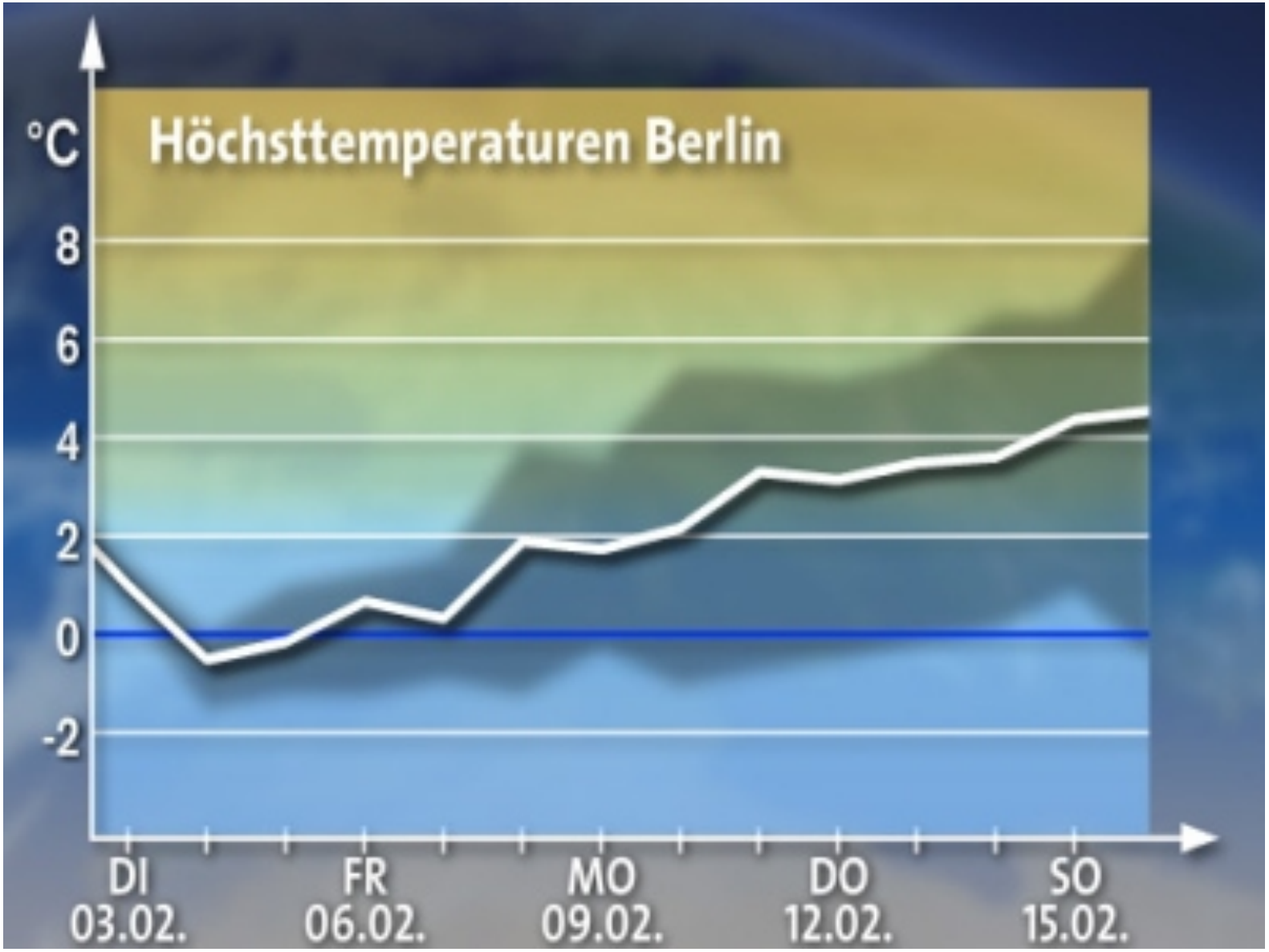

*(MacEachren et al. 2012)*

# Visual Variables



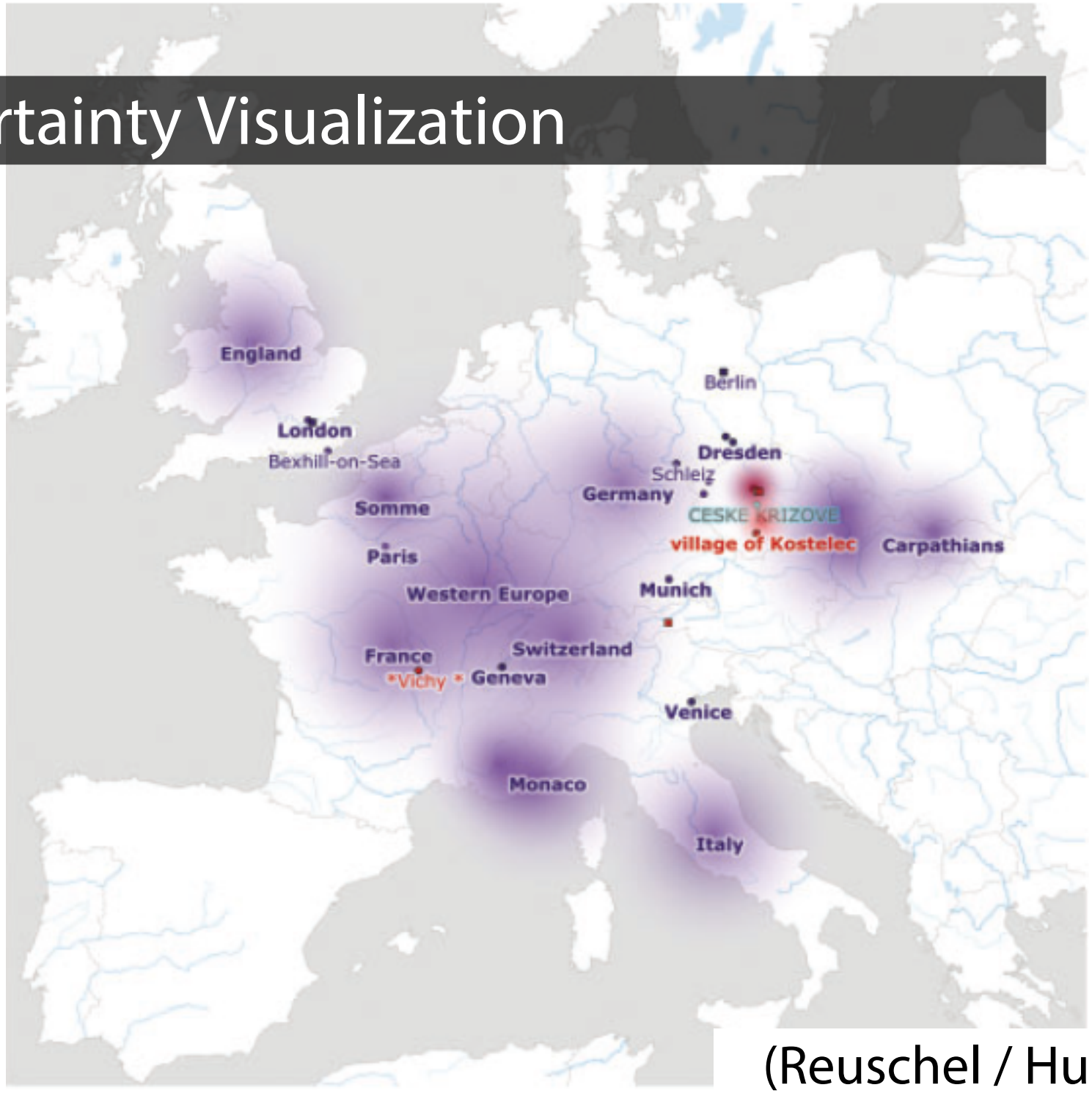

(Seccia et al. 2014)

# Uncertainty Visualization



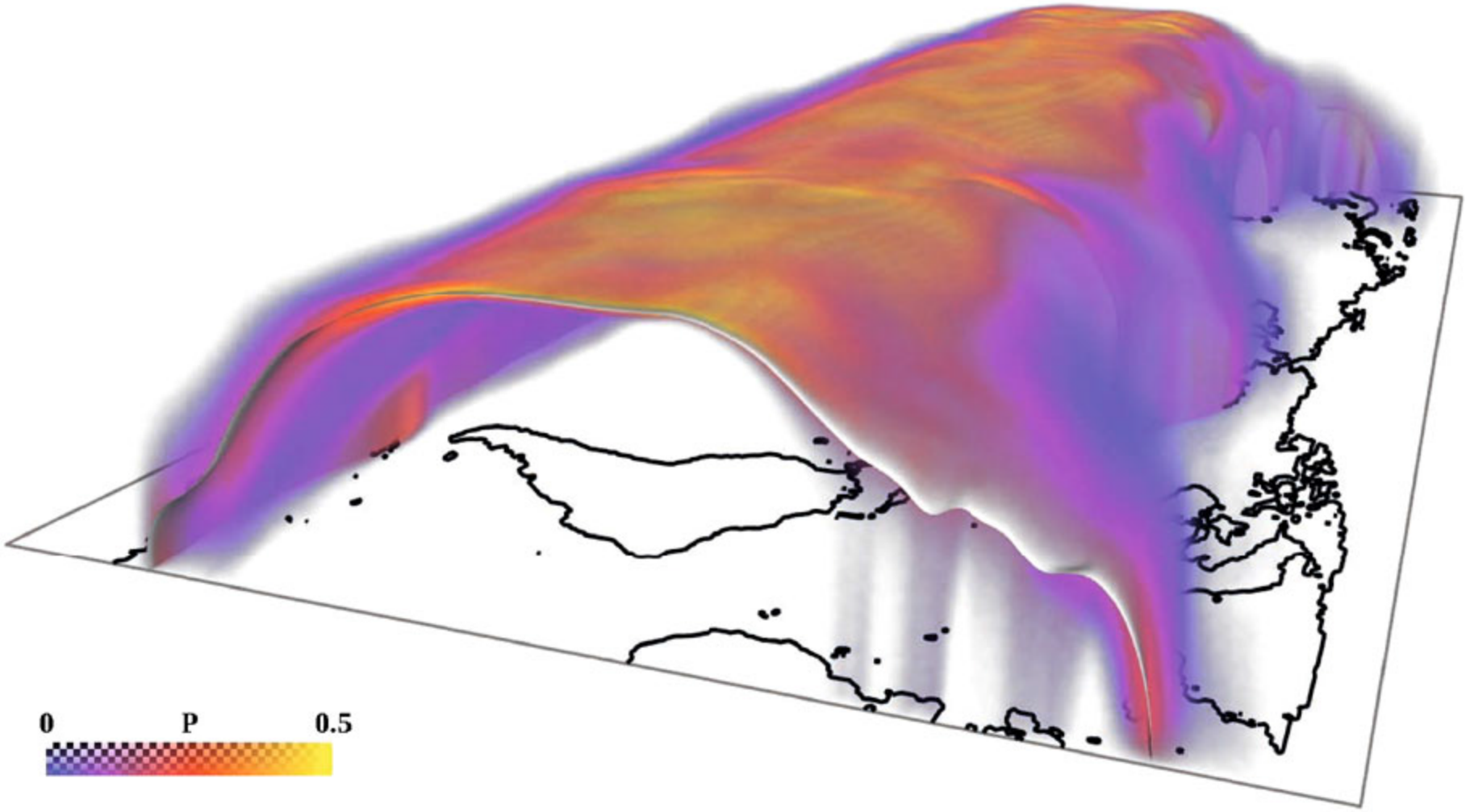

http://wetterbilder.daserste.de/

# Uncertainty Visualization



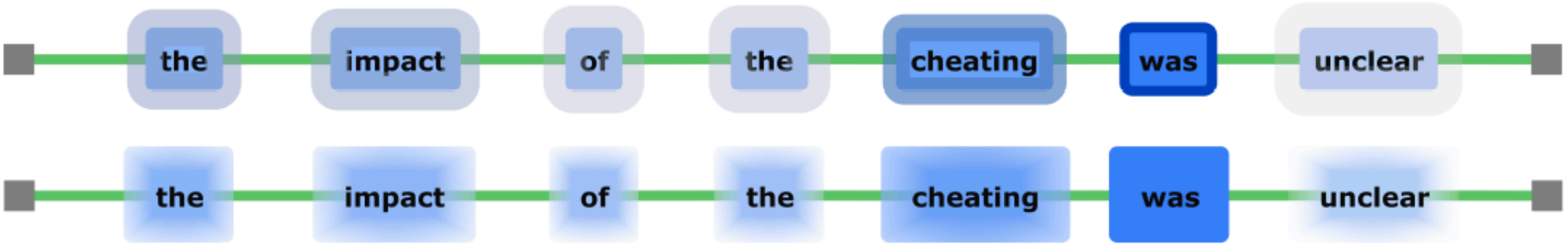

(Reuschel / Hurni 2011)

# Uncertainty Visualization




*(Pöthkow et al. 2011)*

# Uncertainty Visualization

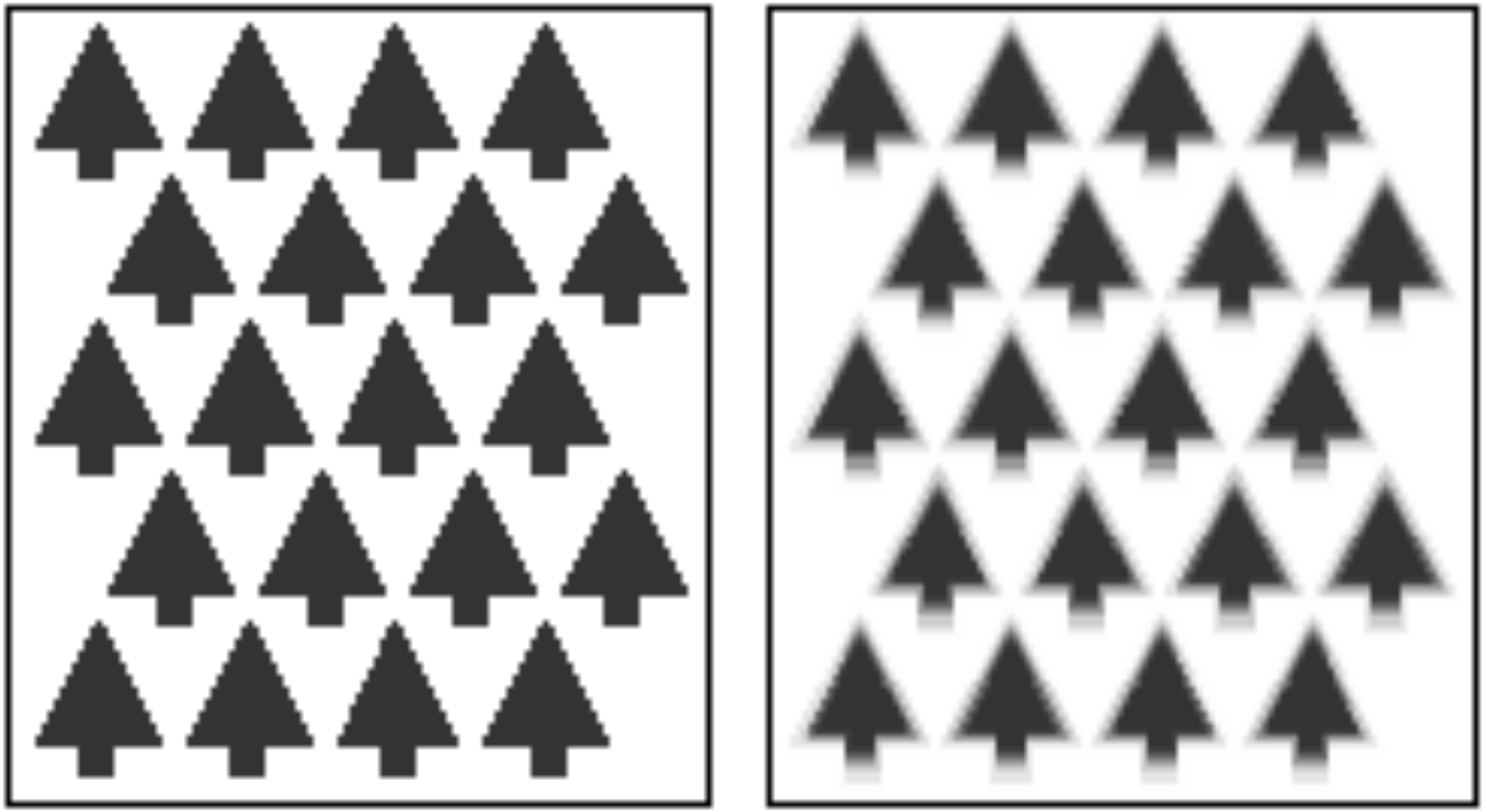



(Collins et al. 2007)

# Metaphors

Metaphors (fog, clarity…) are deemed to increase intuitiveness of uncertainty displays

# “Clarity“



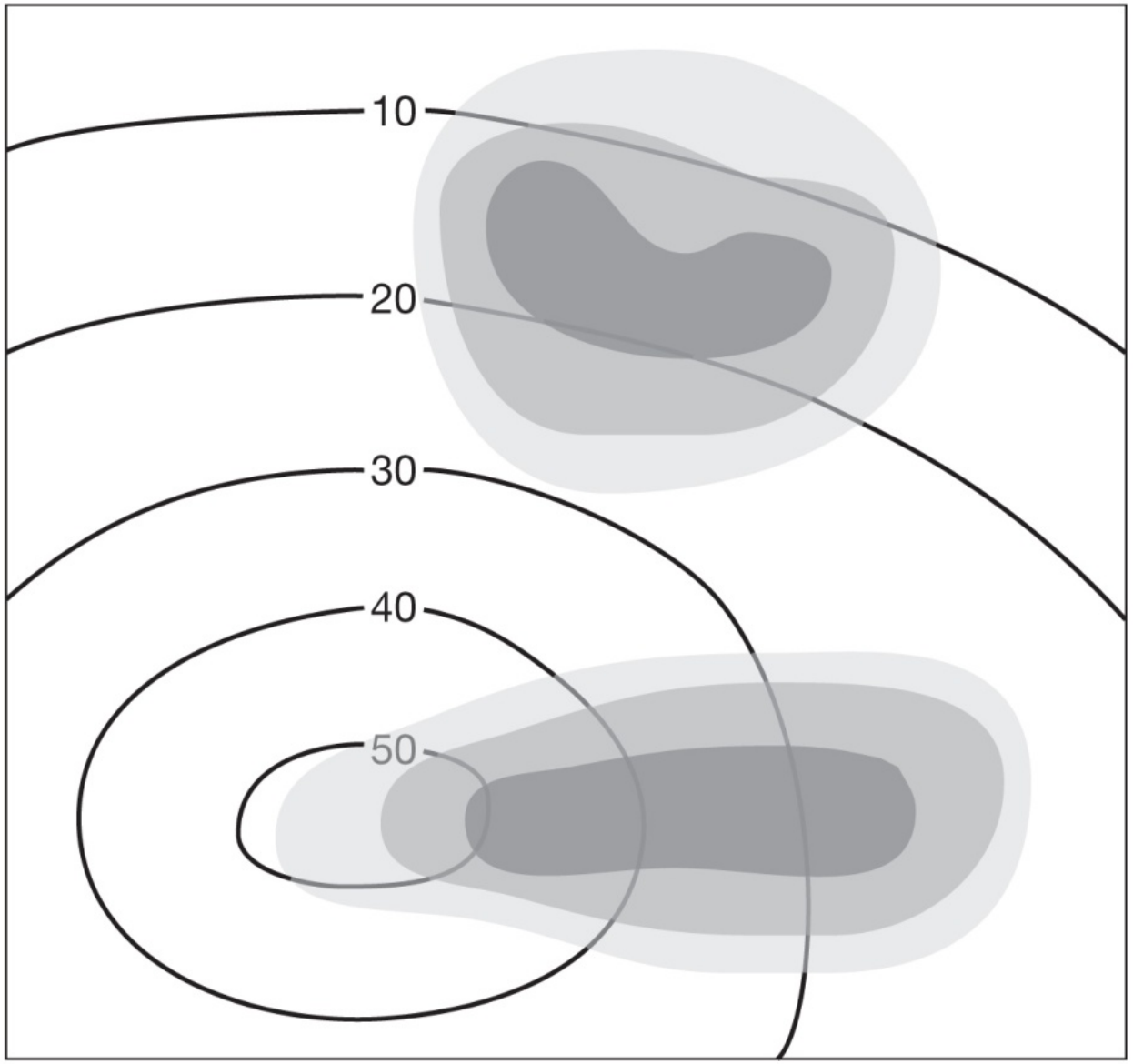
(MacEachren 1992)

# Fuzzy border




(MacEachren 1992)

# Transparency (“fog“)






(MacEachren 1992)

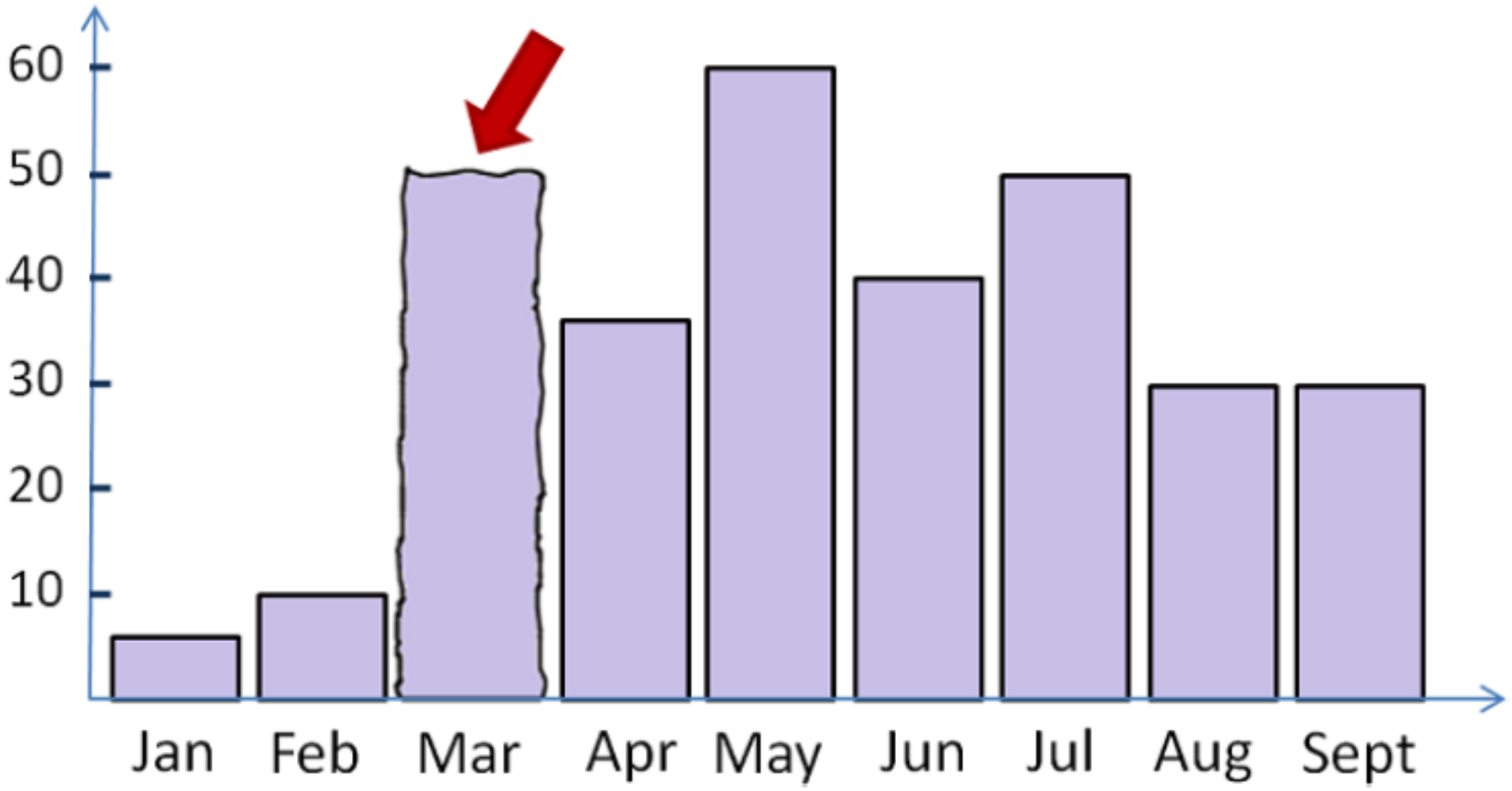
# Resolution

High Resolution

Low Resolution

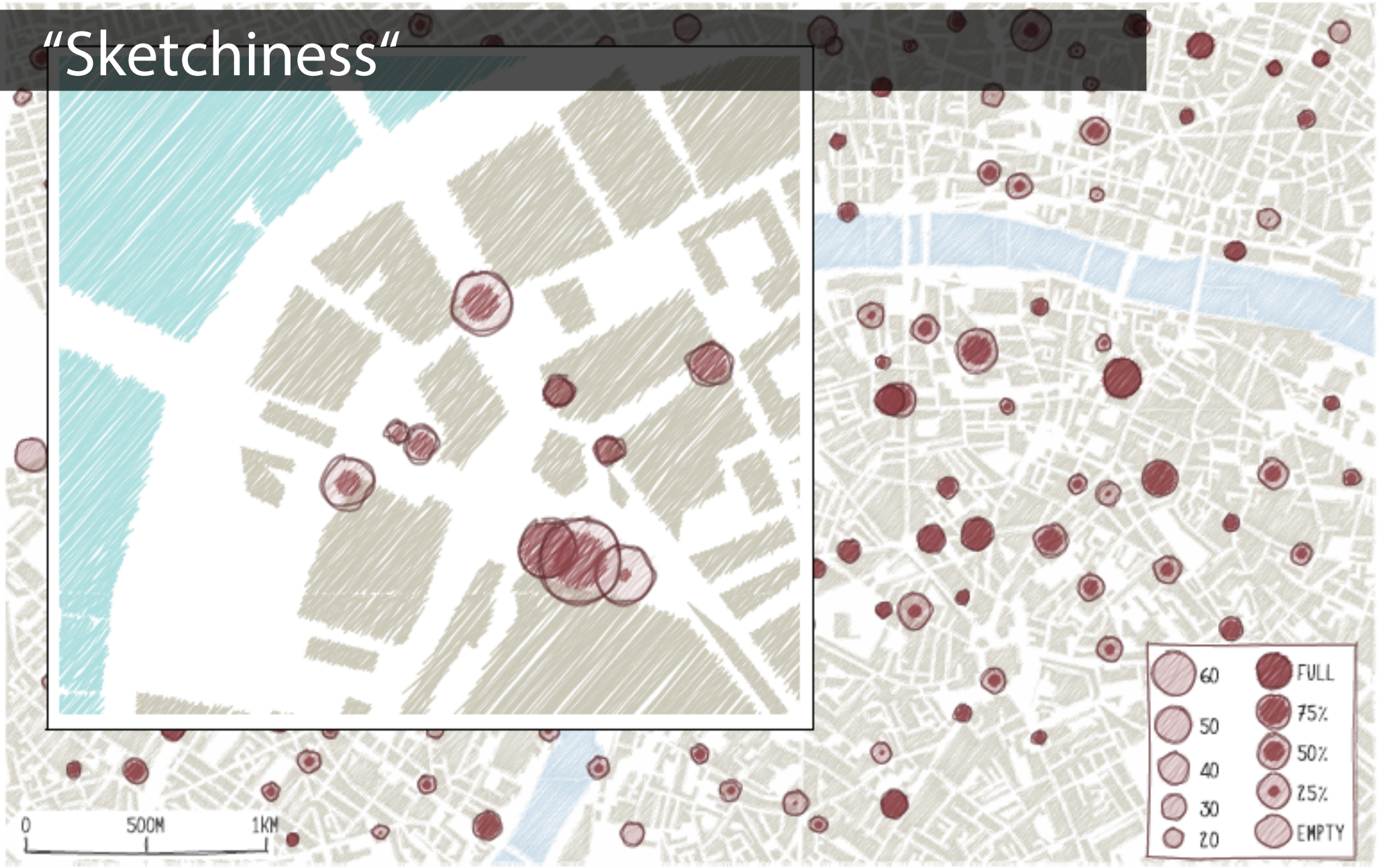
(MacEachren 1992)

# “Sketchiness“




(Boukhelifa et al. 2012)

# “Sketchiness“




*(Wood et al. 2012)*

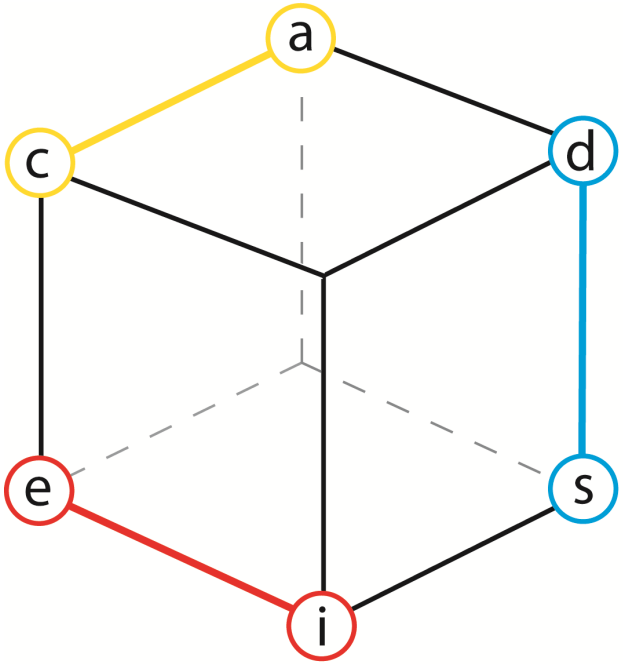
# Uncertainty Visualization

A wide range of techniques exist but:

*What are the best ways to depict uncertainty visually for analytical tasks?*

→ accuracy, speed, intuitiveness, user confidence, preference

# Uncertainty Vis Cube (UVis$^3$)



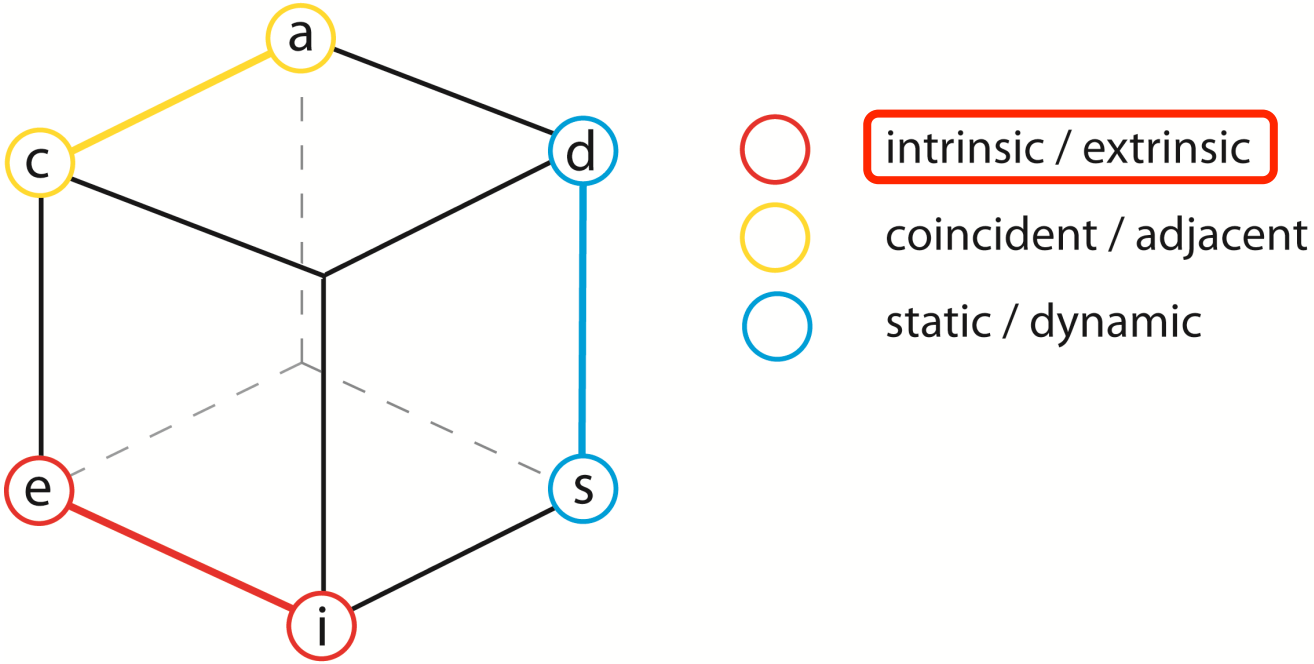

intrinsic / extrinsic

coincident / adjacent

static / dynamic

*(Kinkeldey et al. 2014b)*

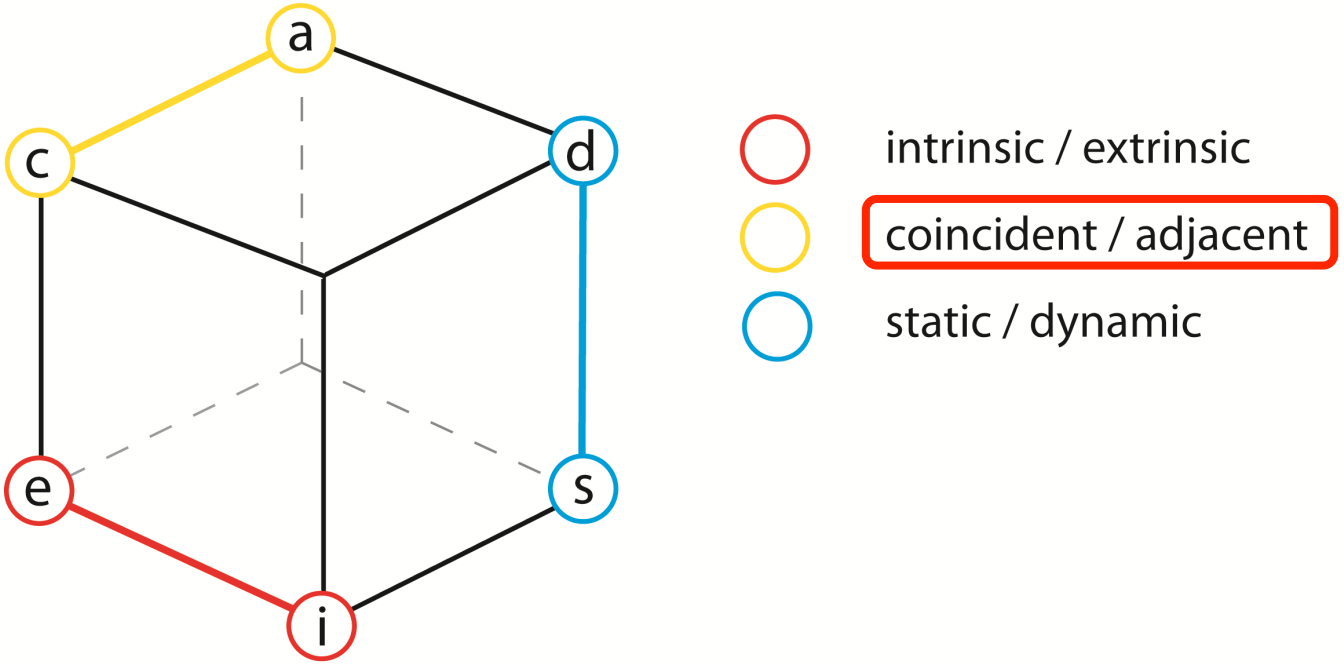
# Uncertainty Vis Cube (UVis$^3$)




- Intrinsic: Existing objects in the display are manipulated
- Extrinsic: Uncertainty is represented by additional objects in the display, e.g. symbols or grids

*(Kinkeldey et al. 2014b)*

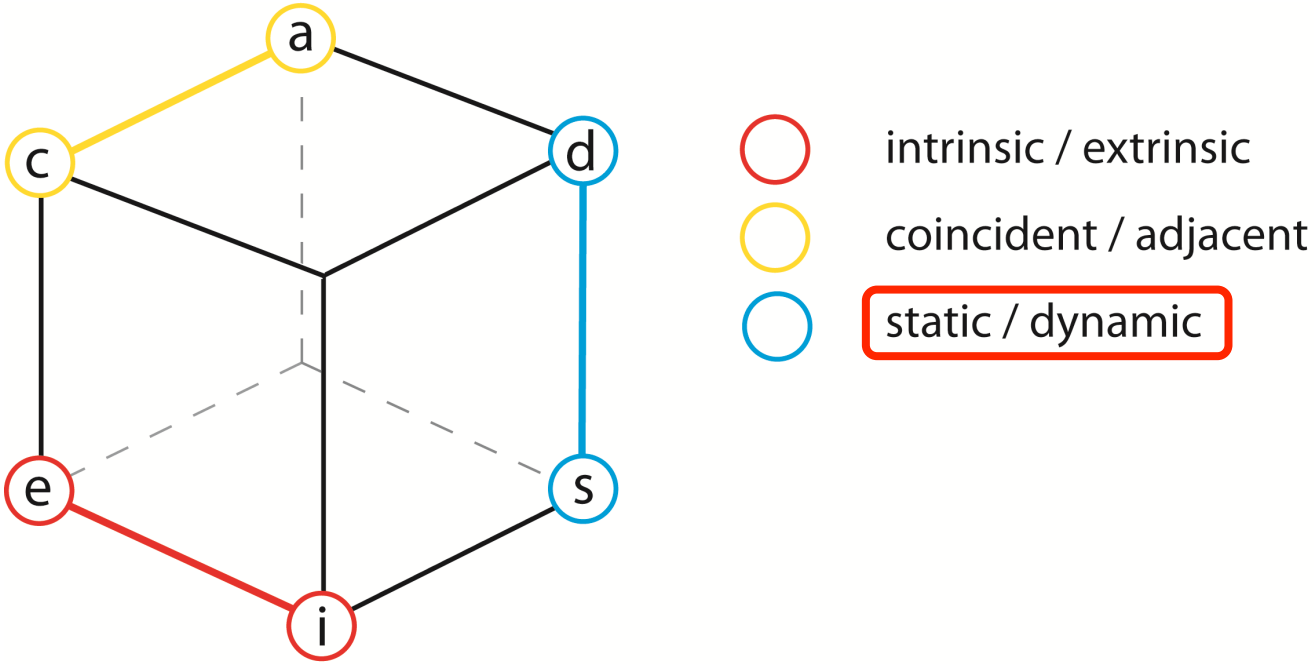
# Uncertainty Vis Cube (UVis$^3$)




- Coincident: data + uncertainty in one view
- Adjacent: data + uncertainty in separate views

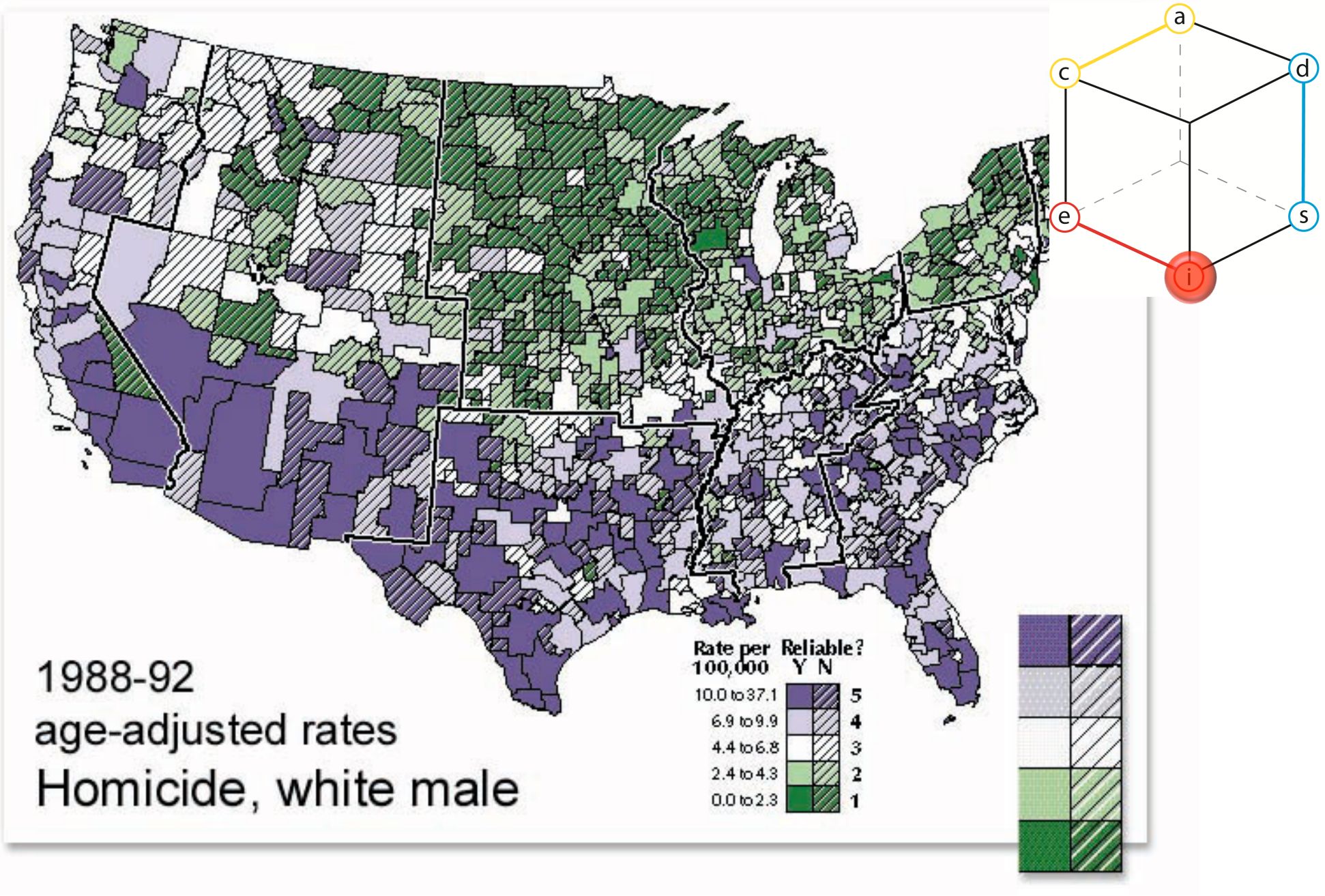
*(Kinkeldey et al. 2014b)*

# Uncertainty Vis Cube (UVis$^3$)

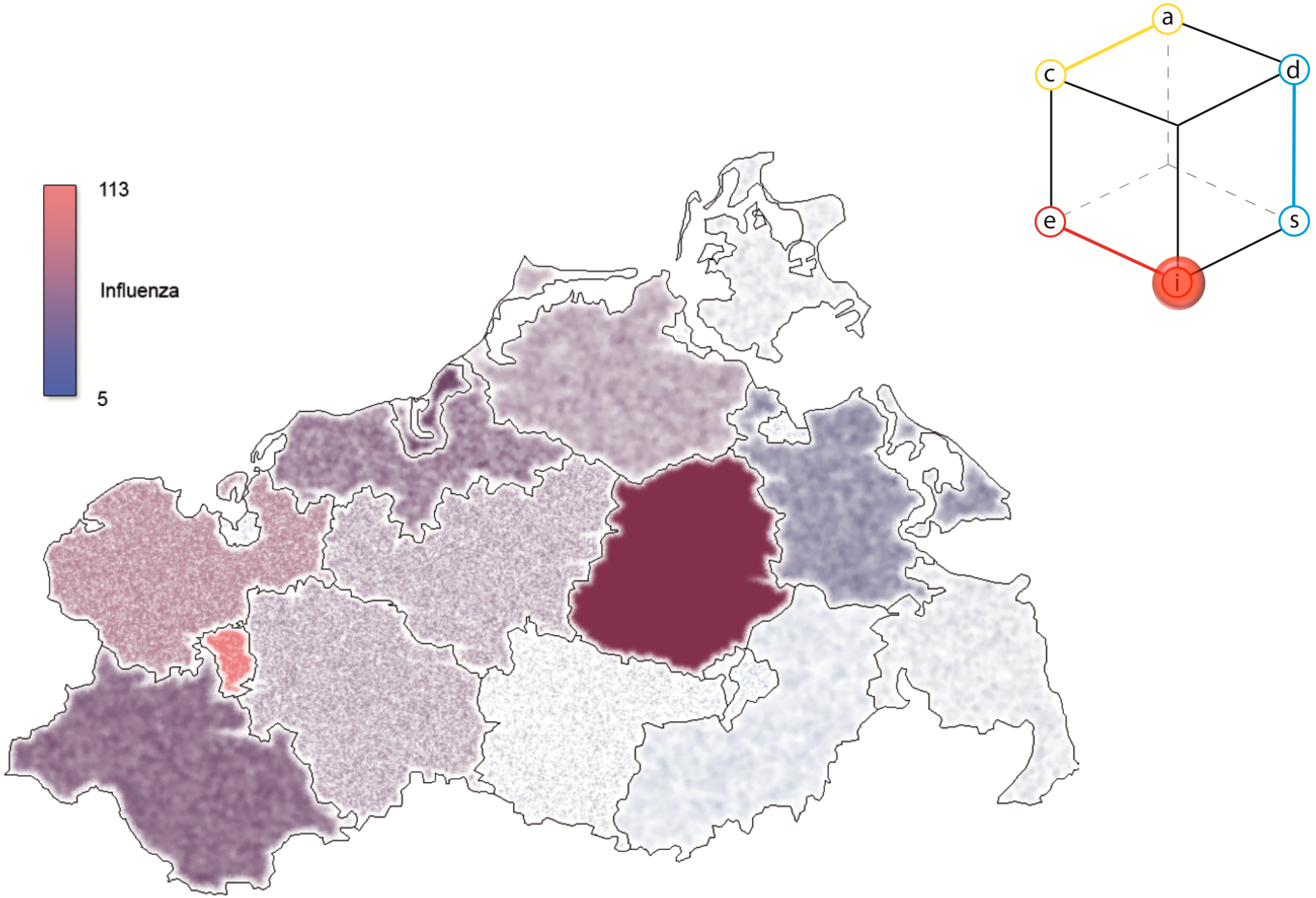



- Static: uncertainty visualization is static
- Dynamic: uncertainty visualization uses animation and/or interaction

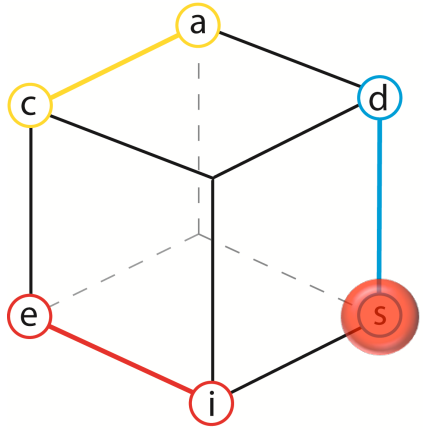
*(Kinkeldey et al. 2014b)*

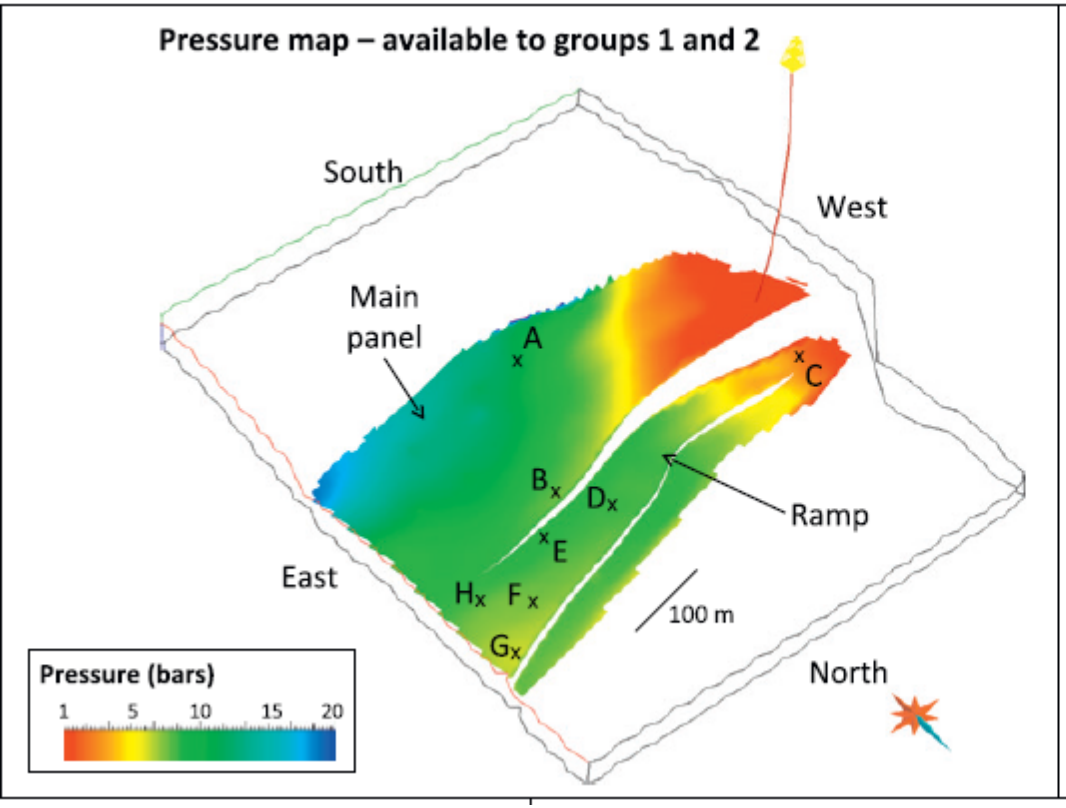


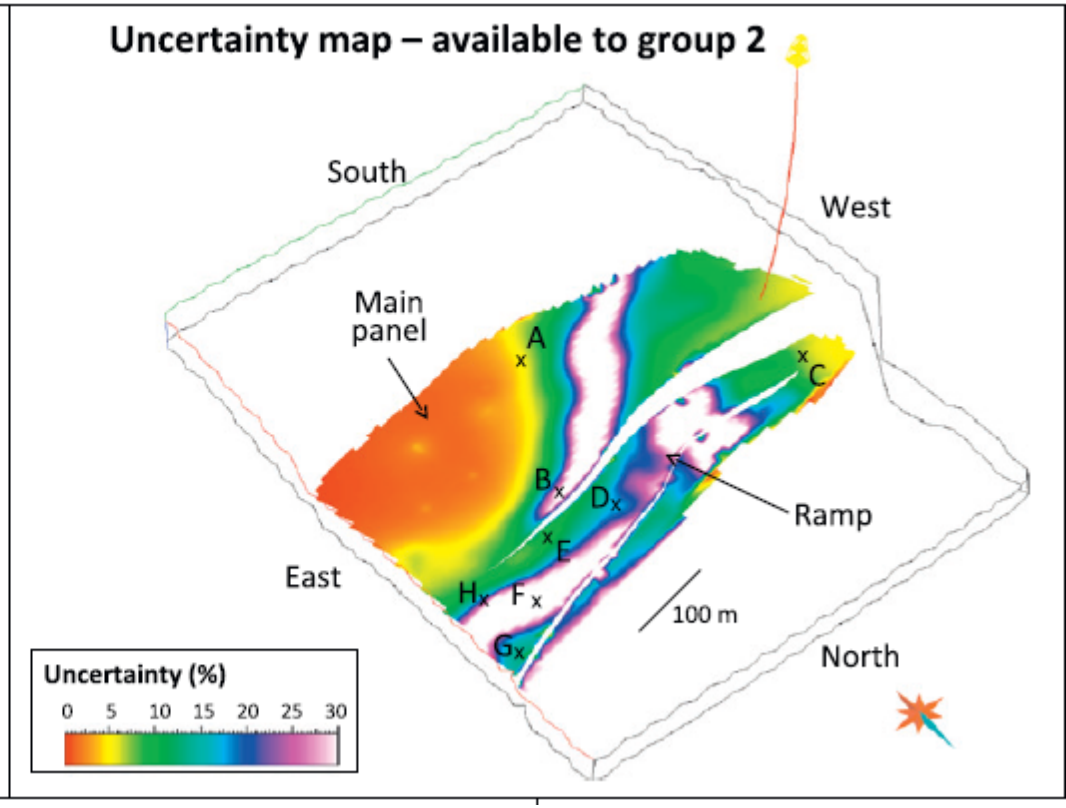

*(MacEachren et al. 1998)*



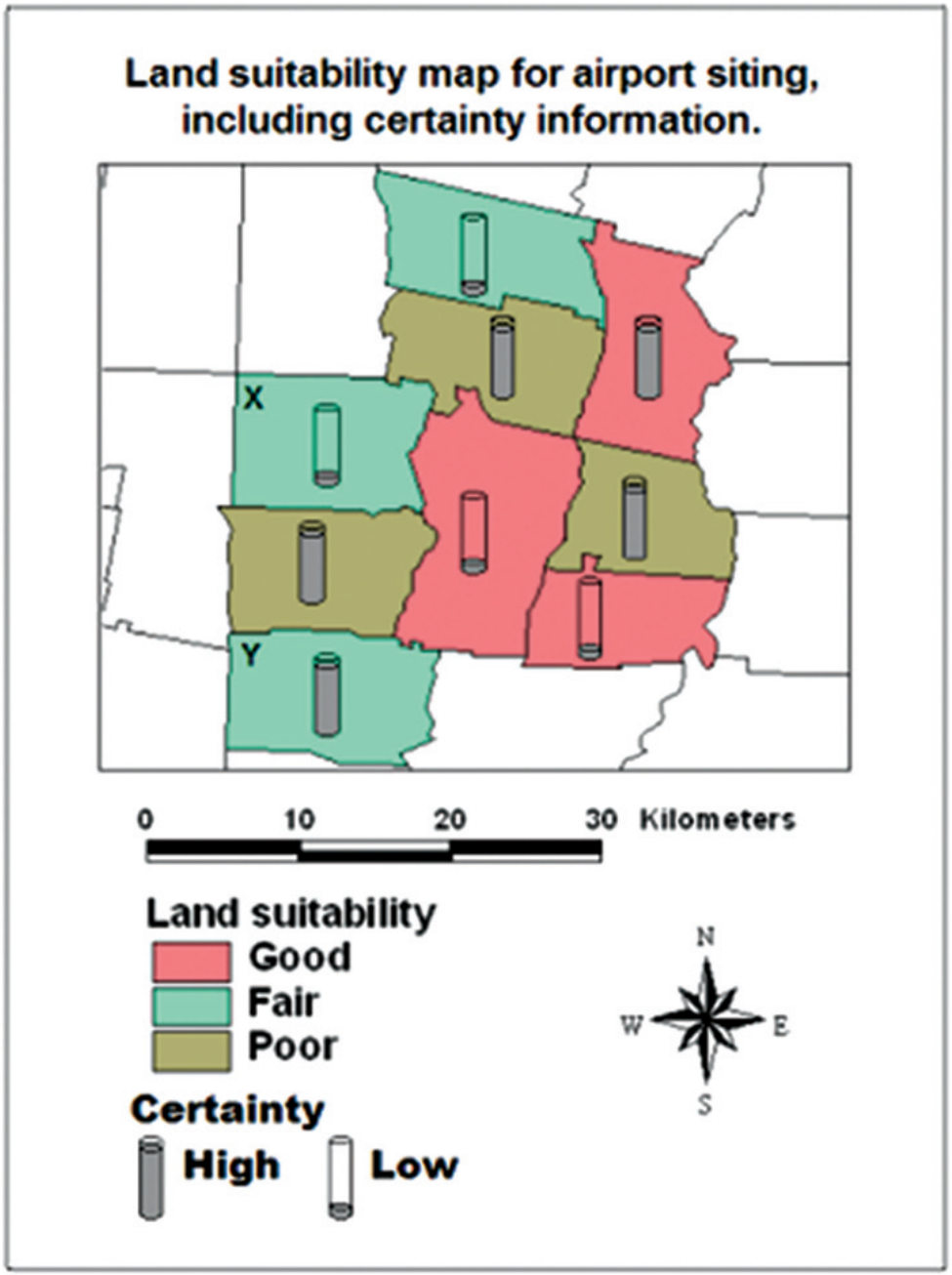

*(Luboschik et al. 2010)*

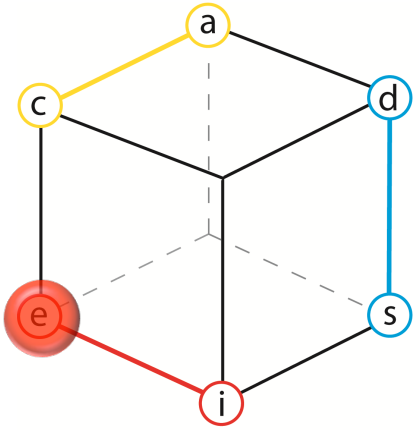


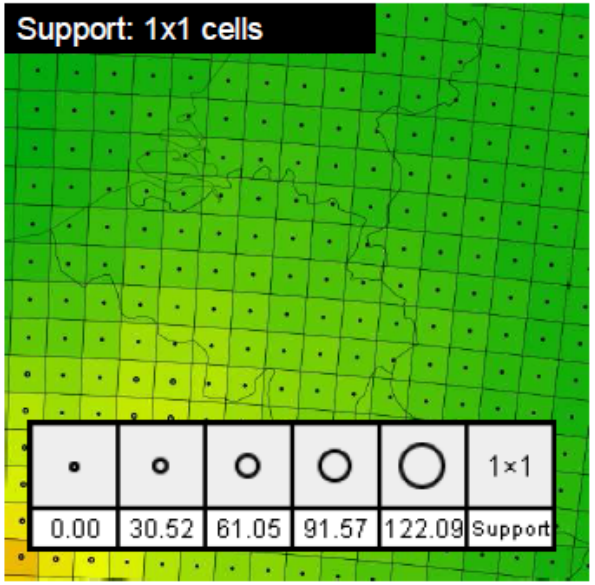



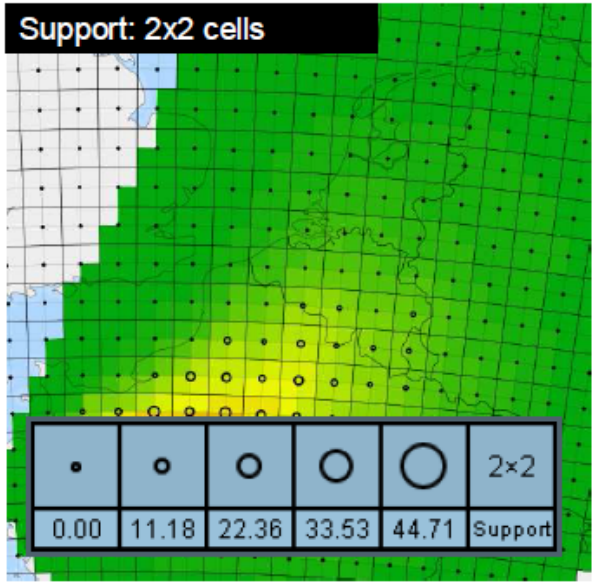

*(Viard et al. 2011)*

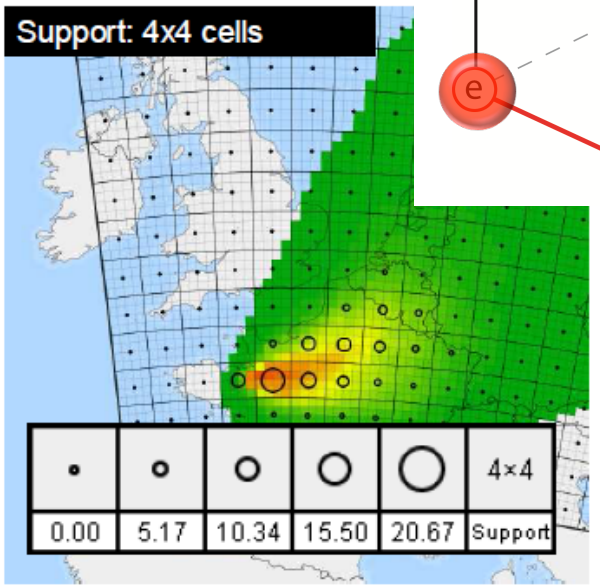


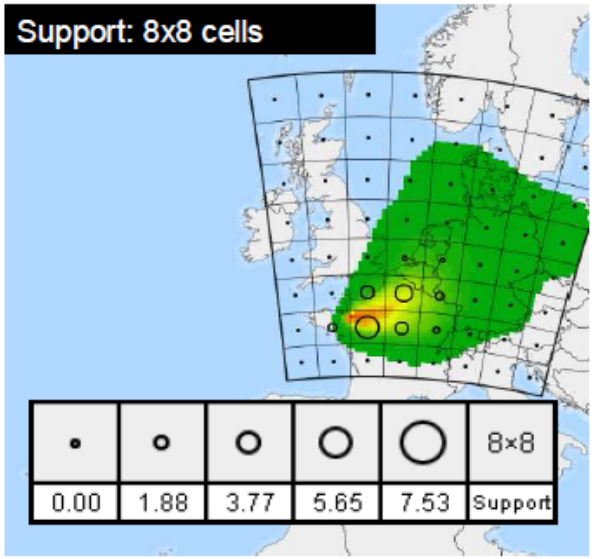



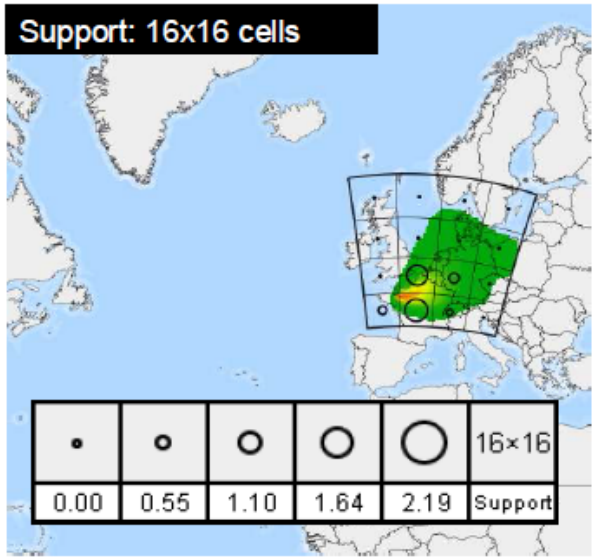

(Hope and Hunter, 2007)

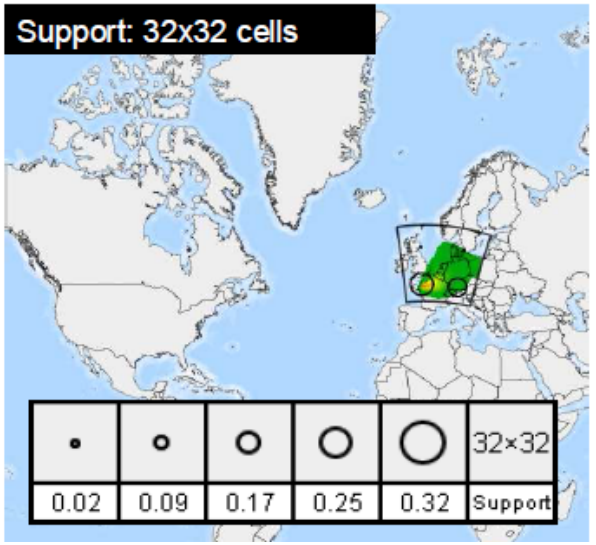


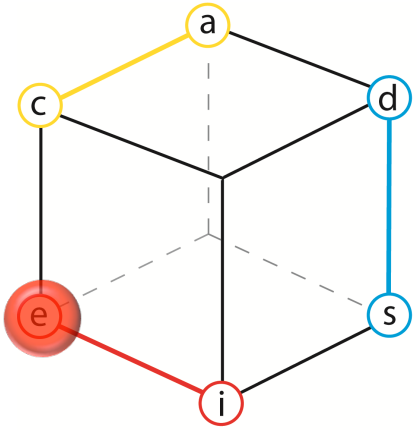




*(Alberti 2013)*

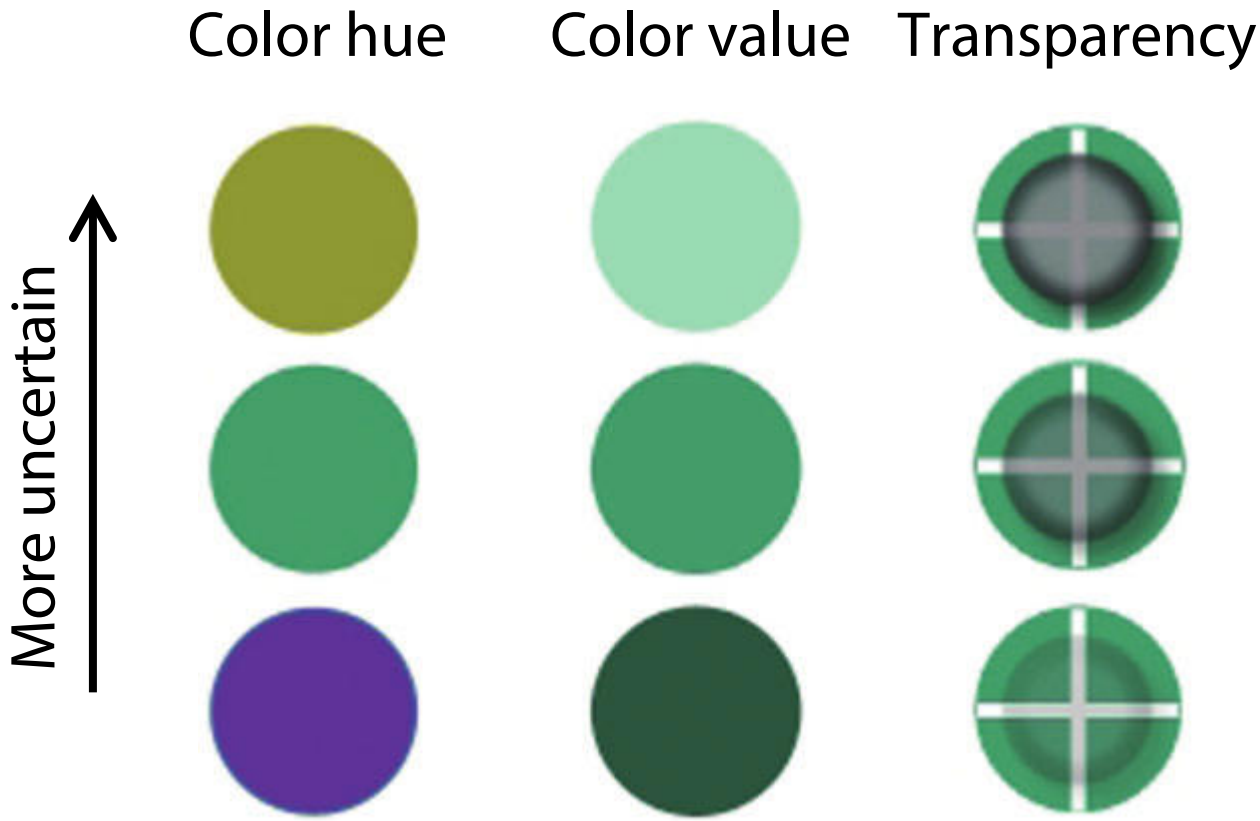
# Typologies

Data types ➔ uncertainty visualization techniques

| Value | Visualization Extent | |
|---|---|---|
| | Discrete | Continuous |
| Scalar | Glyphs (error bars, box plots, Tufte quartile plots) | Pseudo-coloring, difference images, side-by-side, contour lines, blinking |
| Multivariate | Chernoff faces, scatter plots | Side-by-side, difference images |
| Vector | Glyphs (modified tensor probes) | Modified streamlines/ ribbons/tubes, modified line integral convolution (LIC) |
| Tensor | Glyphs (modified tensor probes) | Modified hyperstreamlines |

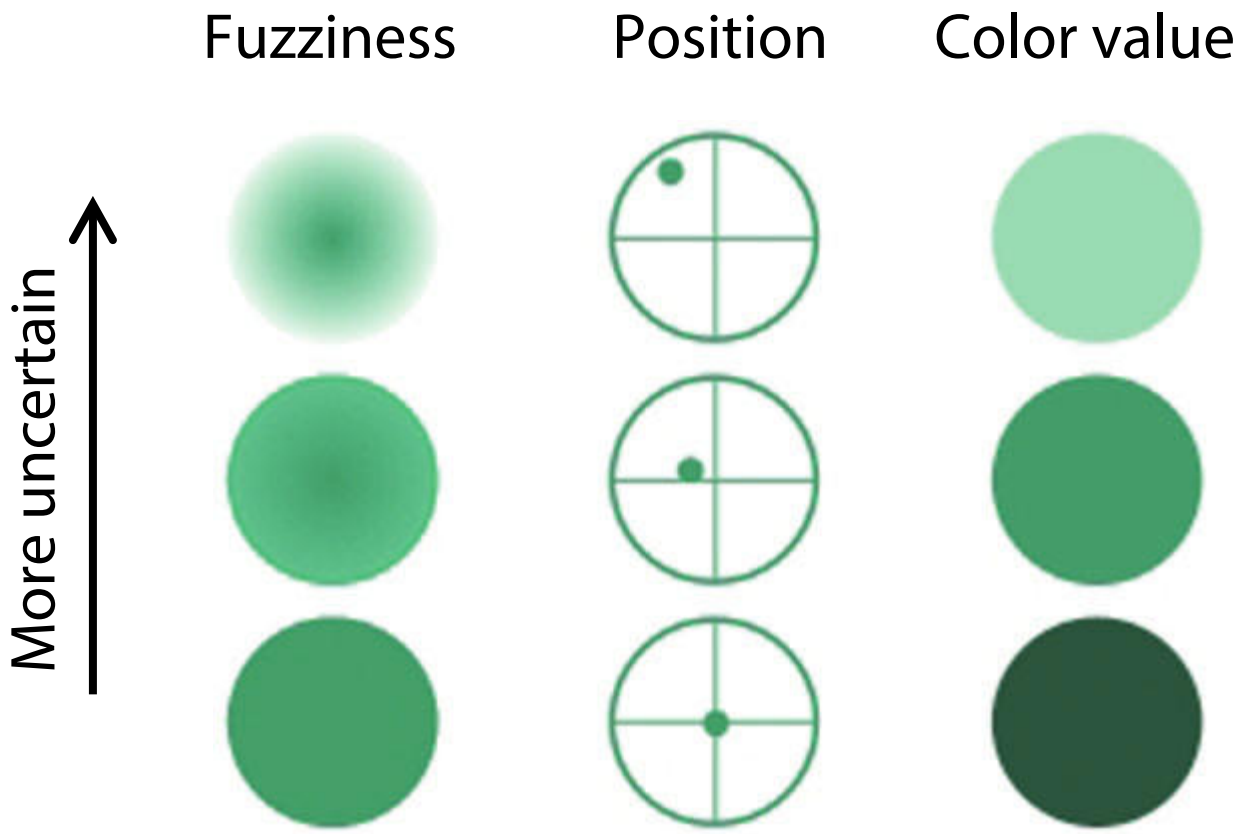
# Some results

Good results w.r.t. user performance (accuracy):



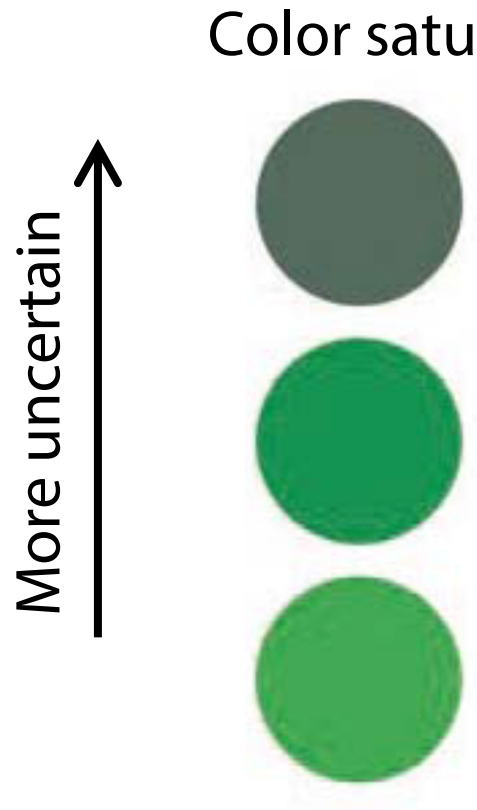
# Some results

Good results w.r.t. intuitiveness:



# Some results

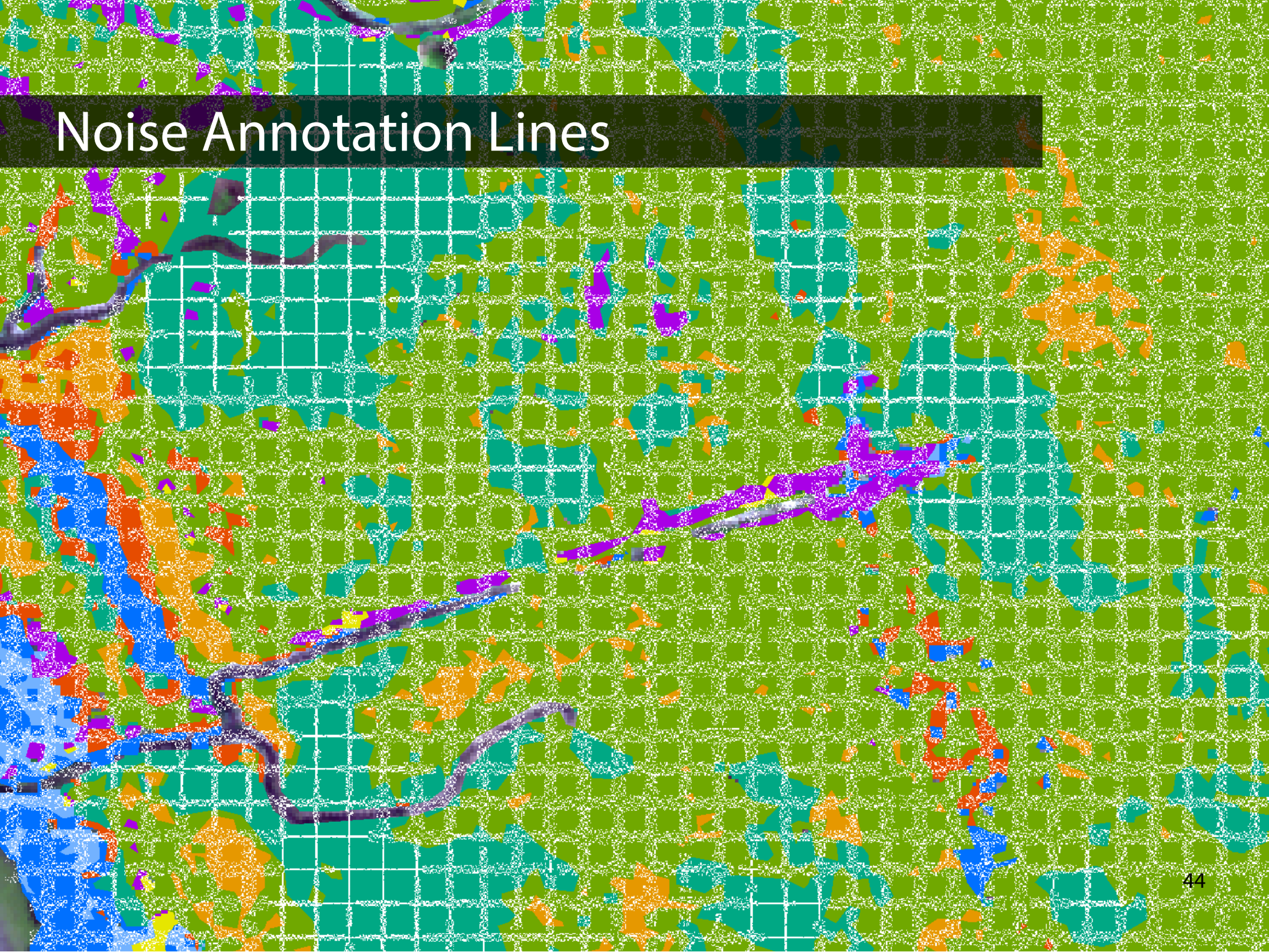
User preference often does not correspond to performance (accuracy)

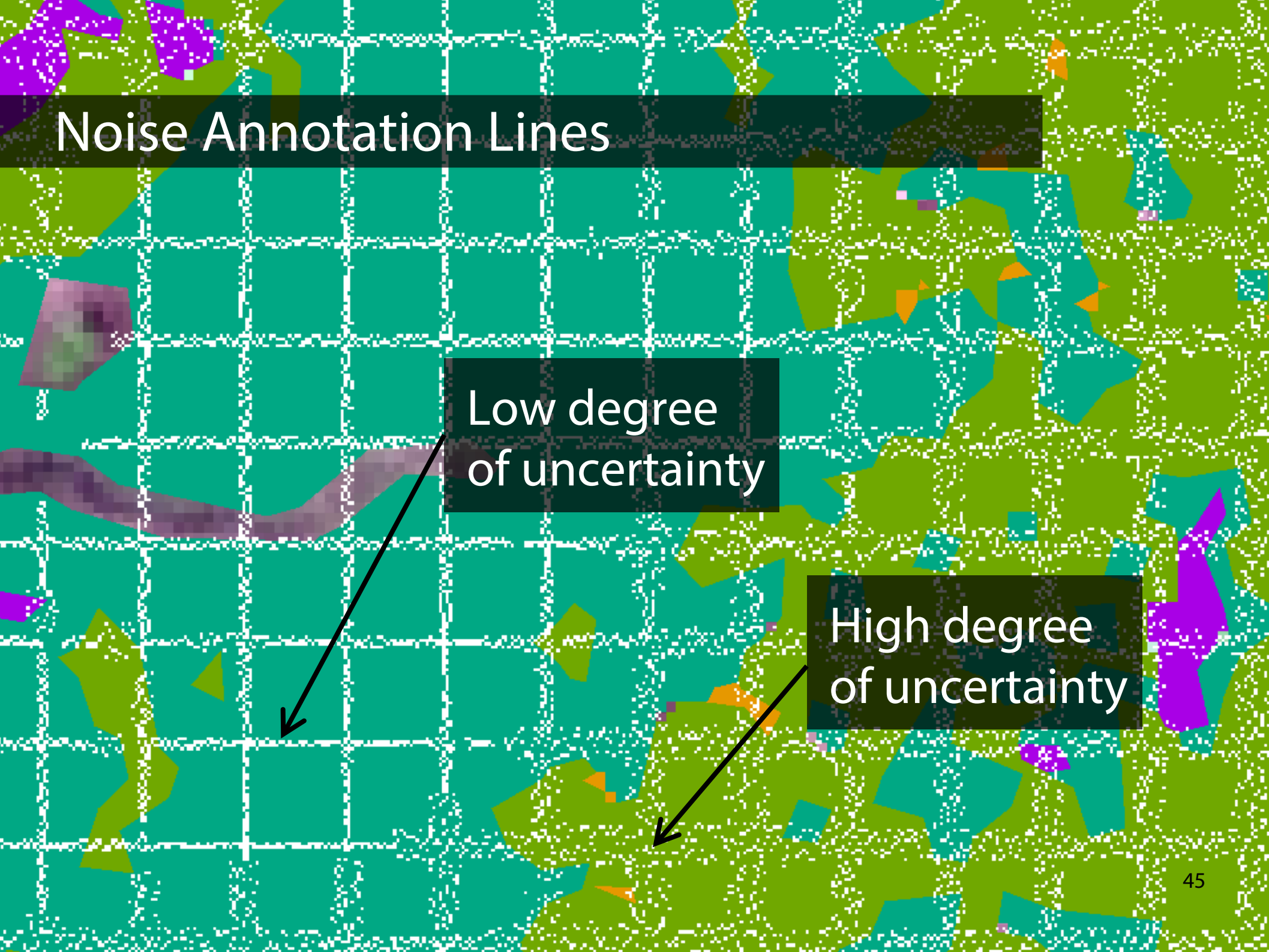


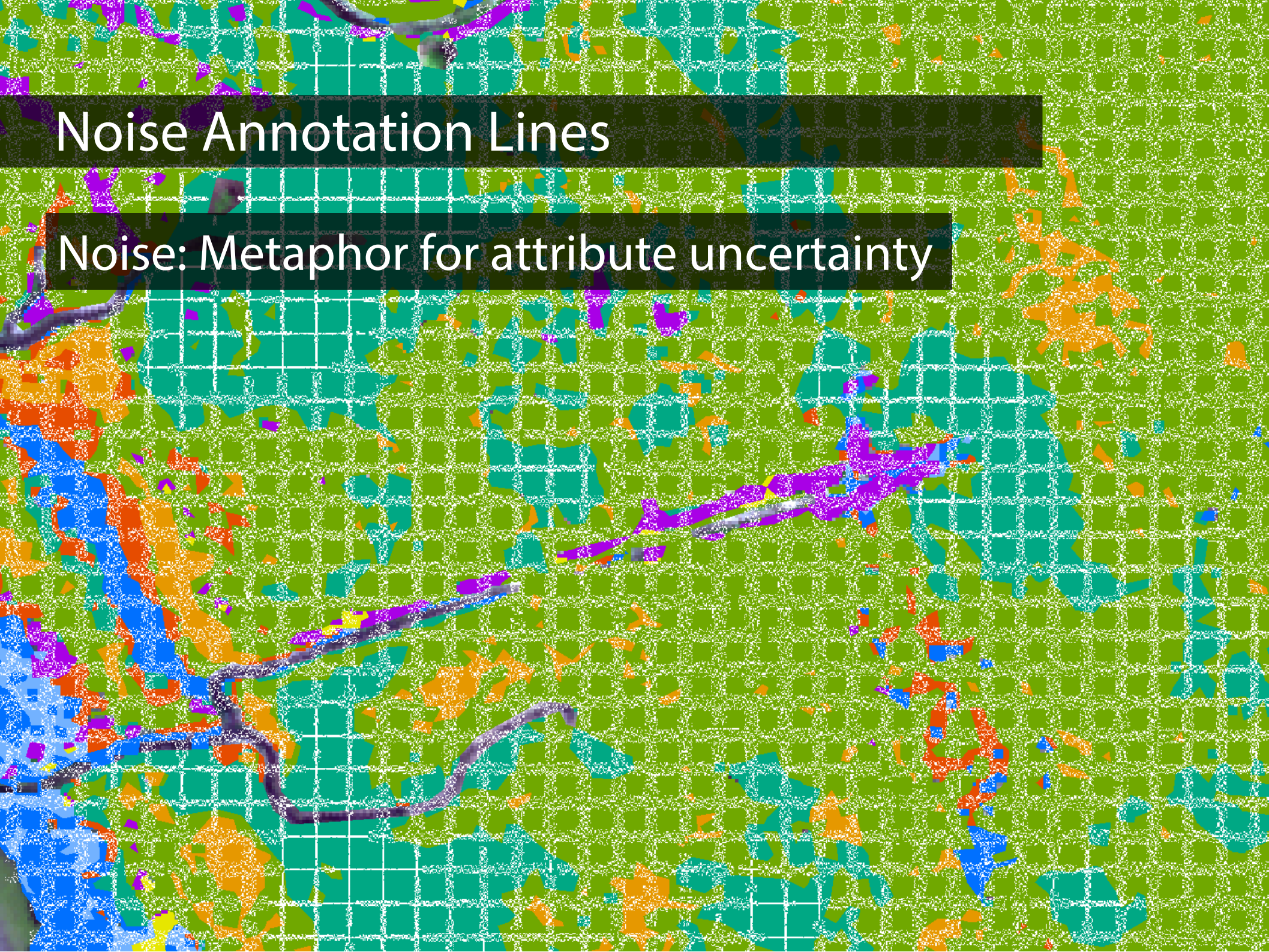

(+) Preference

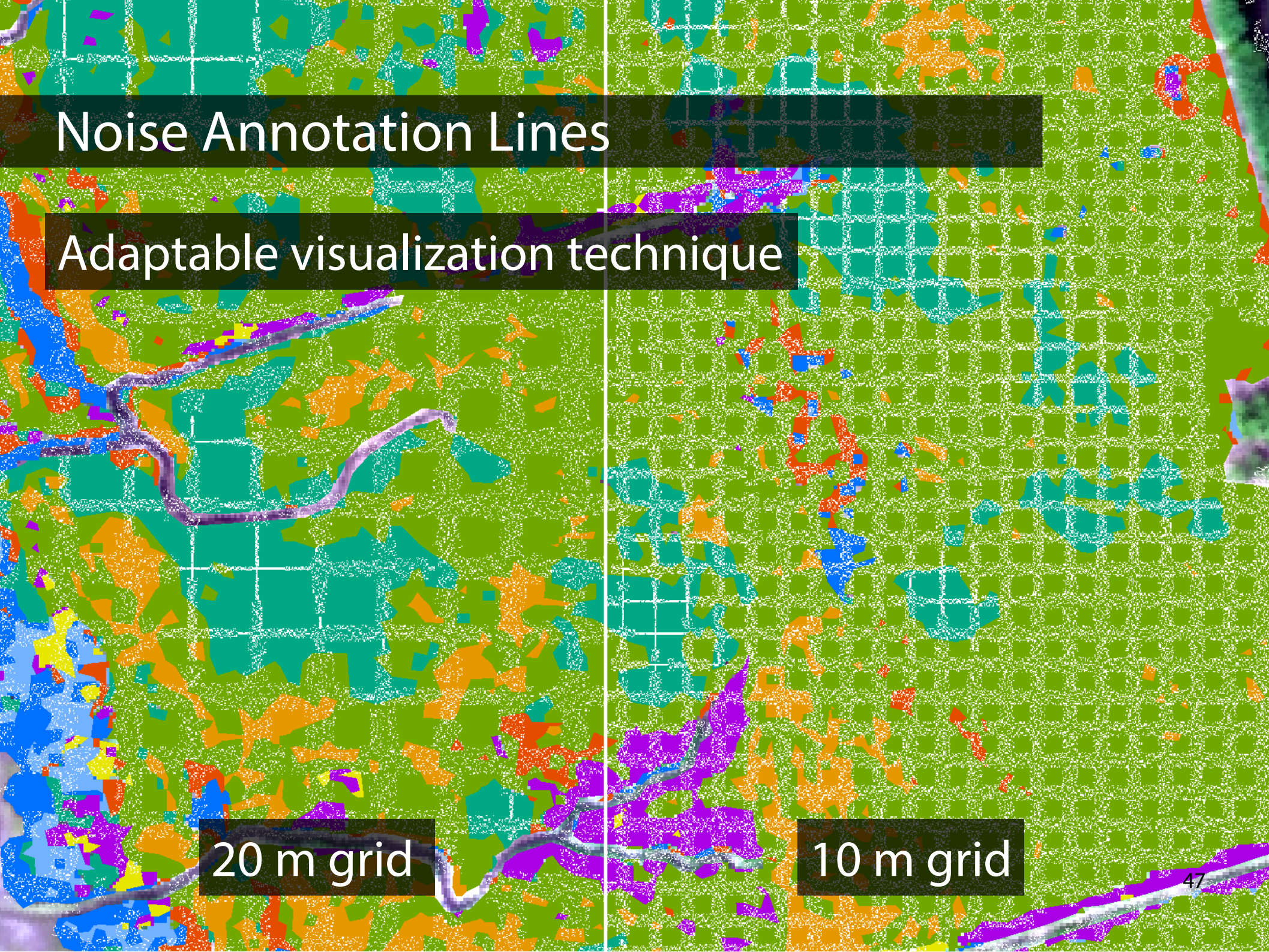
(-) Performance

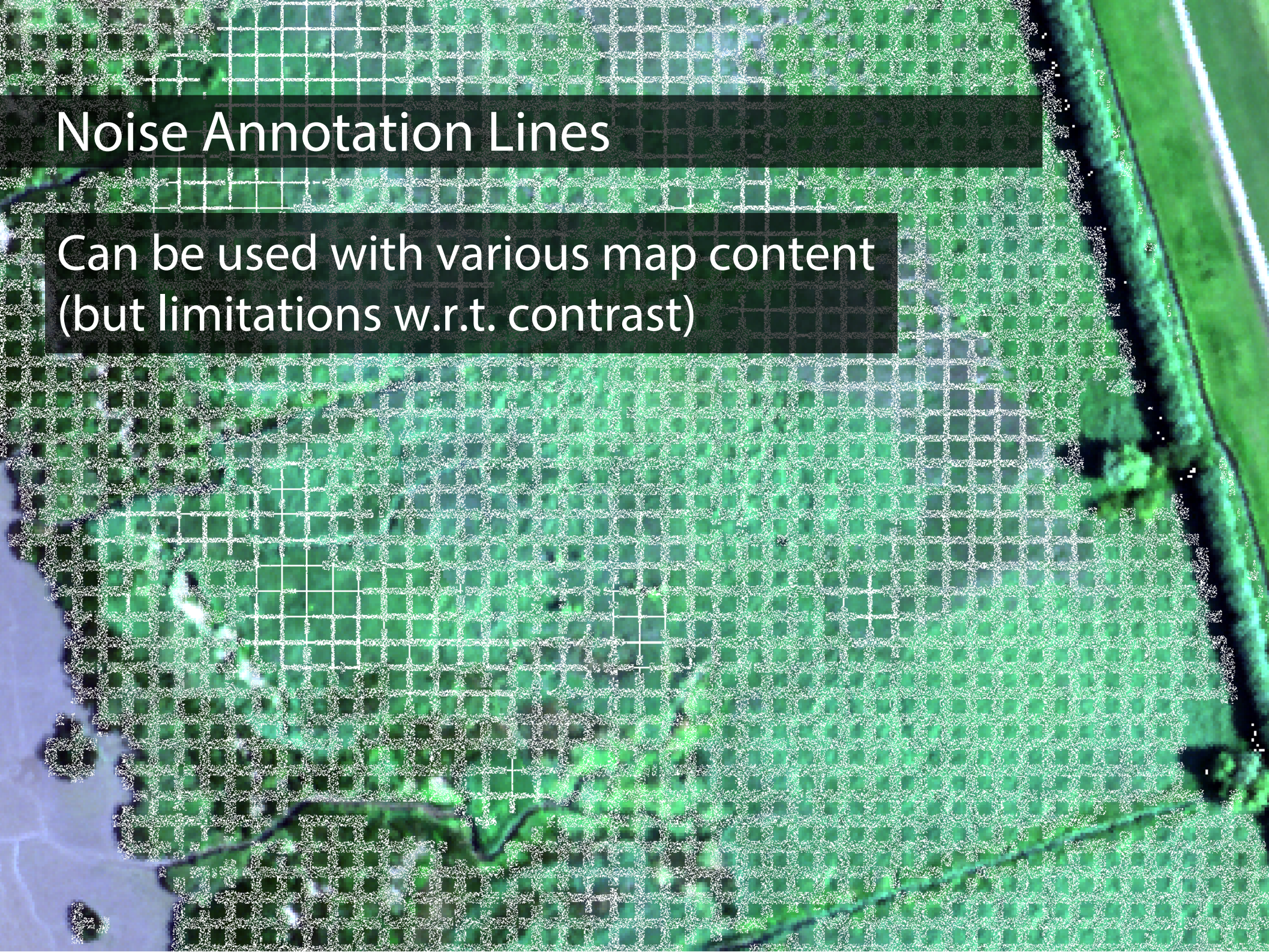
# Noise Annotation Lines

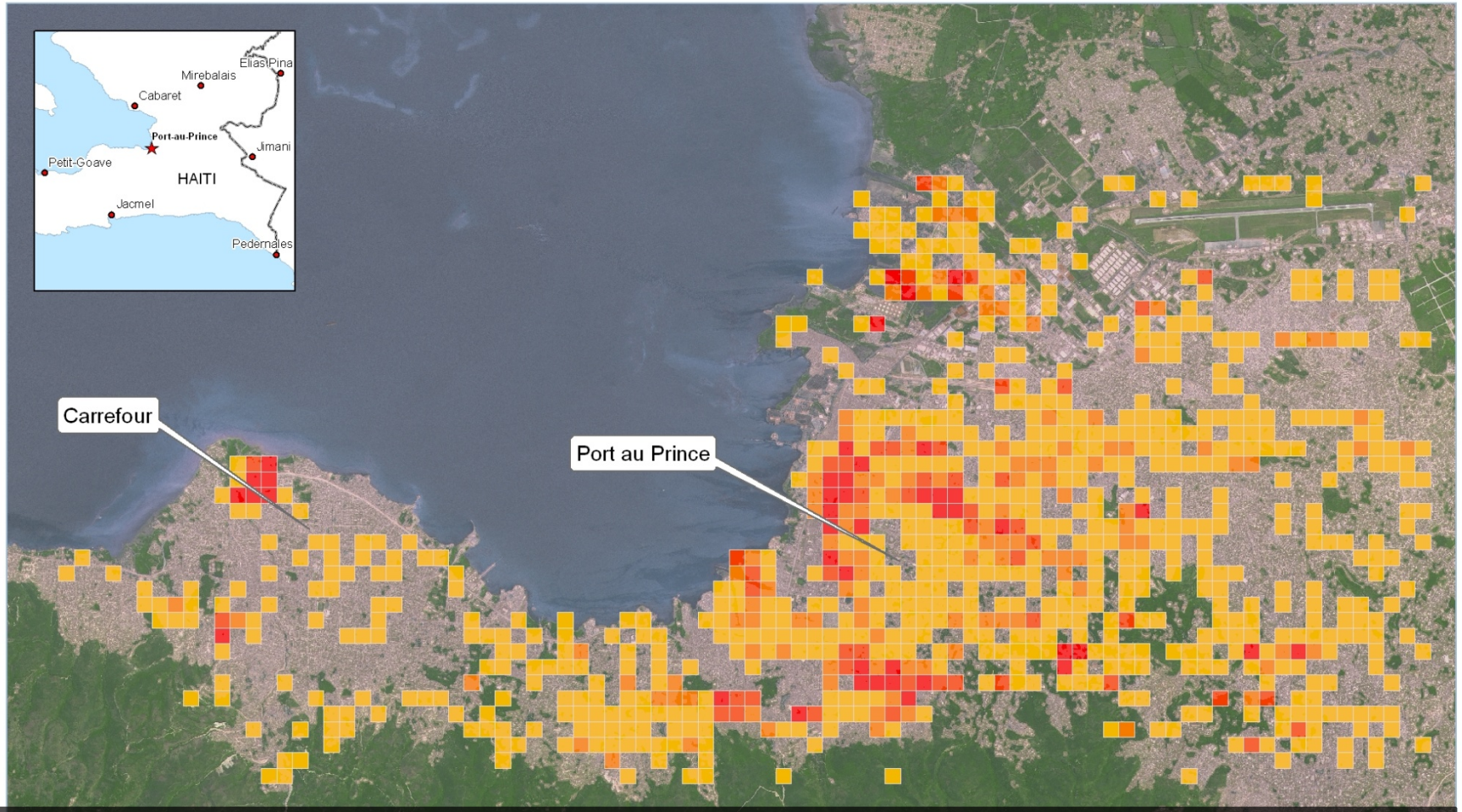



Noise Annotation Lines
Low degree of uncertainty
High degree of uncertainty


Noise Annotation Lines
Noise: Metaphor for attribute uncertainty


Noise Annotation Lines
Adaptable visualization technique
20 m grid
10 m grid


Noise Annotation Lines
Can be used with various map content
(but limitations w.r.t. contrast)




# Land cover change analysis

## *Building damage, Port-au-Prince, Haiti*

## *"before / after" 12 January 2010*

**HAITI Earthquake January 2010 Damage Assessment Map**

## Legend

**Total Number of Damages or Destroyed structures per 200 meters Grid cells**

- Up to 5
- 6 - 10
- 11 - 15
- More than 15
- Built up extent

Coordinate System: WGS 1984 UTM Zone 18 N
Projection: Transverse Mercator
Linear Unit: Meter
Datum: WGS84

N 0 550 1.100 2.200 Meters

Scale 1:55,000 for A3 Printing
1 centimeter = 550 meters

Vector Data: JRC
GIS Analysis: JRC/Isferea Team

Satellite Data: GeoEye1
Resolution: 0.5 meter
Date: 12 January 2010
Copyright:© GeoEye
provided by Google Earth

Satellite Data: Spot5
Resolution: 2.5 meter
Date: 03 July 2007
Copyright:© Spot Image

**Analysis carried out by the European Commission's Joint Research Centre with the support of the Instrument for Stability**

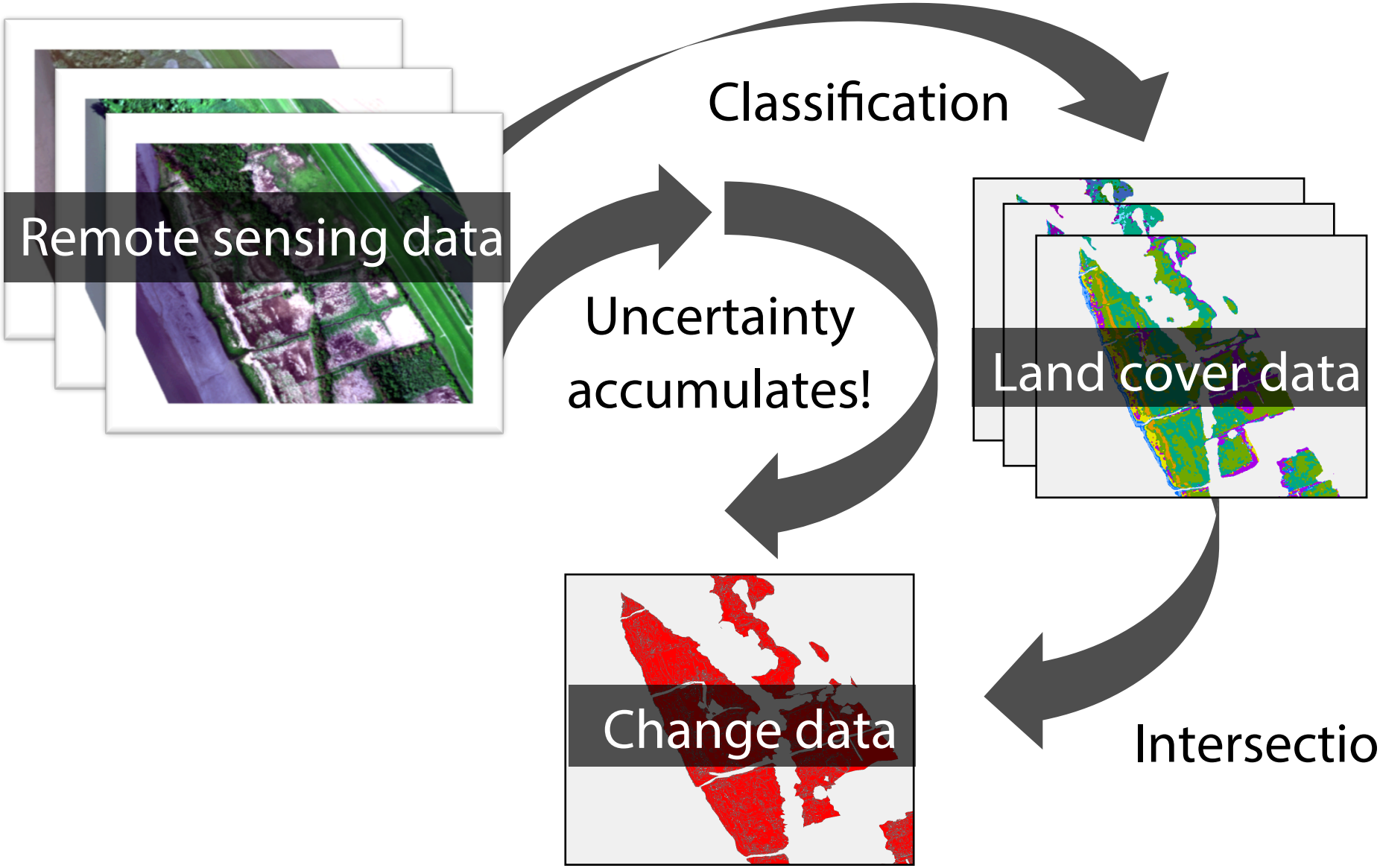


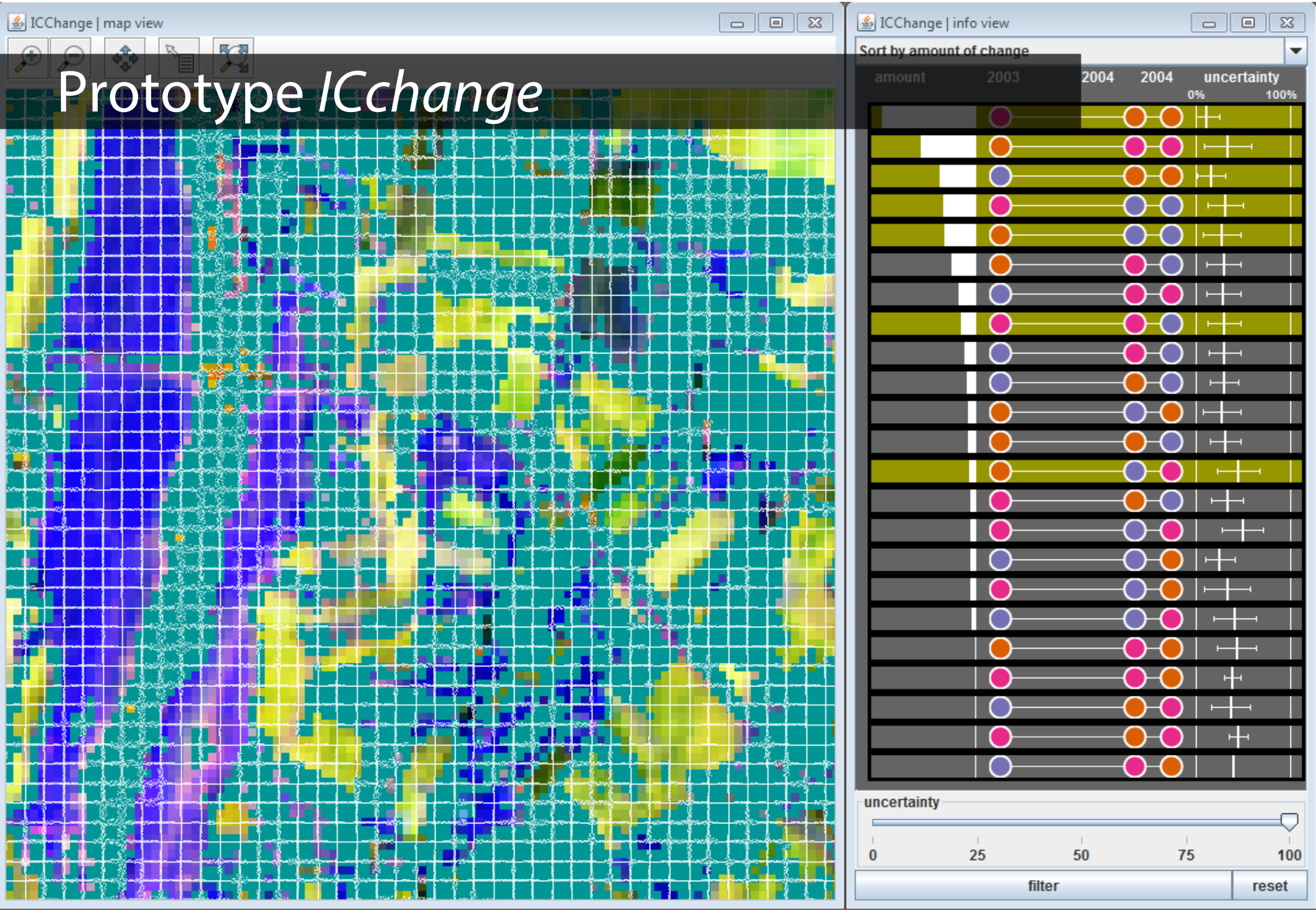
European Commission

Date: 15 January 2010
Global Atlas ID: 1874
Version: 2
Producer: Joint Research Centre

The JRC has carried out a rapid preliminary damage assessment in Port au Prince (Haiti). The JRC has interpreted GeoEye satellite imagery for Port-au-Prince acquired on 13 January 2010. As pre-crisis reference several QUICKBIRD satellite data sets available in Google Earth were used.
The JRC counted damages in Port-au-Prince and flagged those that are either damaged or destroyed. The counting and classification of damages was done as manual labelling following visual interpretation of the pre- and post-event images. Individual damages have been stored as single points with relevant attributes. Statistics have been generated for aggregation by category and attribute.

# Uncertainty in Change Analysis




ICChange | map view
ICChange | info view
Prototype ICchange
Sort by amount of change
amount
2003
2004
2004
uncertainty
0%
100%
uncertainty
0
25
50
75
100
filter
reset

# Expert study // Findings: General

- All groups interested to see uncertainty depicted for their data
- Some participants could confirm intrinsic knowledge about uncertainty with map display
- Possible applications:
  - Assessment of change detection algorithms
  - Optimization of parameters for change detection
  - Identification of erroneous change

# Expert study // Findings: Reasoning

- It was seen as beneficial to have uncertainty information for interpreting change data

“Information that the settlement growth in the north is more uncertain than the reduction in the south would have helped us with the interpretation“

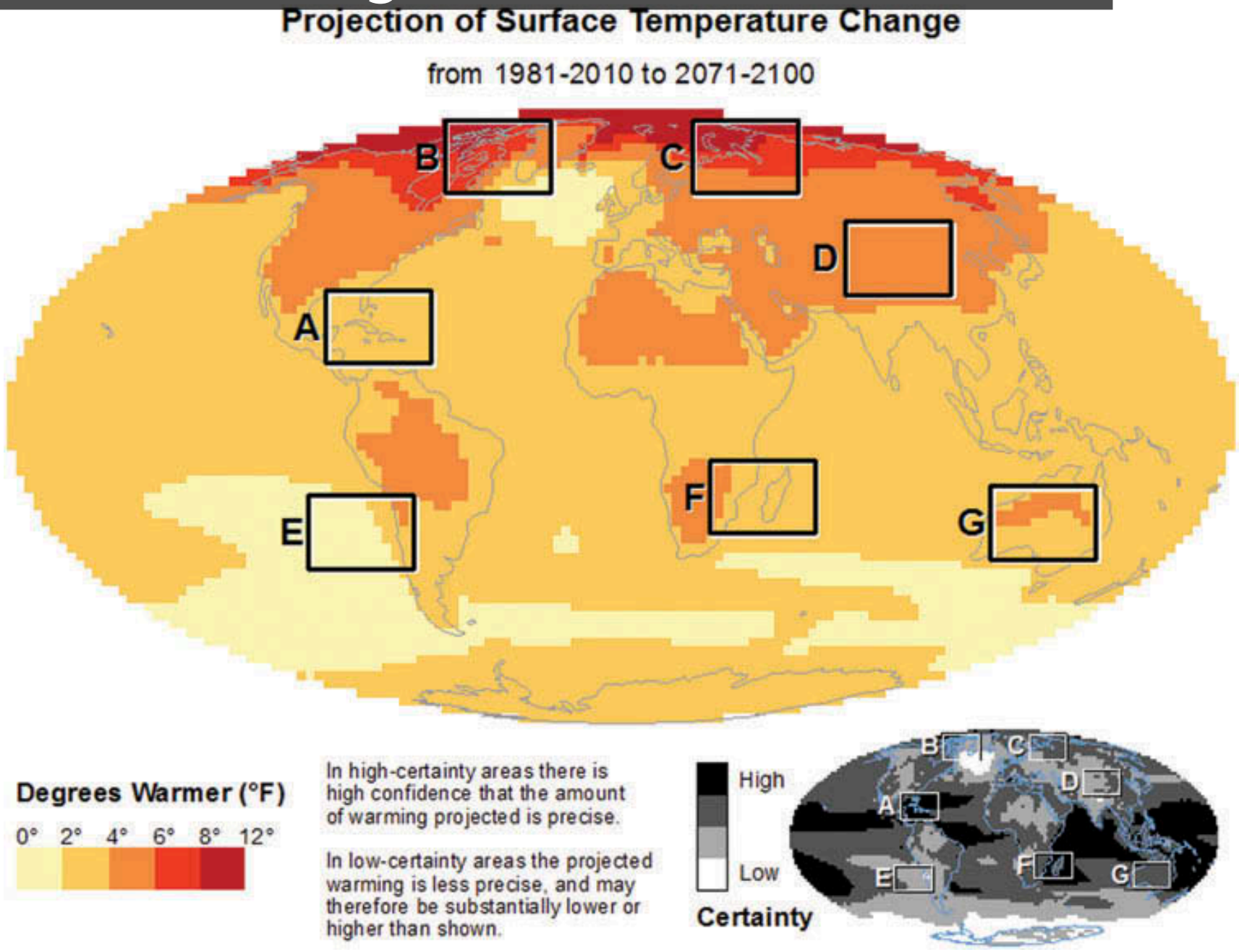
“Estimation about change in population could be better founded with knowledge about uncertainty of the detected change areas”

# Expert Study // Findings: Communication

- Most experts were skeptical about communicating uncertainty to their users
- Could raise doubts about the quality of the data
- “will be hard in the beginning to create acceptance for this”
- Not convinced that decision makers could effectively use uncertainty information

→ depends on the users' role and expertise if uncertainty should be communicated with data

# Decision making



# Decision making

(1) in decision outcomes,
(2) in correctness of decisions,
(3) in kinds of errors made,
(4) in decision time,
(5) in confidence in a decision,
(6) in willingness to make a decision,
(7) in how much workload decision-making causes, or,
(8) in how a decision is made.

# Uncertainty in a nutshell

*»[T]here are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – there are things we do not know we don't know.«*

Donald Rumsfeld, former United States Secretary of Defense, February 12, 2002