LECTURE SEVENTEEN

UNCERTAINTY AND ERROR IN GIS AND GEOGRAPHIC DATA

WHAT IS QUALITY DATA?

Data quality describes the overall suitability of a dataset for a specific purpose

- Error
- Accuracy
- Precision
- Resolution
- Generalization
- Complete
- Compatible
- Consistent
- Applicable

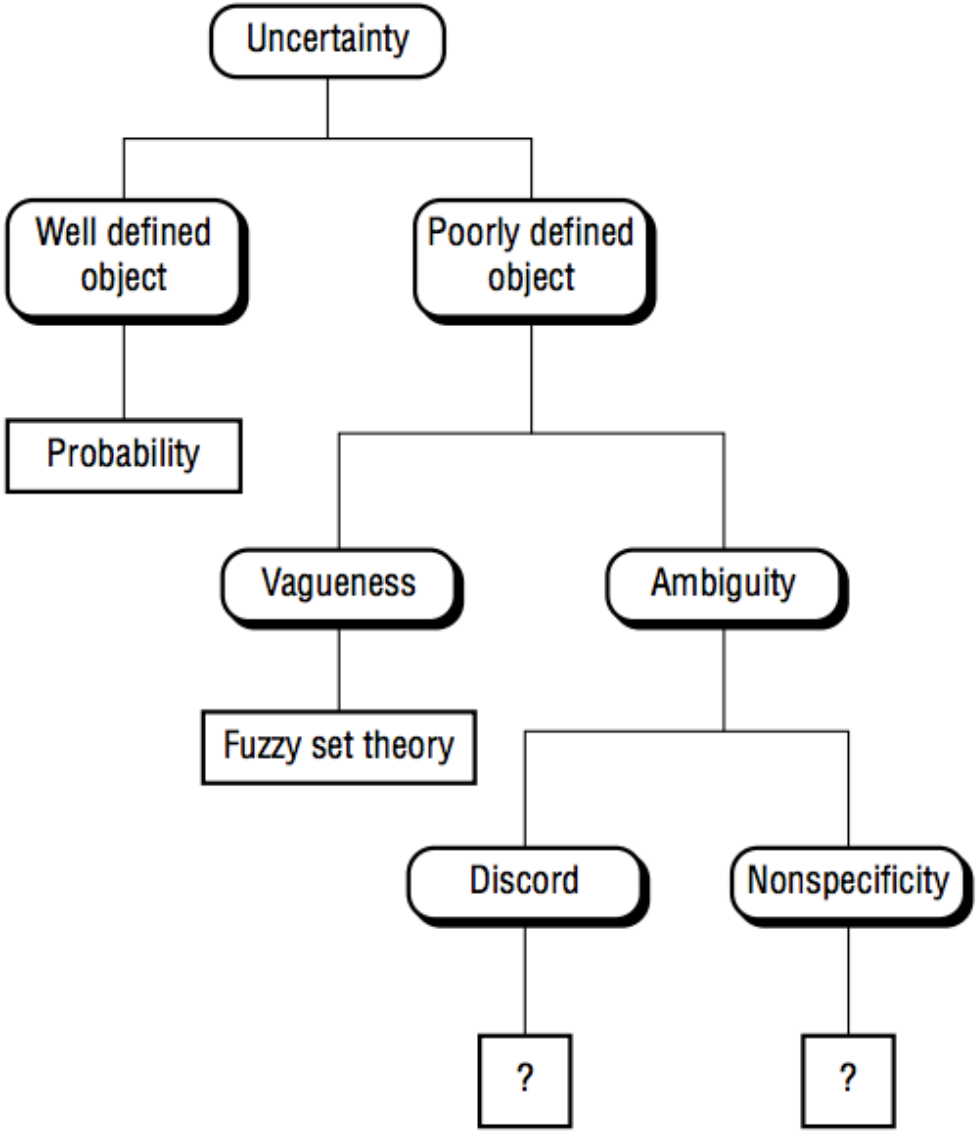


Fig 1. A conceptual model of uncertainty in spatial data (adapted from Klir and Yuan 1995: 268).

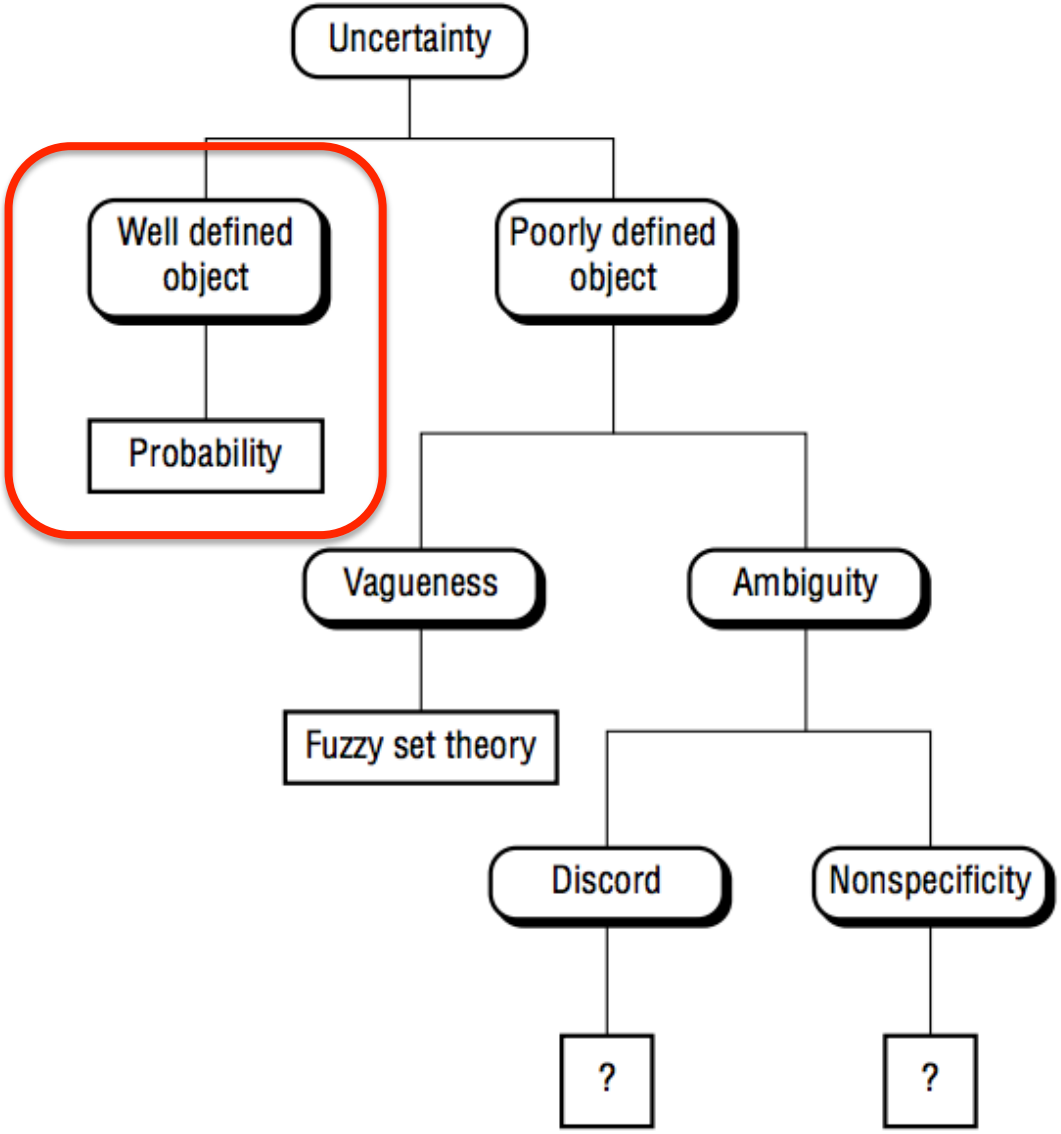


Fig 1. A conceptual model of uncertainty in spatial data (adapted from Klir and Yuan 1995: 268).

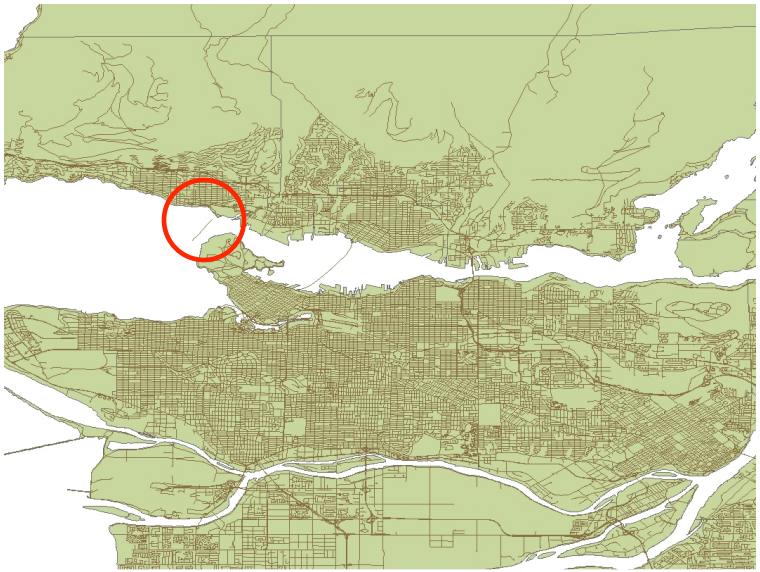
- **Errors** can enter GIS in many ways
- **Errors** in input data can be compounded during analysis
- **Errors** in classification or interpretation may create misleading information

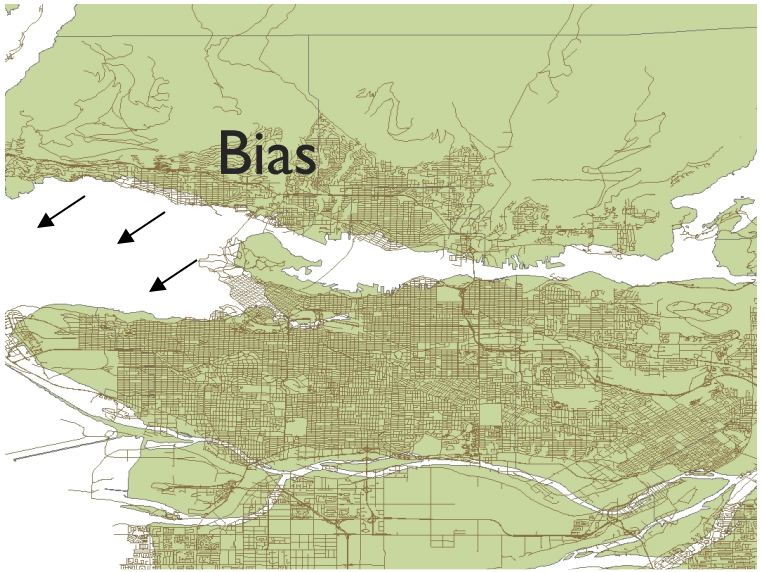
" …the results of analysis are only as good as the data put into the GIS in the first place."

UNCERTAINTY FROM ERROR

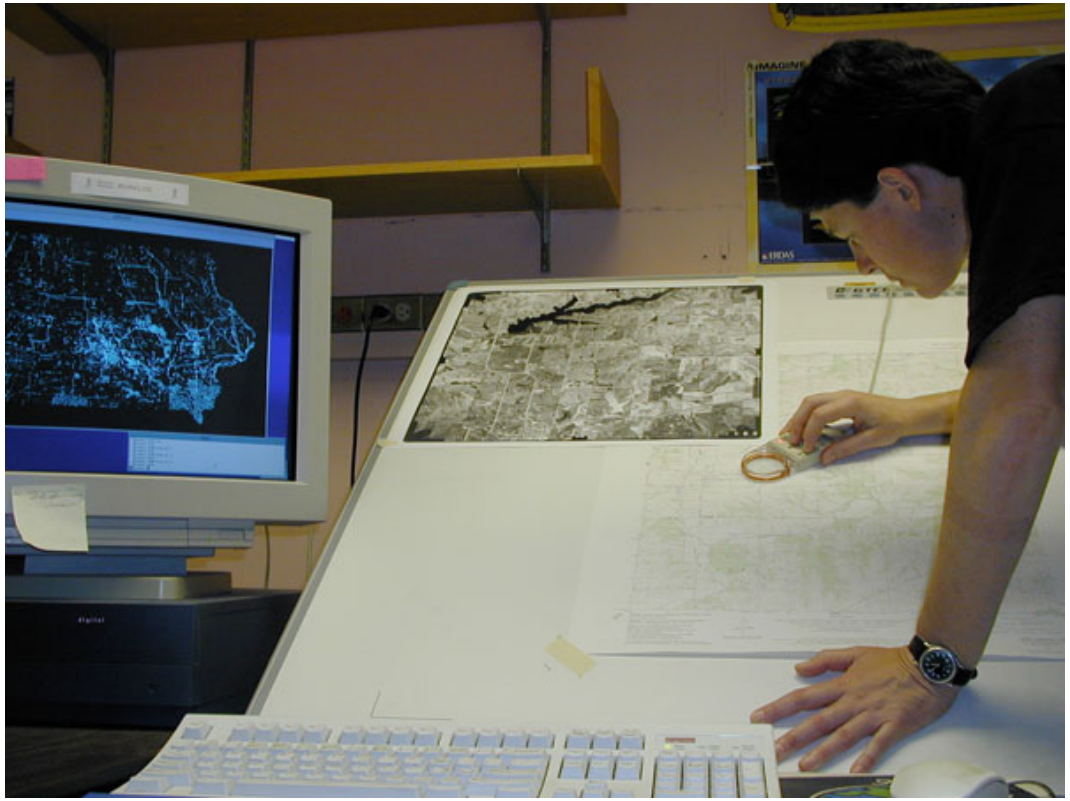
Deviations from the real world as presented in a spatial database

Random Error vs Systematic Error





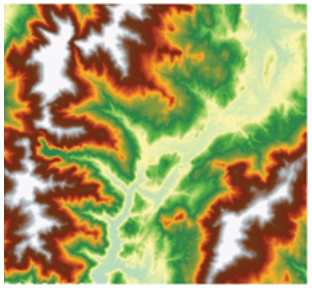
SOURCES OF ERROR

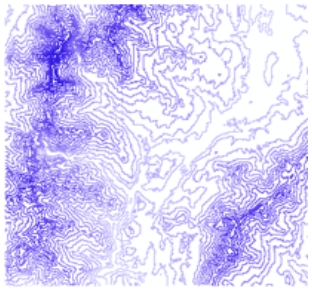


SOURCES OF ERROR

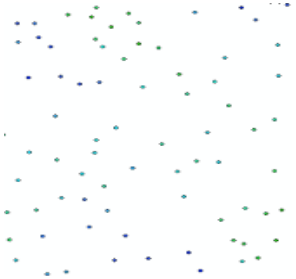
Data transfer and conversion

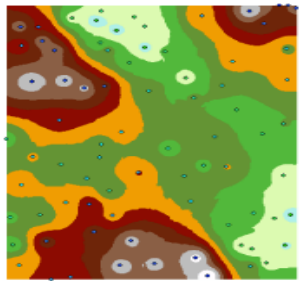
Vectorization





Rasterization





Transferring between software

ESRI

MapInfo.

CLARK LABS
IDRISI Andes
GIS and Image Processing Software
Meeting the Challenges of Environmental Decision Making with GIS

GeoDa

ERROR FROM RASTERIZATION

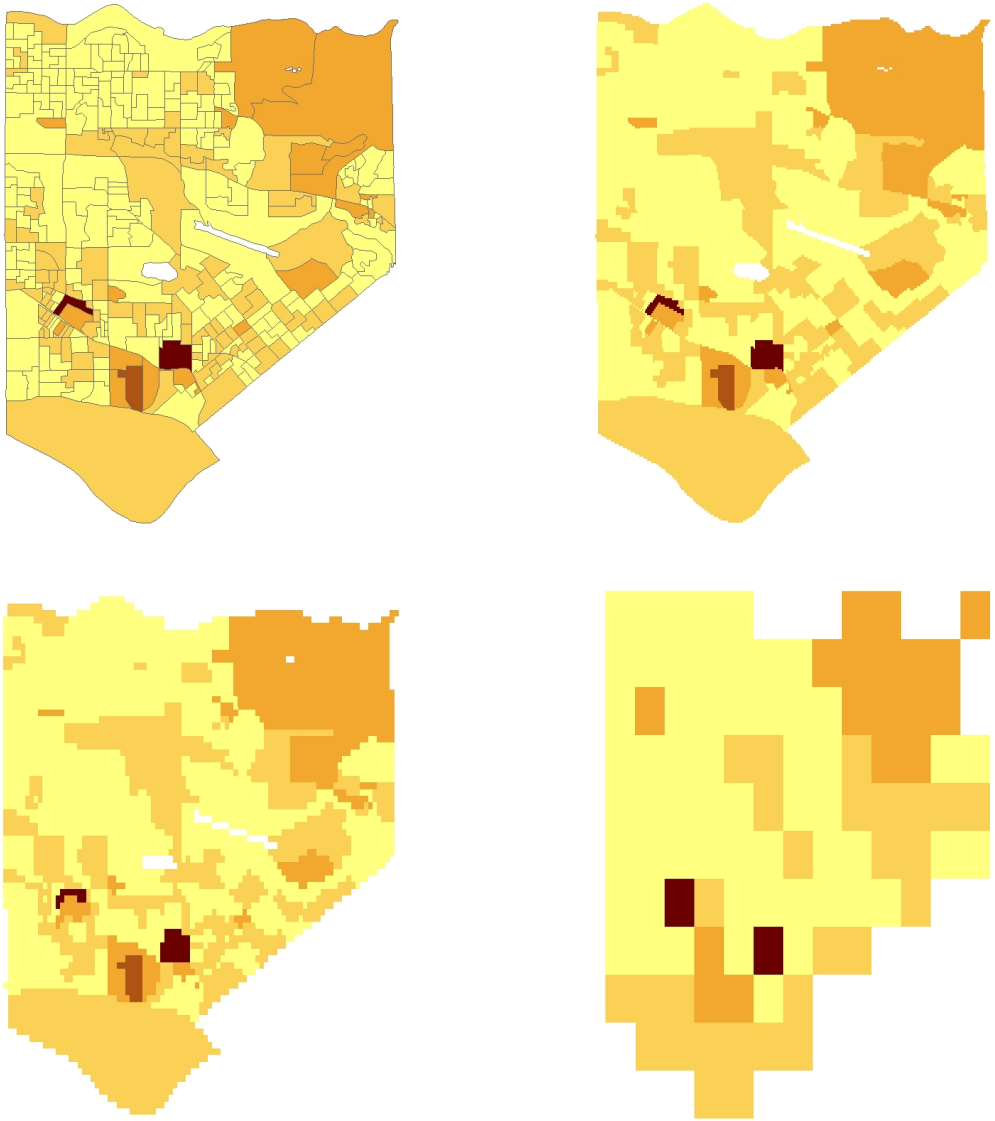
Loss of Polygons:

When polygons are less than half the area of the chosen raster cell size

Topological Errors:

Loss of connectivity or creation of false connectivity where polygons are smaller than the chosen raster cell size

ERROR FROM RASTERIZATION



ACCURACY AND PRECISION

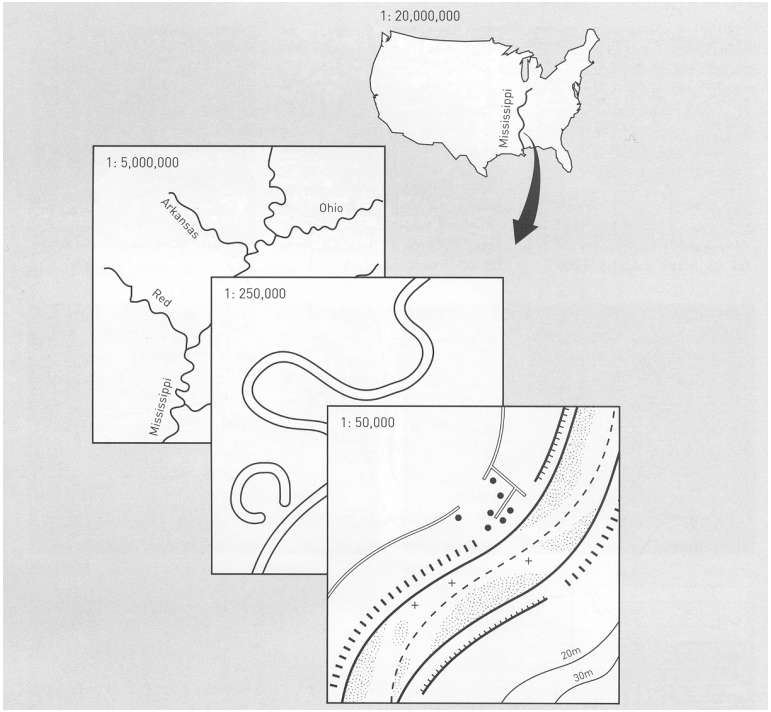
Accuracy: the extent to which an estimated value approaches its true value

Precision: the recorded level of detail in the dataset

GENERALIZATION

The simplification of real-world features

Features need to be generalized because data are required a certain scale



Heywood et al., 2006

TYPES OF GENERALIZATION

Selection: selecting the real-world features to be included in the map

Simplification: features are altered in order to enhance visibility and reduce complexity

Displacement: Altering the location of overlapping or adjacent features in order to make them both visible

Smoothing and Enhancement: altering the jagged or rough edges of features in order to provide a better representation

ERRORS FROM GENERALIZATION

Positional errors:

Changing the location of features in order that they are visible on a map or in a dataset

Representation errors:

(1) The selection of certain features and negating others can produce an erroneous representation of reality

(2) Changing a feature to simplify or enhance its visual appearance

OTHER CHARACTERISTICS OF QUALITY

Completeness: data must be spatially and temporally complete

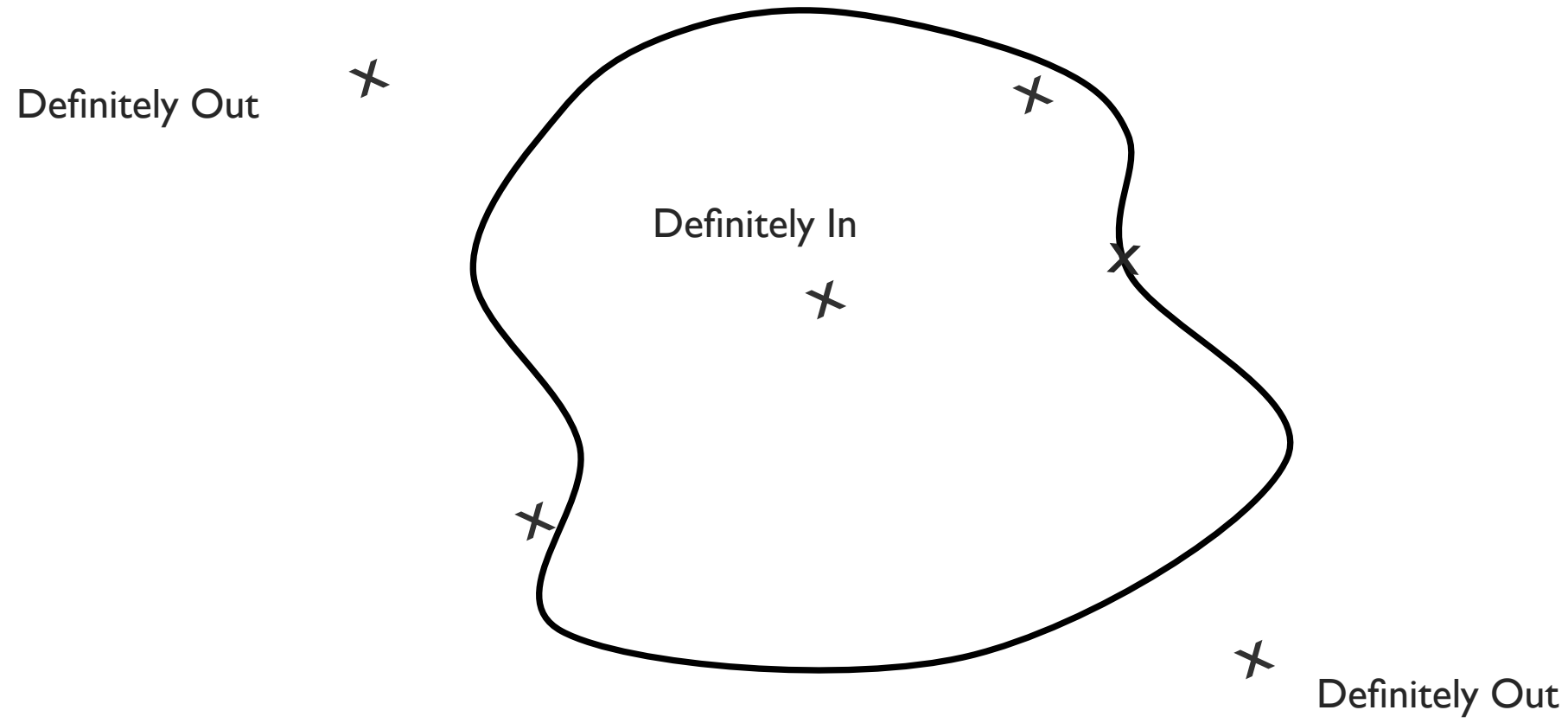
Compatible: Multiple datasets used in the same project must be of the same format, scale and extent

Consistent: Multiple datasets should undergo consistent methods of data capture, storage, manipulation and editing

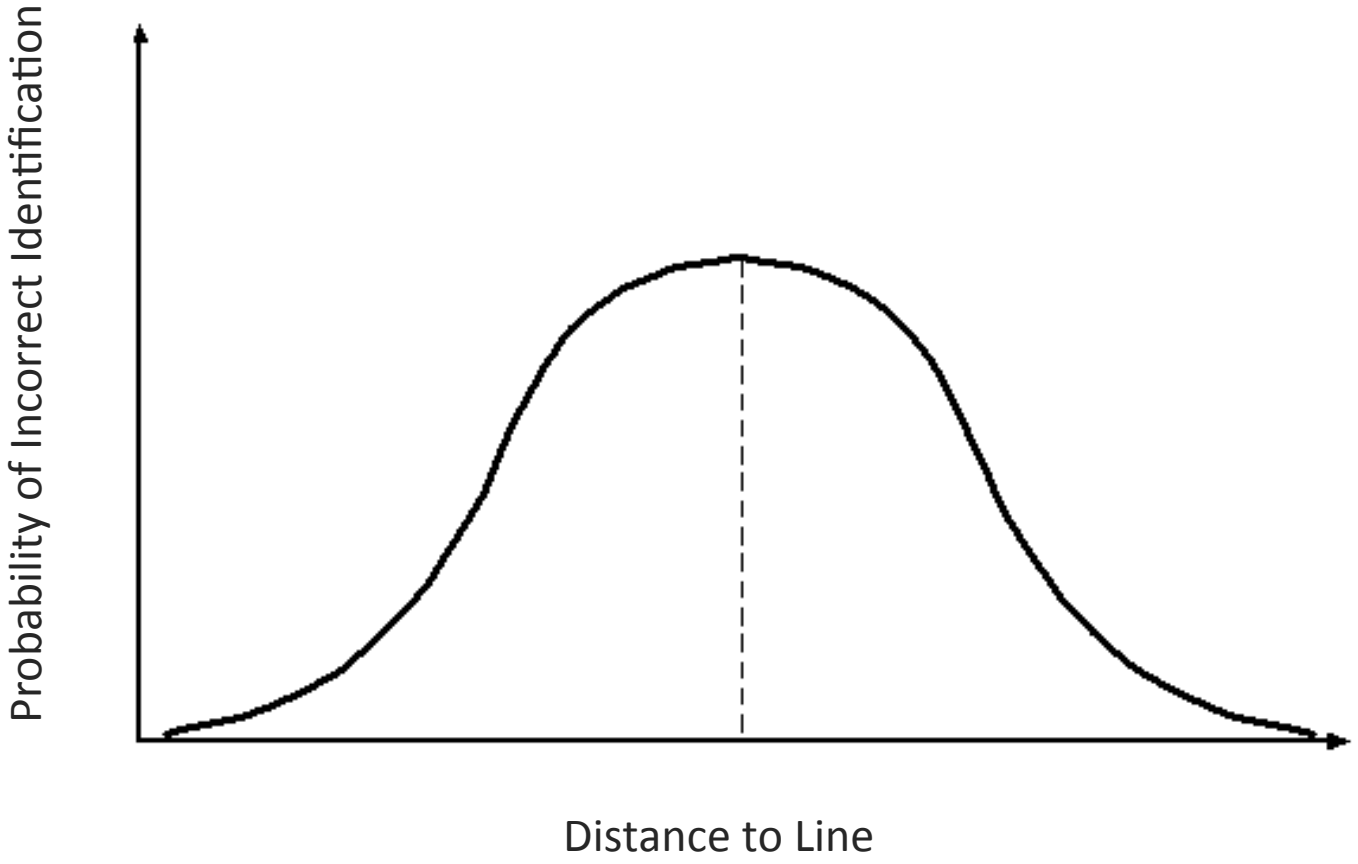
Applicable: data must be suitable for the analysis or project

ERROR IN LOCATION

Uncertainty in object location can hinder GIS operations such as point-in-polygon analysis

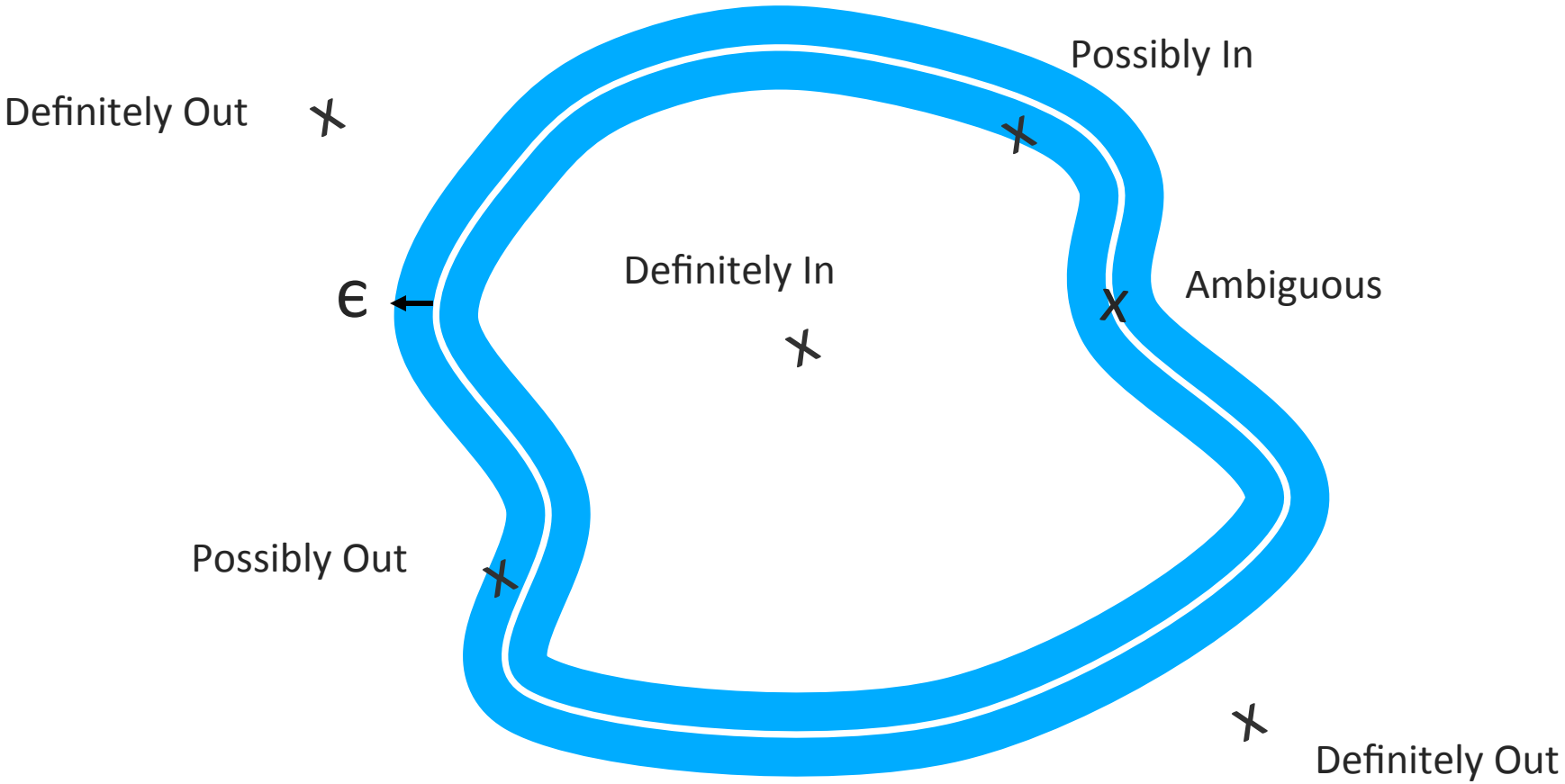


EPSILON BANDS



EPSILON BANDS

Epsilon bands represent widths (ϵ) of uncertainty of whether points fall within or outside a polygon



EPSILON BANDS

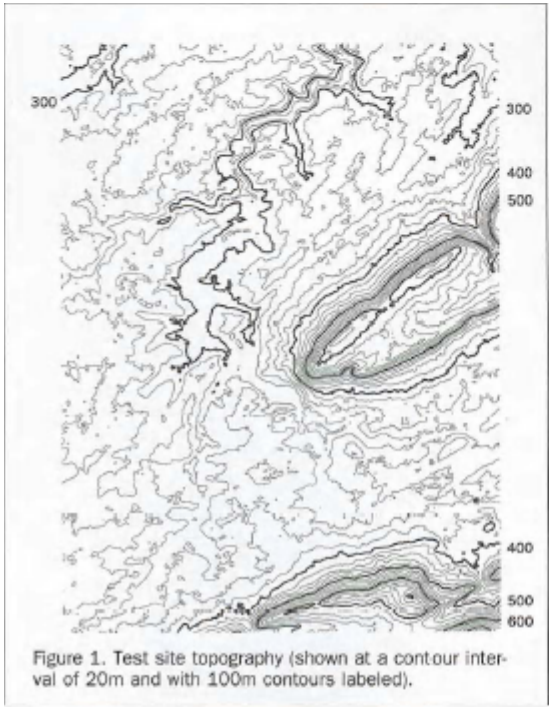
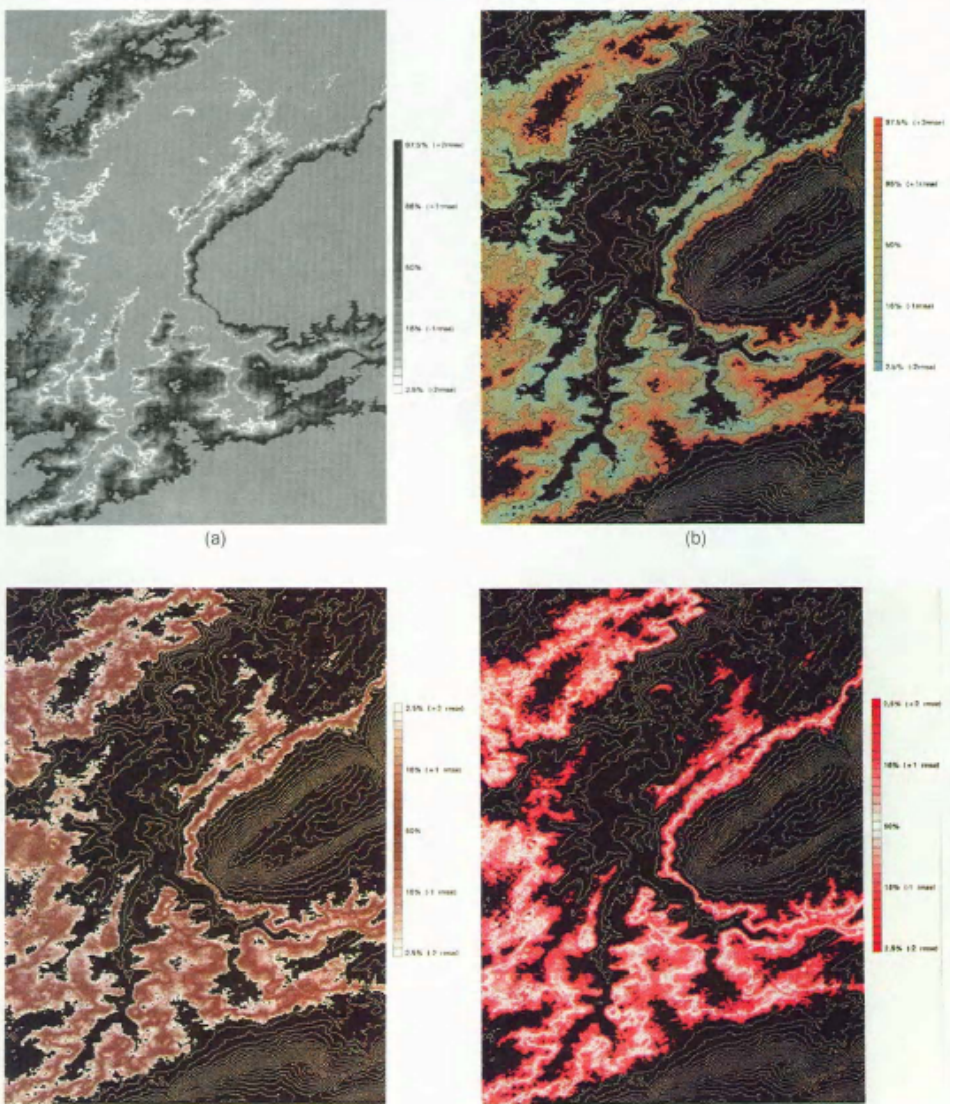


Figure 1. Test site topography (shown at a contour interval of 20m and with 100m contours labeled).



Hunter and Goodchild 1995

UNCERTAINTY FROM VAGUENESS AND AMBIGUITY

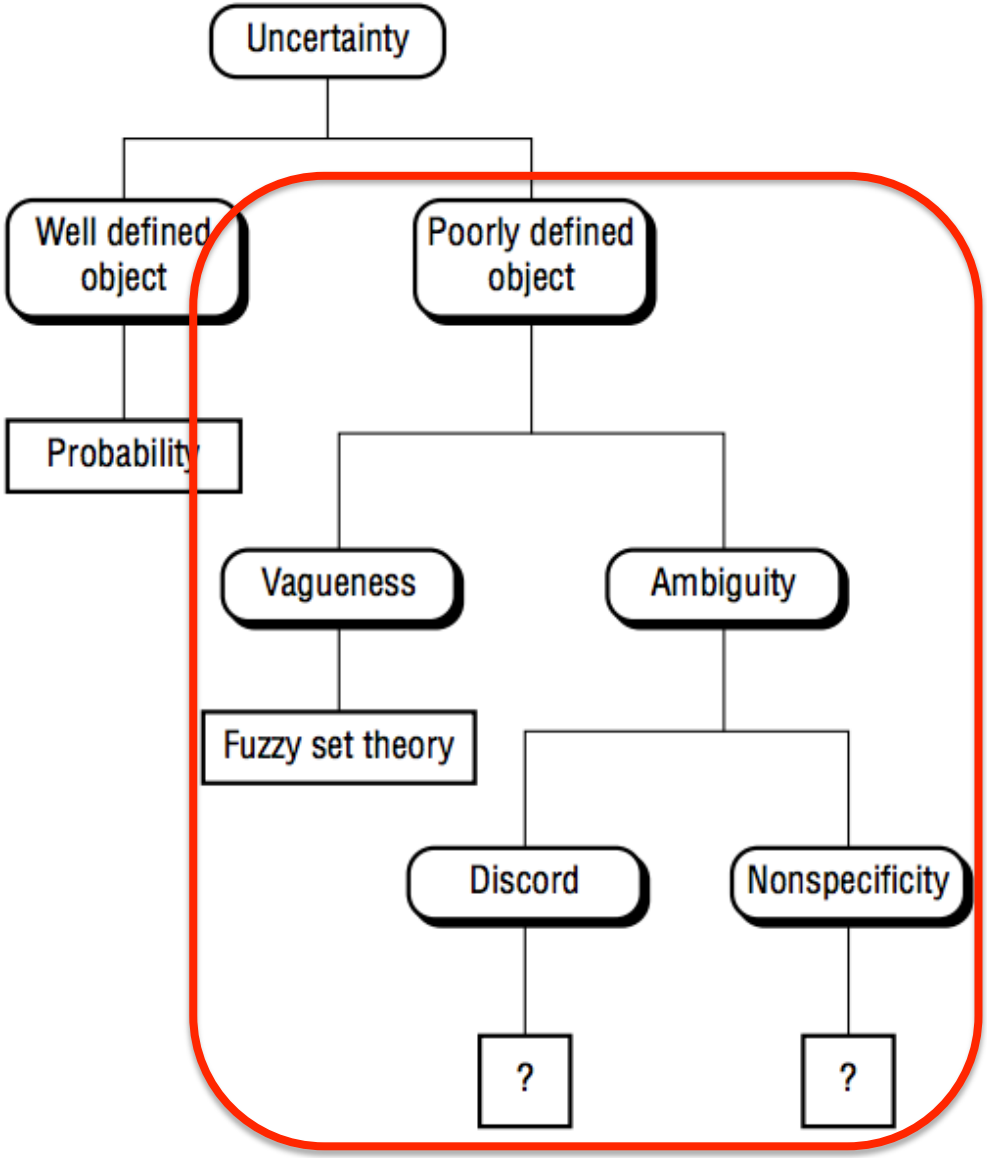
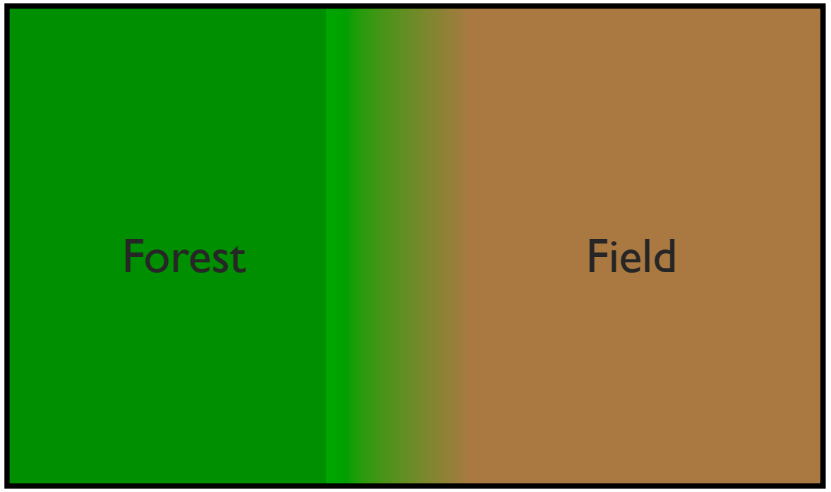


Fig 1. A conceptual model of uncertainty in spatial data (adapted from Klir and Yuan 1995: 268).

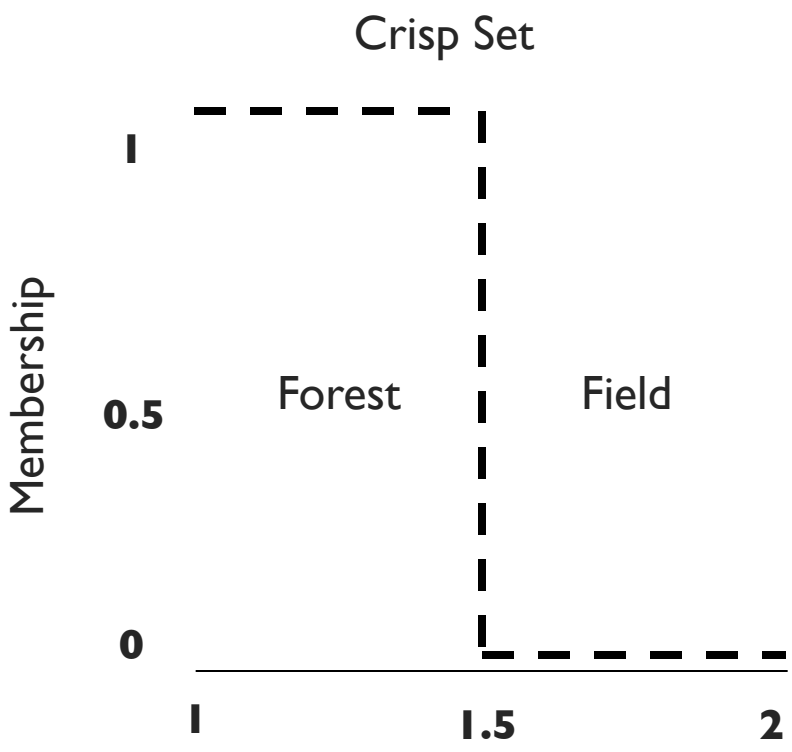
UNCERTAINTY FROM VAGUENESS

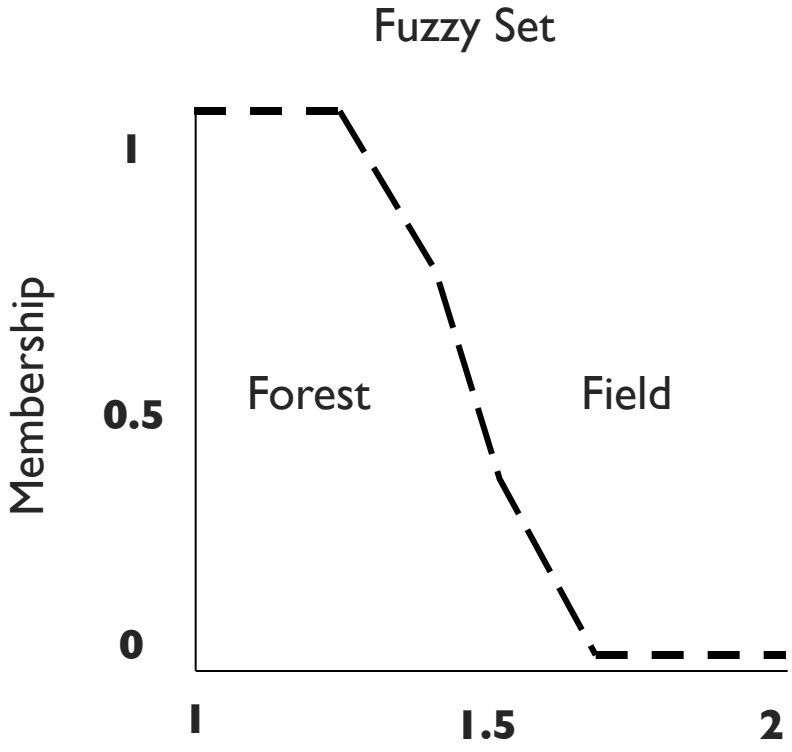
- One of the greatest limitations of traditional GIS methods is the way in which spatial objects are represented in a computer
- We are forced to take phenomena of continuous nature and impose a crisp representation
- Thus, we ignore the inherent uncertainty present in space



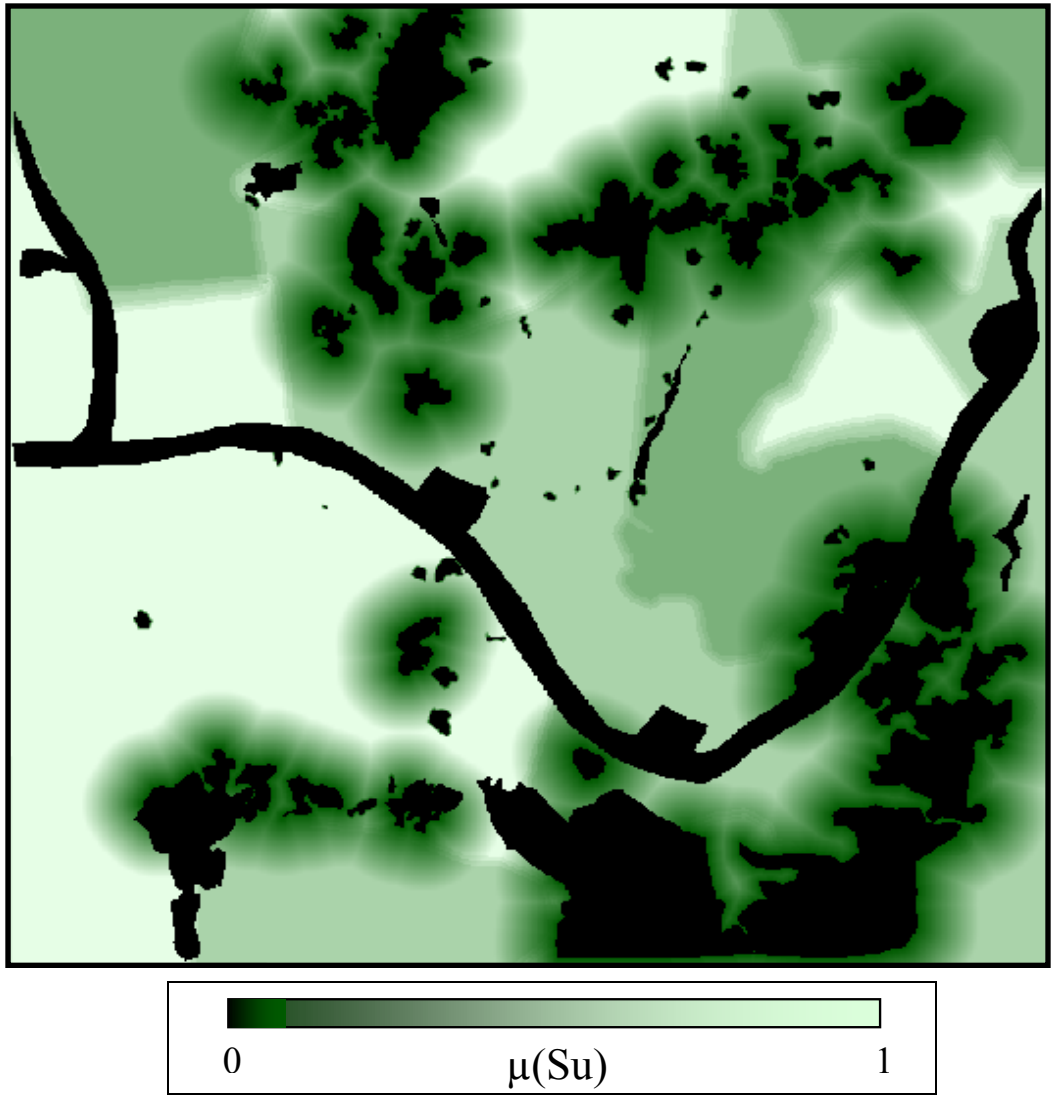
CRISP VS FUZZY REPRESENTATION

Fuzzy sets provide a method for representing the continuous nature of spatial phenomena





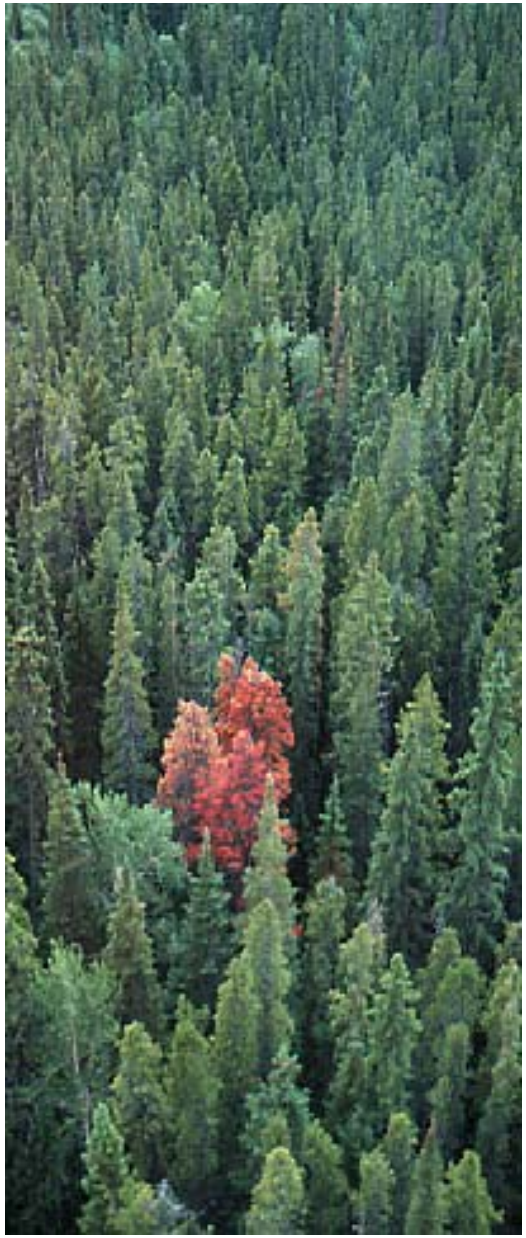
USE OF FUZZY SETS



Case Study

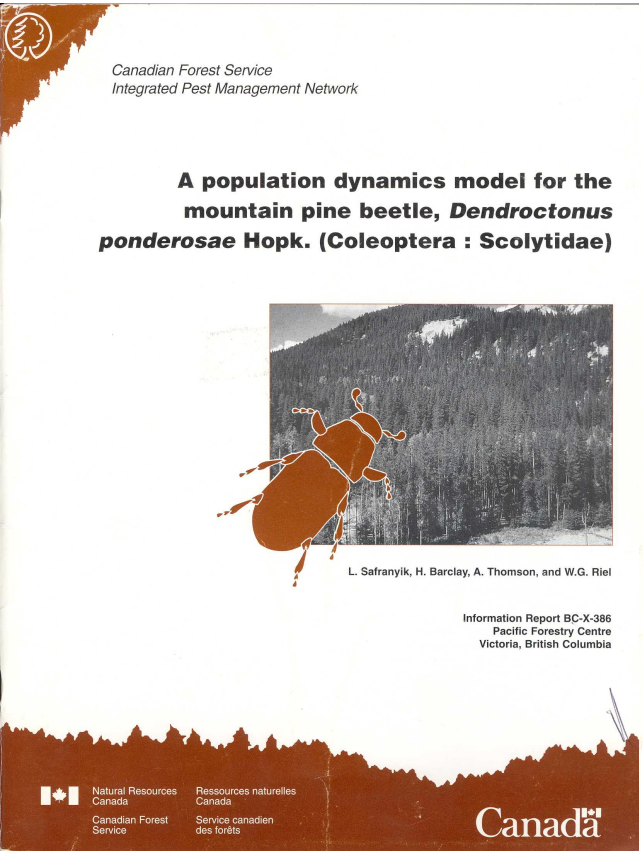
Patterns of lodgepole pine mortality caused by outbreaks of mountain pine beetle (MPB) in the central interior of British Columbia, Canada.

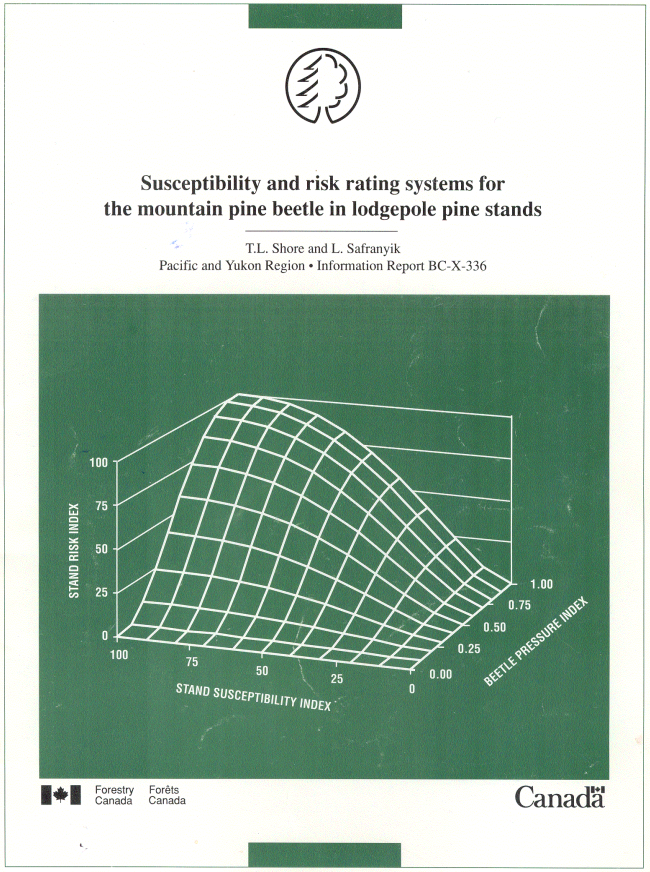
Photos from CFS, 2004

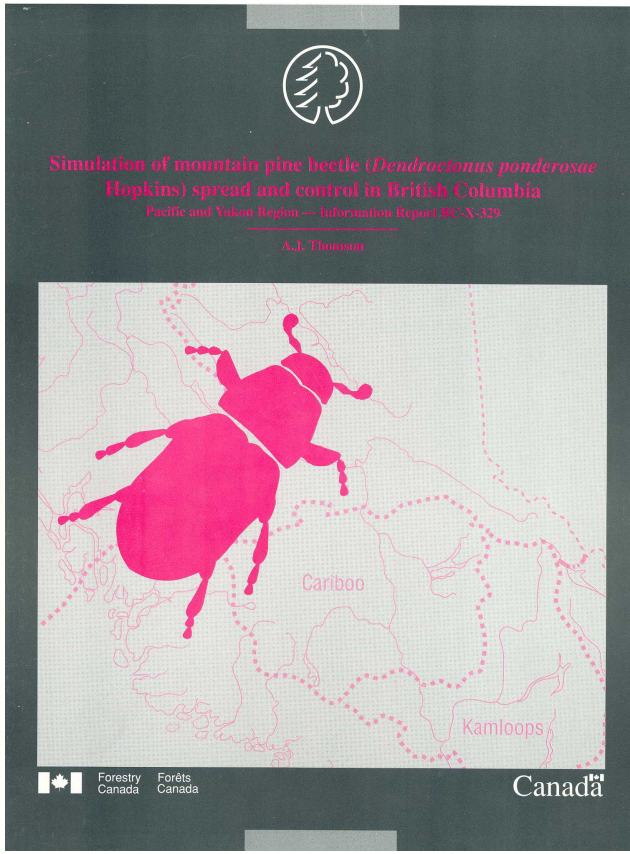


Photos from CFS, 2004

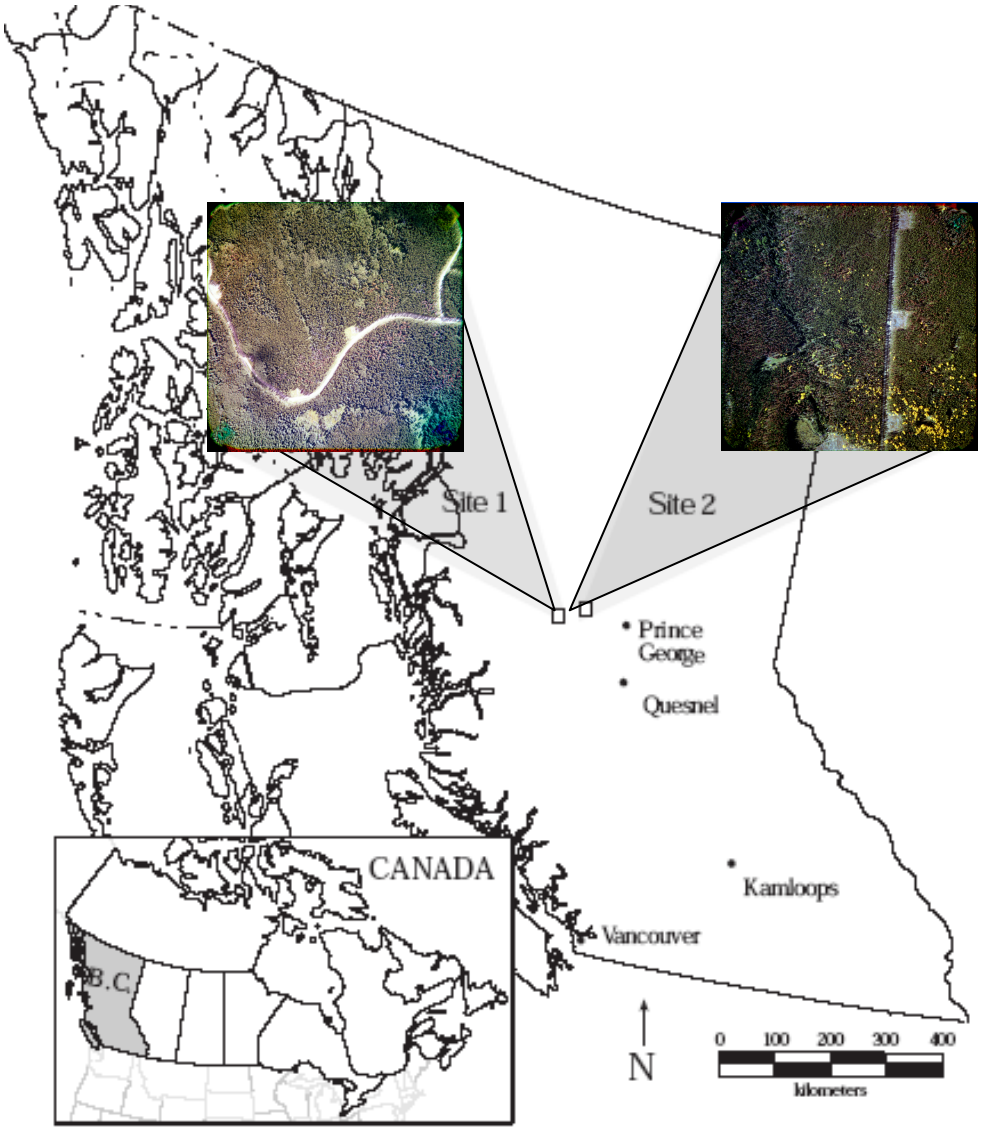
MPB Research







Study Sites and Data



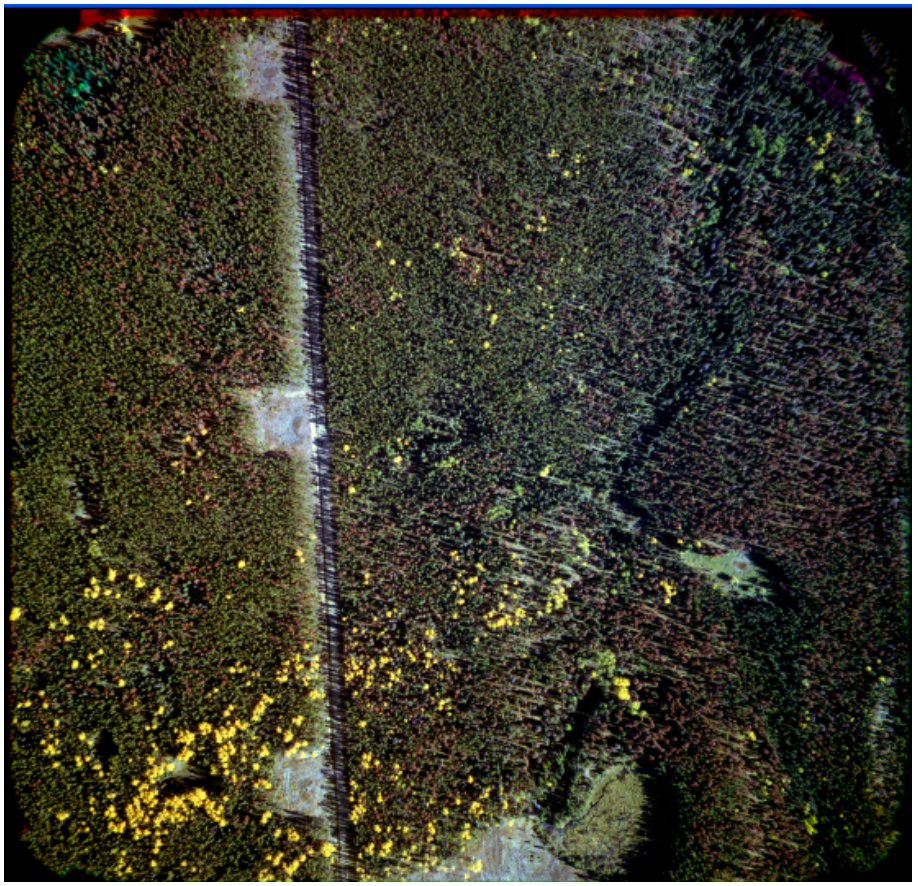
Step 1: Identify Susceptibility Variables

Stand Diversity	Amman and Baker 1972 Thomson 1991 Shore and Safranyik 1992
Distance to Previously Attacked Tree	Safranyik 1989 Thompson 1991 Shore and Safranyik 1992
Average Tree Size	Hindmarch and Reid 2001 Perkins and Roberts 2003
Distance to Large Deciduous Stand	Analysis

Step II: Image Interpretation

Site 1

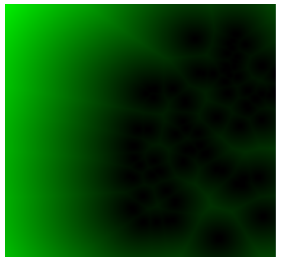
Site 2



Step II: Image Interpretation

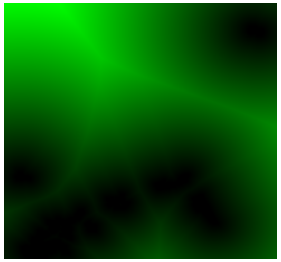
Stand Diversity

Distance to Previously Attacked Tree



Average Tree Size

Distance to Large Deciduous Stand

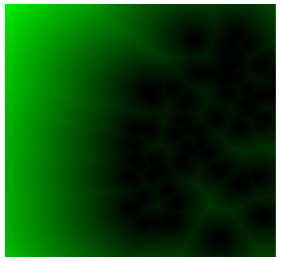


Step III: Fuzzification of Variables

Stand Diversity

$\mu(LP)$

Distance to Previously Attacked Tree

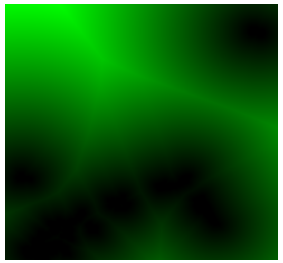


$\mu(AT)$

Average Tree Size

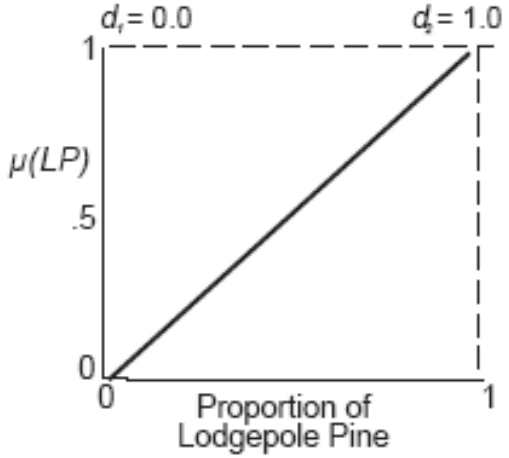
$\mu(LT)$

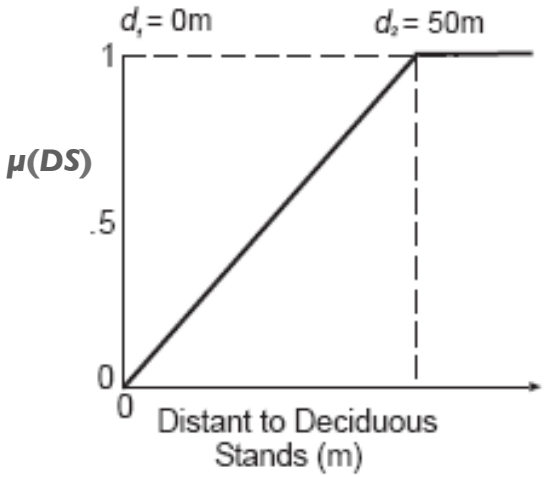
Distance to Large Deciduous Stand

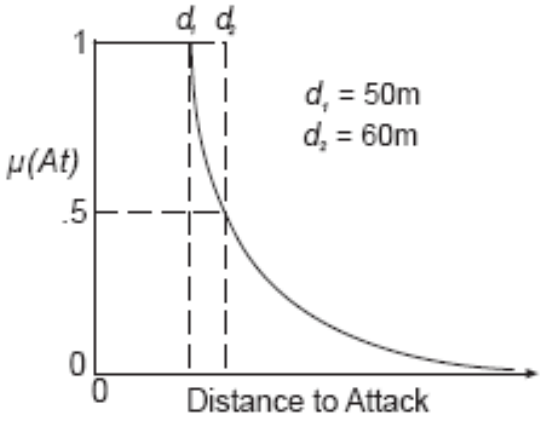


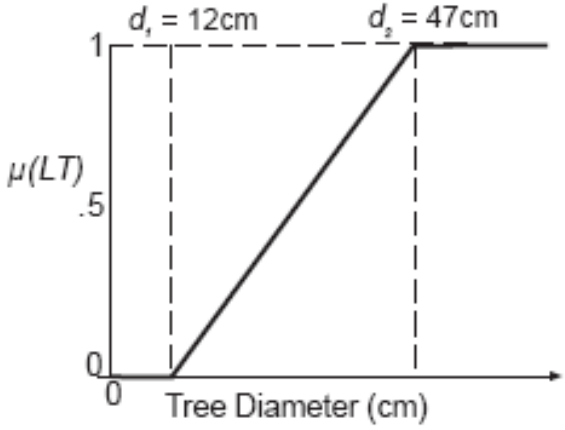
$\mu(DS)$

Step III: Fuzzification of Variables







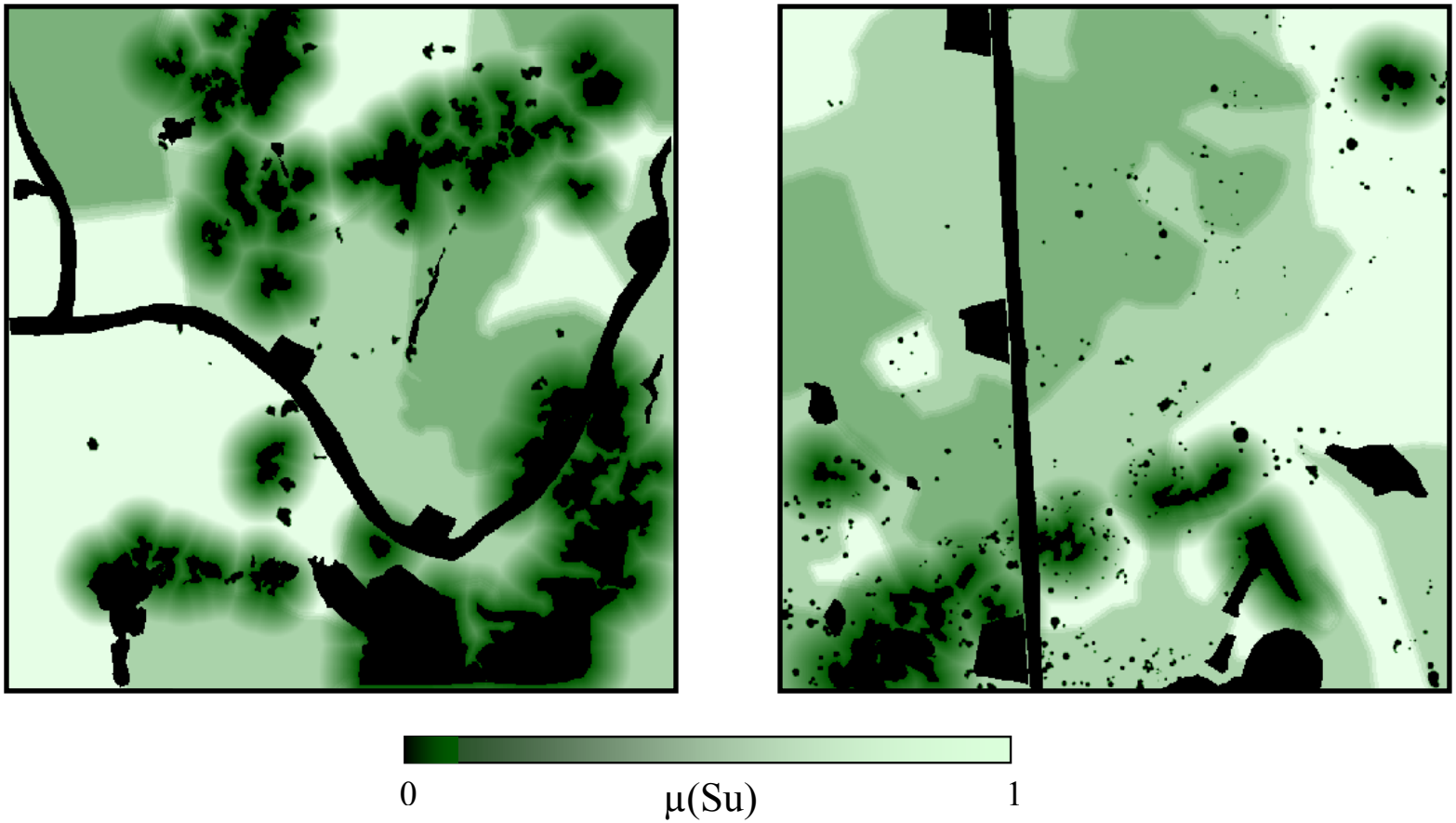


Step IV: Fuzzy Operation

Defining the Degree of Susceptibility $\mu(Su)$ of a Tree

$$\mu(Su) = \left[\mu(LP), \mu_{FZ}(LP)\right] \times \mu(DS) \times \mu(AT) \times \left[\mu(LT), \mu_{FZ}(LT)\right]$$

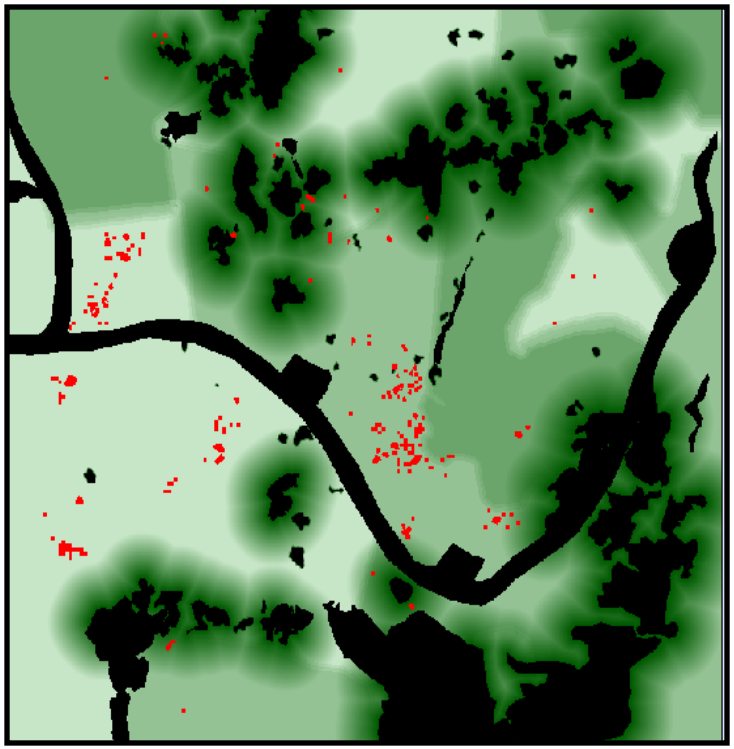
Susceptibility Model Output



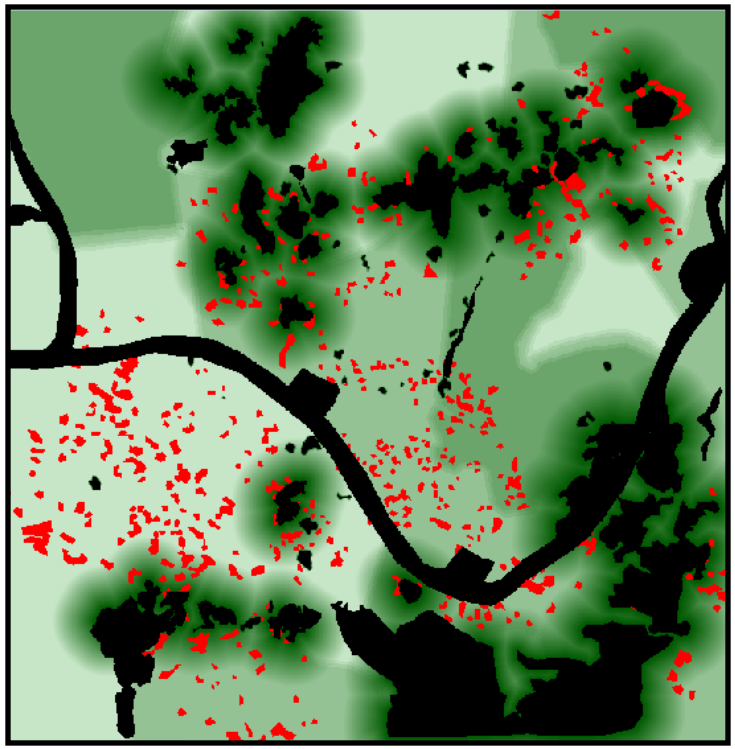
Step V: Validation

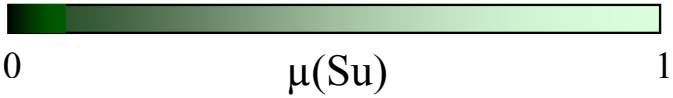
Site 1

Trees Killed in 2001



Trees Killed in 2002

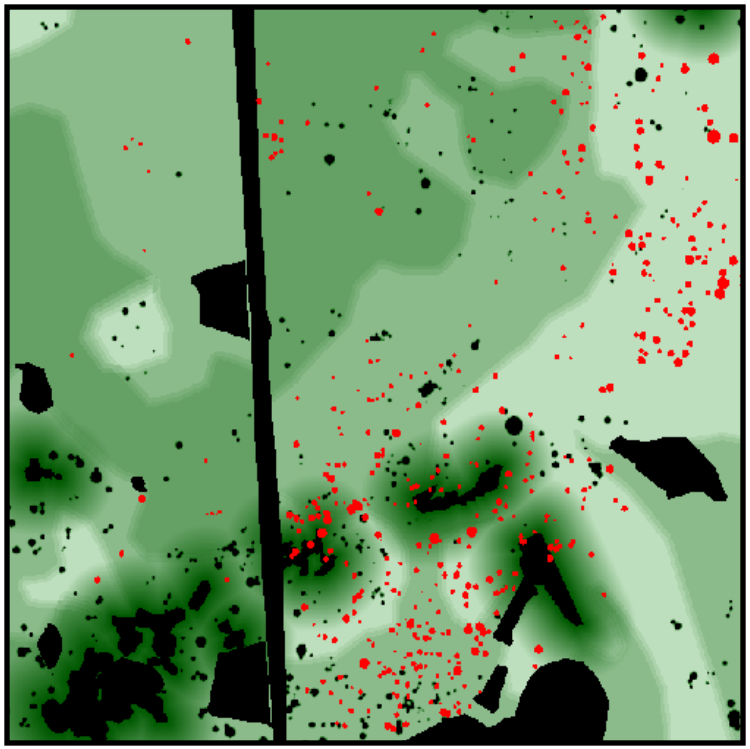




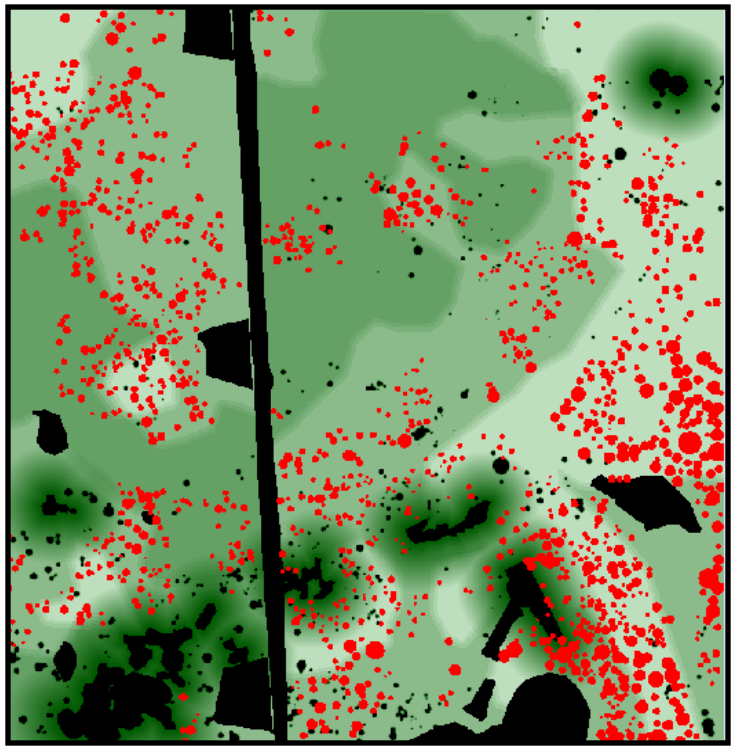
Step V: Validation

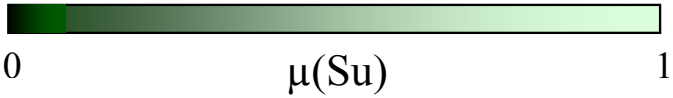
Site 2

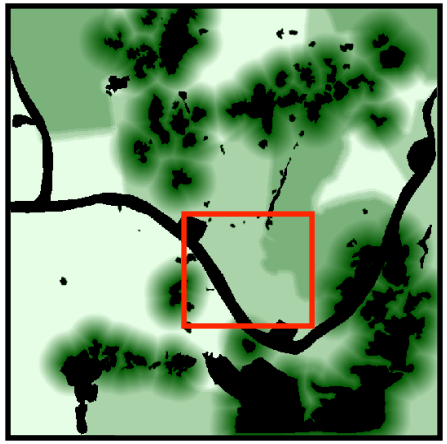
Trees Killed in 2001



Trees Killed in 2002





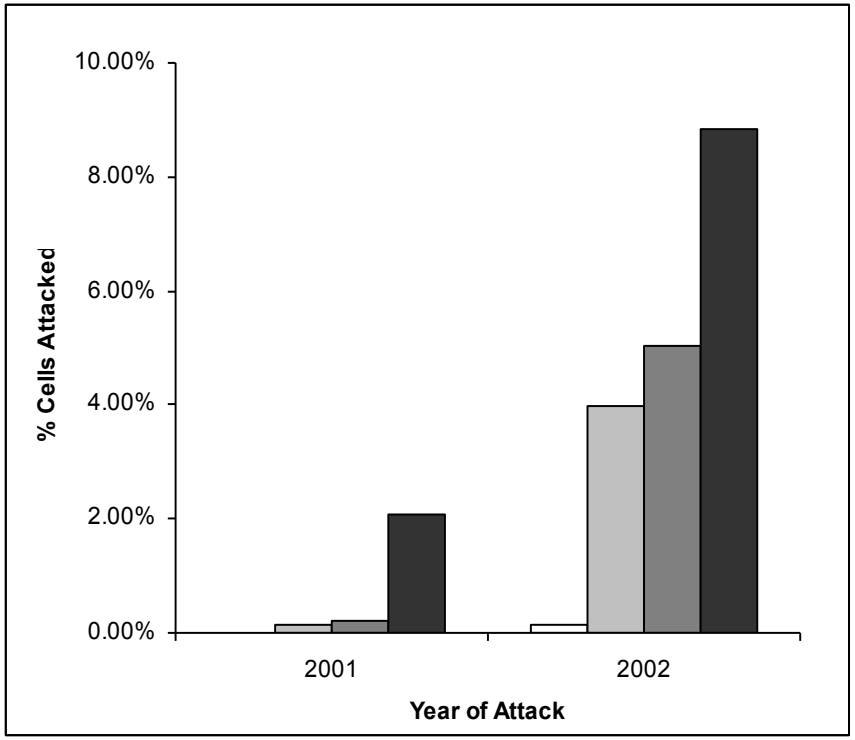


Trees Killed in 2001

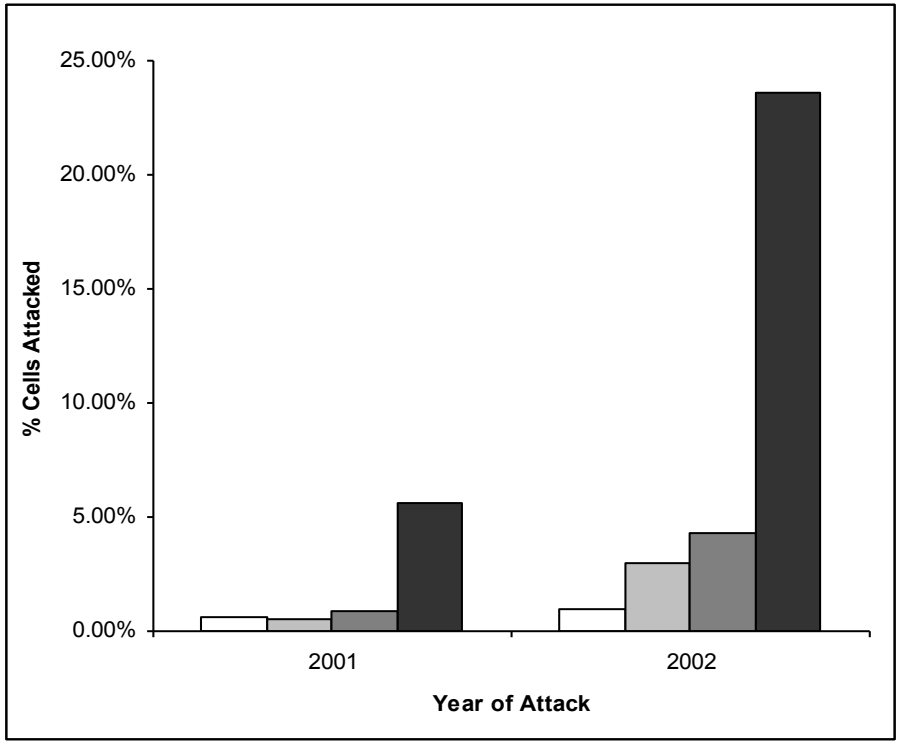
Trees Killed in 2002

Step V: Validation

Site 1



Site 2



Zero Susceptibility

Low Susceptibility

Medium Susceptibility

High Susceptibility

QUESTIONS TO THINK ABOUT…

- Where does there exist the potential for error in my project
- Is there any uncertainty in data representation?
- Is there any uncertainty in data location?

**How does all this affect the results of my analysis?