Data Quality in GIS

Contents

- Introduction
- Sources of inaccuracy/errors in spatial data
- Factors affecting reliability of spatial data

Introduction

- Quality of GIS products is usually judged by visual a**ppearance of** end-products on the computer screen, plotter, or video device
- Quality control by visual appearance is insufficient if data presented wrong or corrupted by errors
- Data accuracy is often grouped in terms of ***thematic accuracy***, ***positional accuracy*** ,*and* ***temporal accuracy***
  - But errors in spatial data can occur at various stages from observation to presentation.
- It is useful to know how errors and uncertainties occur, how they can be managed and possibly reduced and how knowledge of errors and error propagation can be used to improve our understanding of spatial patterns and processes

Introduction

- A good to understanding of errors and error propagation as leads to active **quality control**
- It is therefore very important to understand the nature of errors in spatial data and the effect they may have on quality of the analyses made with GIS
- Many GIS users conduct analysis under the implicit assumption that all data are totally 'error free'.
  - E.g. carefully drawn boundaries and contour lines on maps are elegant misrepresentations of changes that are often gradual, vague or fuzzy.
- NB: The term "**error"** in this discussion denotes both **'faults'** and the statistical concept of error, i.e. '**variation'**

Sources of errors in spatial data

- **Accuracy of content**
  - Whether the attributes attached to the points, lines, and areas of the geodatabase are correct or free from bias
  - **Qualitative accuracy:** whether nominal variable or labels are correct
  - **Quantitative accuracy:** level of bias in estimating the values assigned
  - Accuracy can be ensured by having reliable, documented input and transformation procedures.
- **Measurement errors**
  - Unreliable, inaccurate or biased observers or apparatus
  - Reader should understand between difference between **accuracy** and **precision**
- **Field data**
  - Well designed data collection procedures and standards help reduce observation bias
  - The quality of data sometimes depend on the interpretation capabilities of the surveyor
  - Automated sampling device linked to electronic loggers is ensuring that accuracy and precision of data are good

Sources of errors in spatial data

- **Laboratory errors**
  - Analysis carried out in different laboratories may produce different results
  - Affected by also poorly prepared samples e.g. soil samples
  - Automation in lab analyses is reducing errors from labs
- **Locational accuracy**
  - Determined by judgement of a surveyor e.g. the boundary between vegetation types
  - Positional accuracy can result from poor field work, distortion of original paper base map, poor quality vectorizing after raster scanning
  - Local errors can be corrected interactively on the screen while others are through transformation
- **Natural spatial variation**
  - Thematic mapping usually assume homogeneous distribution of a phenomena within crisply delineated areas
  - Information about gradual change within boundaries could not be represented on conventional chorochromatic maps
  - It is important to realise that unseen spatial variation of phenomena like soil, lithology, or water quality can contribute greatly to the relative and absolute errors of the results produced by modelling and map overlay

Factors affecting quality of spatial data

- **Age of data**
  - Are data up to date?- reliability of data decreases with age
  - Cost of data collection is influencing the use of already collected data
- **Areal coverage**
  - Is it partial or complete?
  - No complete cover of certain kinds of thematic information
    - Leads to using different methods to fill, or to generalization of detailed data to match less detailed data
- **Map scale and resolution**
  - It is necessary that scale of source map matches that required for study
  - User should choose the appropriate scale for the task at hand.
- **Density of observations**
  - There is need for publishing data that was used in generating maps such as the density and location of sampled points
  - Knowledge of density of observation is good to be able to resolve the spatial patterns of interest

Factors affecting quality of spatial data

- **Relevance**
  - Not all data used in geographic information process are directly relevant for the purpose for which they are used, but have been chosen as surrogates because desired data does not exist or are too expensive
- **Data format, exchange and interoperability**
  - Affects when data are created in different formats which have different ways of recording essentially similar entities
  - Need for reformatting so as to exchange
  - Need for interoperability: open GIS Consortium (OGC)
- **Accessibility**
  - Not all data are accessible: data could be state secret
  - Costs and formats can affect accessibility

Factors affecting quality of spatial data

- **Cost and copyrighting**
  - Method adopted to collect data so as to cut on costs
  - It is also necessary to know the legal situation in each country e.g. when digitizing maps or using spatial data for research or commercial application.
- **Numerical errors in computer**
  - Ability of the computer to store and process data at the required level of precision
  - This has consequences for both arithmetical operations and for data storage
  - Errors such as rounding off, and truncation have consequences in some GIS operations such as scaling, rotation