WORKSHOP
Spatial Uncertainty Propagation

Gerard Heuvelink
Wageningen University and Research Centre

James Brown
National Oceanic and Atmospheric Administration
GIS  = software tool for storage, analysis and presentation of geographical data
Analysis includes deriving new maps from existing maps:

- Input maps: $A_1, A_2, \ldots, A_m$
- Output map: $U$
output map = \( g(\text{input maps}) \)

\[ U = g(A_1, A_2, \ldots, A_m) \]

for instance:

slope angle = \( g(\text{elevation}) \)

erosion risk = \( g(\text{landuse}, \text{slope}, \text{soil type}) \)

soil acidification = \( g(\text{deposition}, \text{soil physical and chemical characteristics}) \)

crop yield = \( g(\text{soil properties}, \text{water availability, fertilization}) \)
Fact of life: data (maps) stored in the GIS database are rarely if ever error-free

Causes: generalization, digitization, measurement, classification and interpolation errors

Consequence: errors will propagate through GIS operation

Key research question: given the errors in the inputs to the GIS operation, how large are the errors in the output?
Error propagation analysis involves three steps:

1. DEFINITION of an error model for spatial objects and attributes
2. IDENTIFICATION of the error model (estimate its parameters)
3. Perform the actual ERROR PROPAGATION ANALYSIS
Before addressing these three steps, let us first discuss:

- What is error?
- What is uncertainty?
- What is the difference between error and uncertainty?
- How can we represent uncertainty statistically?
What is error?

- Error is the difference between reality and our representation of reality (assuming reality exists and is clearly defined).
- Example: population size Germany. We may estimate it as 110 million. Perhaps in reality it is 103,132,483; hence the error is $-6,867,517$.
- Error is usually not known because reality is not known. If error would be known, we would simply eliminate it!
- Definition of error needs slight modification for categorical variables: dominant car brand in Germany (my guess would be Volkswagen, but perhaps it is BMW).
What is uncertainty?

- Uncertainty arises when we are not sure about the ‘true’ state of the environment; it is an expression of confidence based on limited knowledge.
- Uncertainty is an acknowledgement of error: we are aware that our representation of reality may differ from reality itself and express it by being uncertain.
- Uncertainty is subjective; one person can be more uncertain than another.
- In the presence of uncertainty, we cannot identify a true ‘reality’. But perhaps we can identify all possible realities and a probability for each one.
Uncertainty can be described statistically with a probability distribution function (pdf)
Possible realities of objects with positional uncertainty

Rigid object

Deformable object
Possible realities of an uncertain spatial attribute
Example of an uncertain categorical spatial variable: possible realities of soil type
EXERCISE 1
1. DEFINITION of the error model, focus on quantitative attributes

\[ A_i(x) = b_i(x) + V_i(x) \]

\( b_i(x) \) is our (deterministic) estimate of the attribute (map stored in GIS), \( V_i(x) \) is the (stochastic) error about it (typically zero mean, but non-zero variance and spatial autocorrelation)

Error is difference between reality and our representation of reality:

\[ V_i(x) = A_i(x) - b_i(x) \]
Statistical model of error $V_i(x)$ must include:

- Marginal pdf at each location (shape and parameters)
- Spatial correlation (correlogram or semivariogram, usually invoke stationarity assumption)
- Temporal correlation (for dynamic variables)
- Cross-correlation with other uncertain inputs
2. IDENTIFICATION of the error model

- Measurement error from instrument and lab specifications or by taking replicates
- Sampling error using sampling theory from statistics (e.g. standard error of the mean, confidence intervals)
- Use of ground truth verification data (e.g. soils data bases)
- Interpolation error using geostatistics (kriging)
- Errors in transfer functions such as regression: R-square
- Classification error using multivariate statistics (e.g. maximum likelihood classification remote sensing imagery)
- Expert judgement (last resort?)