Scaling effects in a multi-layered DSS-structure

Steffen Bender, Tilman Mieseler, Till Rubbert, Stefan Wohnlich
Chair of Applied Geology, Ruhr-University Bochum, D-44780 Bochum, Germany

Keywords: DSS, scaling, nitrate, hard rock, water management

INTRODUCTION

Decision support systems (DSS) are a specific group of computerised information systems that support management, business and organisational decision making activities. In case of environmental questions, these information systems include data implementation in a Geographical Information System (GIS) to visualise results. Due to the availability of information of various scales, the DSS-structure is built of several layers containing miscellaneous data clusters of different significance. One of the main problems that have to be accounted for is caused by scaling effects, which are mainly influenced by 1) existing environmental heterogeneity, 2) regionalisation of an inhomogeneous data pool, and 3) scale differences between input parameters and resulting output. Even though sampling and analytics can be controlled by the use of guidelines for quality assurance, the main source of errors is imposed by the compilation of punctual information into data of spatial distribution. Data affected by different anthropogenic or geogenic impacts may cause high fuzziness of results, which are then unsuitable for a DSS.

INPUT DATA

The main basic requirement of each information system is to guarantee high quality of input parameters. With regard to a pan-European platform, a second requirement is necessary: a standardised input frame to provide a) comparability of different sites, b) assignability of information from well-investigated reference areas to other test sites, and c) the possibility of cross-border activities e.g. hydrogeological work in transboundary catchments. The density of data per area is controlled by the type of monitoring tool and monitoring strategy, depending on local or regional interest. For many subjects, there exist several monitoring networks on different scales such as a) state-wide monitoring networks (broad range of parameters), b) regional networks (selected group of parameters) and c) special networks for local problems (indicating parameters) (fig.1). Only in very few cases there exist tandem networks (FRAPPORTI et al. 1995) or standardised information systems such as LOWRGREP-GIS (BENDER et al. 2001b), where data can easily be combined. For the TRANSCAT-project dealing with the management of transboundary catchments and containing heterogeneous data pools characterised by a strongly varying density of spatial information, the first step to a GIS based decision support system is to evaluate and classify information from different sources. While the selection of standard procedures for data collection is a minor problem, the implementation of regionalised spatial data based on different scales within one capacious database is nontrivial.

![Proposed monitoring network (springs) in Bavaria](image)

Fig.1: Integrated monitoring network using three different scales (BENDER et al. 2001a)
Due to varying demands of information, it is normally useful to work in different scales to decrease monitoring and analytical efforts, which lowers costs and helps to avoid the impact of geogenic heterogeneities. But when working with a GIS offering different scales depending on data density, users are tempted to zoom in and out between different scales, which normally means to change between independent systems. To prevent false interpretations, reducing the quality of the whole DSS, there must be visible limitations to point out the difficulties when working in different scales. There are two main influencing variables which are responsible for these scaling effects: a) incompatible clusters of data due to environmental heterogeneities or logistic differences, and b) variation of spatial distribution of input information. Both factors of anisotropy avoid the possibility to create spatial distribution maps out of point data. According to ANLAUF & LIU (1988), it is possible to recognise anisotropic effects when using 100 or more measured values. But generally this number of data are only available in few regions. A more feasible approach is suggested by JOURNEL (1986), who uses a justifiable guidance level on the variation coefficient of the data pool to be between 20 and 50%.

SPATIAL DISTRIBUTION OF NITRATE
Nitrate is one of the most important chemical parameters concerning water contamination. Therefore it combines the different characteristics of a statistical variable: on the one hand widespread measurement programmes, important component for hydrochemical analysis and easy analytical determination methods, on the other hand anthropogenic impact due to fertilisation as punctual and diffusive contamination source.

Fig. 2: Detailed investigation of nitrate input near a local water supply (OBERMANN et al. 2003)

The actual state concerning nitrate impacts is generally monitored in a state-wide network. In North Rhine-Westphalia (NRW), the data pool contains nitrate concentrations of untreated water as well as of drinking water. Data of such a scale normally give a rough estimation only, where a contamination may possibly be expected. Investigations on this scale allow the observation of general trends for certain clustered areas. Still, it has to be stated that the practical use of such generalised mean values is quite restricted. As opposed to detailed investigations, no statement on problems such as contamination in particular water wells, neither practical suggestions for solutions are possible. To show differences due to variations of scale, results for the state-wide network may be compared with detailed investigations in a selected catchment near a water works in NRW. Whereas in the state-wide network one more or less representative piece of information is used for regions of perhaps more than 100 km², the detailed study focuses on a differentiation of all agricultural pads in crop land and grassland (fig. 2). The analysis of nitrate at the groundwater surface shows a broad variation of nitrate contents in
dependence of type of soil, type of agricultural usage and quantity of fertilisation. So, for local questions the quality of small-scaled data is rather useless, due to differences of mixed samples and depth-specific zonal formation of concentration profiles.

**VISUALISATION OF DATA**

For the visualisation of data, there are two main options: depiction of punctual information (boreholes, outcrops, springs, wells) or regionalised maps of spatial distribution. While the first method shows the current stage of knowledge, this type of information is not useful for a DSS-system, because most parts of the test site are unknown area in terms of existing punctual information. Therefore, data has to be regionalised to offer simulation or calculation options for a GIS-system such as the development of pollution load maps or estimations of the acidification risk (HRKAL et al. 2003). Due to the heterogeneity of data distribution and a data pool structure built of different data clusters as a result of various characteristic impacts, the calculation of spatial maps using kriging or similar interpolation methods produce regionalised maps with high uncertainties. The application of a weighting scheme, which is used to optimise hydrogeological models (MIKULLA 1998), is often impossible due to small amount of data. For areas in hard rock regions the results of spatial interpretation of heterogeneous tectonical structures and of the hydrogeological situation are influenced by the selected scale of investigation (KRÁSNÝ 1993). At LOWRGREP-test site Rötz-Schönthal (fig.3), a typical area for agricultural usage in a forest dominated landscape, the calculation of spatial distribution based on a heterogeneous data pool shows the problem of up- and downscaling of image information. A typical hydrochemical situation is depicted in fig. 3. The spatial distribution of nitrate was calculated according to the inverse distance method. It shows three centres of anthropogenic impact with increased concentrations. As a result of direct neighbourhood of fairly low and highly increased values, distribution maps based on kriging methods produce intensively smoothed out values partly useless for a DSS. Furthermore, it is shown that the selection of monitoring sites for representative characterisation of the regional situation should be done according to detailed investigations only. Otherwise, the test site could be classified either as uncontaminated (nitrate values: 5-10 mg/L) or as highly contaminated (over 50 mg/L). So, for useful interpretations it is necessary to depict spatial maps in combination with the measuring points or sampling locations to get an idea about the quality of the displayed information.

Fig. 3: Spatial distribution of nitrate and location of nitrate-monitoring sites (BENDER 2000)
CONCLUSION
The comparison of information of different scales shows the possibilities and also the problems imposed by the use of low-detail, small-scale regionalised data, giving an overview on a potentially vast area, in contrast to highly detailed large-scale data. Spatial data from small-scaled maps can be used for depiction of regional trends only. Zooming into larger scales seems to be practicable if upper and lower scales which are to be displayed are closely related. Anyway, it has to be guaranteed that predictions based on small-scale data of low detail and possibly affected by intense smoothing, are still based on representative values to be suitable for interpretations. Therefore, it is necessary to show a combined map of spatial distribution and locations of sampling or measurements, which may serve as quality assurance. Thus, it is possible for an expert to identify both density of information and anthropogenic as well as geogenic anomalies or heterogeneity. To prevent misinterpretation, it seems to be essential to include information about the degree of inaccuracy or confidence intervals to restrict the scope of interpretation for the DSS endusers. Finally, it has to be mentioned that scaling effects with importance for vast databases are not only restricted to spatial scale, but that they are equally important concerning the dimension of time, as for sampling intervals.

REFERENCES