

# IMPORTANCE OF QUALITY IN SPATIAL STUDIES

## Aspects of Data Quality

Application fields for spatial data are very broad and quality aspects are important because the usability of results depends upon them. In the international arena there is increasing attention for a numerical approach to data quality in today's GIS. Information produced by the GIS specialist should include a quality-description component, often based on error propagation. The author discusses the importance of data quality in spatial studies.

Data processing involves the storage of data and its quality. Low-quality data will result in "low-quality" output: "garbage in is garbage out". Not all applications require accurate input data; meteorological data and air-quality data are typically collected at relatively few locations. Elevation data, on the other hand, is available at high density. Level of detail or resolution also applies to satellite data; for some applications a few spectral bands will suffice whereas others, in particular mineralogical and biodiversity studies, benefit from many bands, typically in excess of two hundred.

### Positional Definition

The precision of coordinates affects many of the other processing steps in a GIS, such as computation of the length of a pipeline or the area of a parcel. Many decisions, such as cost calculations and subsidy values, are based on such estimates and can thus be affected by them. Let us take as an example that demonstrates the effect of precision of coordinates on derived parameters a rect-angular surface given by four points two metres apart. Ideally, the surface would be 4m<sup>2</sup>. However, when the precision of the coordinates is 1m the area may vary from 2 to 7m<sup>2</sup>, and when precision increases to 0.01m there occur values between 3.95 and 4.05m<sup>2</sup> (Figure 1).

### Reliability

Attributes can be distinguished in terms of qualitative attributes (names) and quantitative attributes (numerical values). The reliability of qualitative attributes can be expressed as the chance that the wrong name has been assigned or that errors have occurred in the nomenclature. The reliability of numerical attributes is often a matter of rounding off or uncertain measurement. Recent research has indicated that this can play an important role when, for example, defining air quality in the Netherlands on the basis of a limited number of measurements. Figure 2 shows the number of days that the environmental standard for ozone is exceeded. Such graphs play an important role in decision-making at national level. In this research environmental models able to calculate numerical values on a finely meshed grid are linked to scarcely available point-data. Given the available means, an error-in-variables approach was useful in estimating the risk as realistically as possible.

### Logical Consistency

A GIS database must be logically consistent; what is stored in a GIS has to correspond with and to represent that reality. For example, a forest within a town is not logical, while a town in the middle of a forest is more likely. Research on metadata, which summarises in a few numbers the contents of a database, is emerging in this field. Current metadata for quantitative maps, for example, also includes diverse information on contents such as average values, standard deviations and numbers of polygons. This field of research is still in its infancy but may receive a boost in the near future.

### Time Factors

With respect to time, typical questions include whether data was collected simultaneously, whether the most recent data has been used and stays so, and whether spatial data quality remains stable over time. The method of data capture is relevant. If satellite images have been used, what is their quality? If precipitation measurements have been used, were they calculated using a consistent method? Scale in time plays the crucial role. Data quality may change according to season; for example the presence of bare ground rather than crop, or due to day-night fluctuation. A recent study on the detection and tracking of forest fires in Portugal by satellite sensors showed that fire was much easier to predict at night than during the day: daytime temperature was so high that thermal bands did not show much difference between the fire and the non-burning background.

## Inherent Uncertainty

The above quality factors relate to data collected from objects. But uncertainty is also introduced by the objects themselves. A building such as a church is clearly defined but objects such as towns, roads and forests offer fuzziness of definition. What is actually implied by distance or the centre of the region for such objects. Such uncertainty is not a matter of precision but is inherent to the objects. Recent research has also focused on this quality aspect; various vague objects were identified using a rigid mathematical approach (Figure 3), algebra was defined and operations investigated to compute, for example, perimeter, distance and length of fuzzy objects. Implementation employed public domain software package GRASS.

## Exchange

A major issue remains the communication of uncertainty to stakeholders. In the end most stakeholders appreciate a crisp statement. Recent discussions have revealed, however, that the general public is also able to distinguish between “low probability” and “high probability” risk assessment. “Fuzzy decision trees” have been suggested as an important step towards the communication of uncertainty and spatial data quality. Differences in statistical and vague/fuzzy approaches to similar problems still present a challenge, and the same is true for data modelling: should a grid or a vector be used to optimally deal with a spatial problem

## Concluding Remarks

Thanks to the ever-increasing exchangeability of spatial data, data quality is a theme once more in the spotlight at the centre of the international stage.

## Further Reading

Artà Dilo, 2006. Representation of and Reasoning with Vagueness in Spatial Information. PhD dissertation, ITC and Wageningen University.

Rodolphe Devillers and Robert Jeansoulin, 2006. Fundamentals of Spatial Data Quality. ISTE, London.

Jan van de Kassteelle, 2006. Statistical Air Quality Mapping. PhD dissertation, Wageningen University.

Pepijn van Oort, 2006. Spatial Data Quality: From Description to Applications. PhD dissertation, Wageningen University.

Daniel van de Vlag, 2006. Modelling and Visualising Dynamic Landscape Objects and their Qualities. PhD dissertation, ITC and Wageningen University.