Spatial analyses and spatial data quality

Václav Talhofer[1], Alois Hofmann[1], Šárka Hošková – Mayerová[1], Petr Kubíček[2]

[1] University of Defence, Faculty of Military Technology, Kounicova 65, 662 10 Brno, Czech Republic

[2] Masaryk University, Faculty of Sciences, Geographical Department, Laboratory of Geoinformatics and Cartography, Kotlářská 2, 601 67, Brno, Czech Republic

{vaclav.talhofer, alois.hofmann, sarka.mayerova}@unob.cz, kubicek@geogr.muni.cz

ABSTRACT

The aim of this contribution is to explore how we might better understand geospatial data using sophisticated analysis tools, to introduce some delightful equations to give objective rigor to otherwise subjective guesswork and to show how we could bring a degree of robustness of analysis. Aforementioned approaches are tested on a pilot study dealing with the selection of a route by measuring the complexity of the terrain and the ability of the Czech Army trucks to navigate such routes, as well as to find the possible way how this task gaining from praxis can be integrated into teaching process.

Keywords: data evaluation, data matching, data quality, spatial analyses

1. SPATIAL DATA QUALITY

The quality evaluation generally results from common schema of quality components, in which both technological aspects and aspects of safety, security etc. are considered. The ability aspects are also evaluated in the relationship with a given process of usage.

Many authors dealt with application of common quality definition into evaluation of *digital geographic data and information* (DGI). Finally, recommendations of International Organisation for Standardisation (ISO) were developed from their results. In the table there is the list of elements and subelements as described in ISO 19113:

Table 1 Elements and subelements of data quality according to ISO 19113

Element of quality	Subelement of quality
Completeness	Commission
	Omission
Logical consistency	Conceptual consistency
	Domain consistency
	Format consistency
	Topological consistency
Positional accuracy	Absolute or external accuracy
	Relative or internal accuracy
	Gridded data position accuracy
Temporal accuracy	Accuracy of a time measurement
	Temporal consistency
	Temporal validity
Thematic accuracy	Classification correctness
	Non-quantitative attribute correctness
	Quantitative attribute accuracy

Each product including DGI has to be made for the specific user and only his satisfaction with the product is the final criterion for quality of this product evaluation. Usability as an expression for a product's potential to accomplish the goal of the user is often mentioned term. Usability can be described as some system which enables to combine different possibilities of expression of quality parameters. The application of the *Value Analyse Theory* (VAT) (Miles 1989) is one of possibilities.

2. GENERAL ASSESSMENT OF SPATIAL DATA QUALITY

The product or a part of the product resultant function utility degree may be assessed based on the criteria k_i for the spatial geodatabase utility value evaluation using a suitable aggregation function F (Talhofer, Hošková, Kratochvíl, and Hofmann 2009):

$$F = p_3 k_3 p_4 k_4 (p_1 k_1 + p_2 k_2 + p_5 k_5) \quad (1)$$

The chosen form of the aggregation function concerns also the case the user gets data on an area beyond his interest or data obsolete so that their use could seriously affect or even disable the *digital geospatial information* (DGI) functions. The weight of each criterion is marked as p_i, where $i = 1,\ldots,5$. The aggregation function proves the product status at the required content and its utility rate. It is applicable also to experiments to find the ways of how to increase product utility at minimum cost increment.

The DGI are usually developed and maintained by individual partial components of the complete database, such as save units, measurement units, map sheets etc. Therefore, it is quite a good idea to assess their utility value in the above-described system within the established the storing units introducing *individual benefit value*. Similarly the individual benefit value can be applied for the selected part of master databases from given *area of interest* which is used for certain task.

When assessing database utility, the *ideal quality level* is defined at first. The ideal level is used as a *comparison standard* to express each criterion compliance level. Using the comparison standard the individual criteria compliance level and consequently aggregate utility is assessed.

The compliance level of each individual criterion $u_{n,s}$ is given as follows:

$$u_{n,s} = \frac{k_s}{k_s^*} \quad (2)$$

where

- k_s is the value of s^{th} criterion compliance,
- k_s^* is the level of compliance of s^{th} criterion or its group criterion of the comparison standard.

Than the aggregate individual benefit value (*individual functionality – U_n*) of the n^{th} save unit is defined by the aggregation function of the some type as (1). Therefore:

$$U_n = p_3 u_{n,3} p_4 u_{n,4} (p_1 u_{n,1} + p_2 u_{n,2} + p_5 u_{n,5}) \quad (3)$$

The individual criteria weights are identical with the weights in database utility value calculation.

Particular criteria usually consist of several sub-criteria. The authors took 20 criteria into their consideration; hence the equation for calculation the aggregate individual utility value is therefore a function of 20 variables that characterize the levels of compliance for each individual criterion.

The DGI benefit cost assessment including individual benefit cost is a task for a data manager or a geographer-analyst which is responsible to provide demanding project. The system enables him to consider which quality parameters are possible to improve in given time, with given technological conditions, with given sources, with given co-workers etc.

3. SPATIAL ANALYSIS AND DECISION MAKING PROCESS

The number of various spatial databases is increasing in all regions and states and powerful *Spatial Data Infrastructures* (SDI) are built in many countries. SDI development can help to government organisations such as armed forces, fire brigades, emergency rescue systems etc. in their activities, but it can cause a strong dependency on the whole infrastructure. Therefore the SDI development has to respect a *sustainable level of data quality*. If the state budget is restricted in all states and number of workers in geoinformation 'industry' is decreasing the determination of appropriate level of SDI quality is very important task.

Spatial analyses are usually developed as a source for a decision making process that contains demanded information about geographical conditions in the area of interest or responsibility which can have important impact on a given or supposed situation. ***Figure 1*** illustrates the scheme of usual procedure of spatial analyse creation in armed forces.

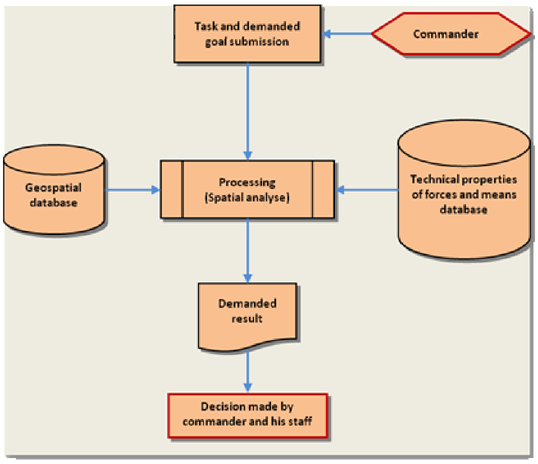

Figure 1 Usual process of spatial analyses for the decision making process

This process has several advantages and disadvantages. The *main advantages* are:

- relatively simple solution,
- unambiguous result,
- the commander does not have to think about received result and its properties.

The spatial analyses are created from given spatial data and by the use of a given mathematical model, so the quality of spatial data and the quality of the mathematical model can significantly affect the result. Therefore the spatial analyses have the following *essential disadvantages*:

- total dependence of spatial data,
- the quality of result is unknown,
- without any supplementary information about quality of demanded result the commander on duty can only make one decision (he has no choice)

In order to overcome the aforementioned disadvantages we propose to take the spatial data quality assessment into account.

If the spatial data quality is considered in the process of spatial analyses creation, not only one result can be delivered to the responsible commander. Geographer-analyst can prepare several solutions, e.g. reliable, less reliable; but according to previous knowledge about given area of interest still bearable within a certain level of risk. If commander obtains given solutions and additional information about quality properties, it is up to him what will be next steps. Either he is satisfied with all information he has obtained or orders to geographers to improve database and prepare a new version of given spatial analyses. Next picture illustrates this process (***Figure 2***).

Two important questions have to be answered if such as this solution is accept.

1. Is it possible to use the risky or less reliable solution when the reliable solution is not possible to use due to tactical situation in the battlefield or timely restricted area which given solution passes?
2. What measures is necessary to take in order to increase reliability of spatial analyses solution and final decision? What expenses will be necessary for it – financial, personal, time etc.?

The system of spatial data quality evaluation and application of the VAT should help to answer most of previous questions. Next figure illustrates a process of spatial analyses for decision making process with the data quality consideration.

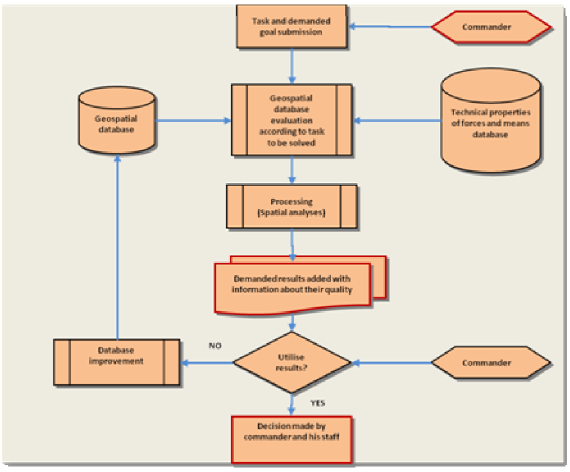

Figure 2 Process of spatial analyses for decision making process with the data quality consideration

4. PILOT STUDY

In order to verify the VAT methodology the task of *Cross Country Movement* (CCM) was chosen as an example. CCM can be solved as a common problem or with consideration of certain types of vehicles. The detailed theory of CCM is in (Rybansky and Vala 2010).

The solution can offer to the commander not only one possibility, but the variants from which he can choose according to his intentions and the current situation at the given area.

4.1 Cross Country Movement

The main goal of CCM is to evaluate the impact of geographic conditions on of a movement of vehicles in terrain. For the purpose of classification and qualification of geographic factors of CCM, it is necessary to determine:

- particular degrees of CCM
- typology of terrain practicability by kind of military (civilian) vehicles
- geographic factors and features with significant impact on CCM

As a result of the geographic factors impact evaluation we get three known degrees of CCM:

- GO - passable terrain
- SLOW GO - passable terrain with restrictions
- NO GO – impassable terrain

The impact of geographic factor can be evaluated as a *coefficient of deceleration 'C_i'* from the scale of 0 to 1. The coefficient of deceleration shows the real (simulated) speed of vehicle v_j in the landscape in the confrontation with the maximum speed of given vehicle v_{max}. The impact of the whole n geographic factors can be expressed as the formula:

$$v_j = v_{\max} \prod_{i=1}^{n} C_i, \qquad n = 1, \ldots, N. \qquad (4)$$

The main coefficients of deceleration are listed in the next table.

Table 2 Main coefficients of deceleration

Basic coefficient	Geographic signification and impact
C_1	Terrain relief (gradient of terrain relief and micro relief shapes)
C_2	Vegetation cover
C_3	Soils and soil cover
C_4	Weather and climate
C_5	Hydrology
C_6	Build-up area
C_7	Road network

Each coefficient consists of several coefficients of 2nd grade. For example C_2 is express after simplification as:

$$C_2 = C_{21} C_{22}$$

where:

- C_{21} is deceleration coefficient by impact of trunks spacing,
- C_{22} is deceleration coefficient by impact of trunks diameter.

The values of deceleration coefficients are counted for given vehicle (its technical properties) from ascertained properties of geographic objects stored in the spatial geodatabase. Using formula (4) it is possible to create a cost map in which the value of each pixel is the final (modelled) speed. The cost map can be as a source for the fastest path, reliable path etc. calculation.

4.2 Spatial database utility value evaluation

The master DGI database is usually utilised as a base for spatial data analyses. The national or international databases as DMU25, VMAP1or MGCP are very detailed, carefully maintained and used in many applications. But nobody can suppose that those databases contain all information he could need.

The task of CCM solution could require more information that is available in the master database. Geographer-analyst has to consider which information and in what quality can he obtain from master database. E.g. all forests in the area of interest are necessary to select for mentioned C_{2i} coefficients. Further he has to find out all their properties and their accuracy or count how many characteristics are missing. Next step is the individual functionality value of given part of master database evaluation.

Not all attributes are available within the used thematic spatial databases. So far the incompleteness of attributes has been omitted. Thus the real state-of-the-art has not been taken into account and the resulting CCM path has been considered as ‘certain’. One of the possibilities to make the resulting path closer to reality is to take the data attribute incompleteness into account and inform the decision maker (commander) about the uncertain parts of the path.

Two variants of the DMU25 database were utilised for the pilot project. The feature properties were defined according to the *Feature Attribute Coding Catalogue* (FACC) adapted as *Catalogue of the Topographic Objects* (CTO) (MTI 2005) in the first variant updated in 2005. The 4th edition of CTO was transformed in accordance with the *DGIWG Feature Data Dictionary* (DGIWG-500, 2010) in 2010 and transformed edition (updated in 2010) was used in the second variants (MoD-GeoS 2010).

The smaller personal database was created in the area of Brno of the size approximately 400 km^2 and all objects necessary for CCM evaluating was selected from DMU25 databases of both variants. The individual utility value was counted for both variants. On the base of statistical analyse 12.65% objects have any problems mainly incomplete attributes in the first variant of DMU25 while 3.45% objects have any similar problems in the second one. The time difference is 5 years between both variants. Hence the individual utility value was calculated by the use of the formula (3) as 0.6887 for the 2005 variant and 0.8825 for the 2010 variant. The ideal quality level is 1.0068. Both variants were used for CCM of TATRA 815 evaluation.

4.3 CCM of TATRA 815 analyses

The common army vehicle TATRA 815 ARMAX was chosen for a particular vehicle evaluation.

Table 3 The technical characteristics of TATRA 815 ARMAX (Tatra 2010)

Parametr	**Value**
Length (m)	7,87
With (m)	2,5
High (m)	3,01
Maximum climbing capability at 26000 kg of cargo	36°
Maximum climbing capability at 41000 kg of cargo	18°
Maximum climbing capability up to rigid step (m)	0,5
Maximum width of trench (m)	0,9
Maximum road speed (kmp)	85
Maximum depth of wade without water streaming (m)	1,2

The process was in progress according to next schema (**Figure 3**):

- CCM evaluation - only reliable information are considered,
- final cost map calculation according to equation (4),
- minimum cost path calculation from one initial point to three destinations placed in the forests as some hidden position.

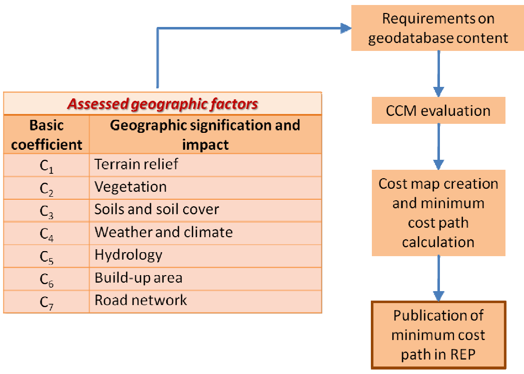

Figure 3 Spatial analyse without database quality evaluation

Only one solution is offered to commander. This solution seems to be appropriate, but geographer-analyst generally doesn't know details about situation on area of responsibility and commander's intentions. Problems should appear when any situation doesn't make the published path possible to use. Commander usually requires a new solution in such a case to miss prohibited area. The quality characteristics of temporally database are than to be considered by geographer to be sure, where are the weakest points of a new analysis. The weak points of analysis have to be sent to commander together with own analysis and it is up to him what will be the final decision. Two tasks for commander appear in CCM example:

1. Use less reliable path and consider that vehicles could stay in front of some obstacle
2. Wait and order to GEO team to improve spatial database (e.g. required properties) as soon as possible and then use new reliable path

The second case was simulated by the second variant of database in which the quality parameters were improved.

ArcGIS 9.3 was used for all calculations and analyses. In the next figures there are the main results – cost maps. The cost of each pixel is symbolized in the gray scale where darker tone signifies higher cost, higher speed in this case.

Figure 4 The cut cost map created from DMU25 2005 version

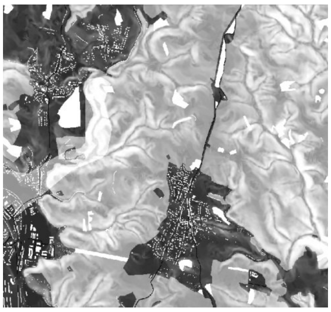

Figure 5 The cut cost map created from DMU25 2010 version

The minimum cost paths were evaluated using both cost maps and the same process created in ModelBuilder were applied. The results are in the next figures.

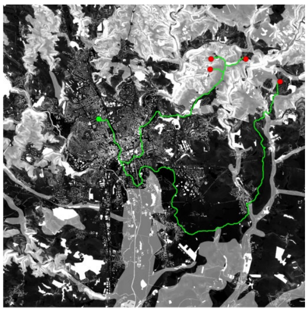

Figure 6 The minimum cost paths in CM of 2005 version. The initial point is green, the destinations are red.

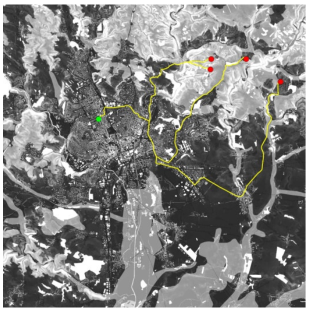

Figure 7 The minimum cost paths in CM of 2010 version. The initial point is green, the destinations are red.

The comparing of both results presented over the topographic situation is shown in the next picture (**Figure 8**).

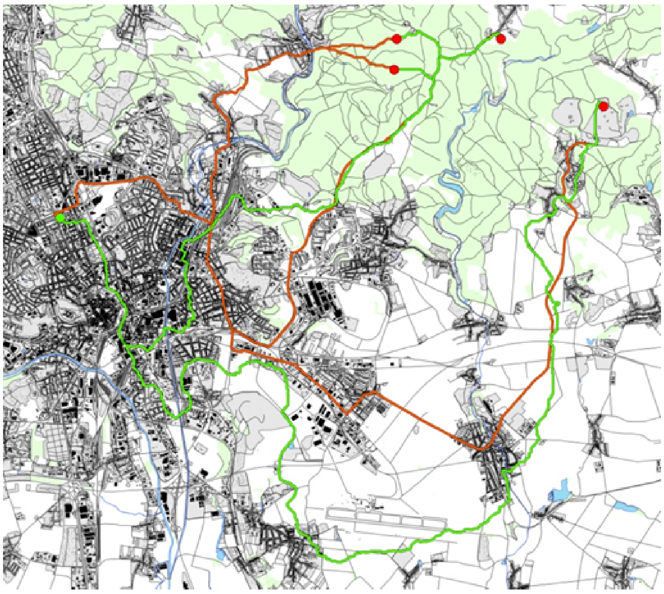

Figure 8 Comparing of two variant of minimum cost paths. Red ones answer to 2005 version and the green ones to 2010 version.

5. CONCLUSION

The pilot project has demonstrated a strong relationship between quality data and the results of spatial analysis. Likewise, it pointed to the problem of defining quality. It is not possible to assess only the technical properties of the spatial database, but it is necessary to consider the quality of the entire complex. In the pilot project, we have dealt only marginally with uncertainty in setting the boundaries of geographic objects and phenomena and with the uncertainty of their thematic properties. The problem of implementation of the principles of uncertainty and their mathematization will be the task of the solution of our future project.

6. ACKNOWLEDGEMENTS

Research results presented above were kindly supported by the project 'The evaluation of integrated digital spatial data reliability' founded by the Czech Science Foundation (Project code 205/09/1198) and also by the Project of Defence Research of the Ministry of Defence of the Czech Republic 'Geographic and meteorological factors on the theatre' (project code METEOR). Partial results are components of research task 'Dynamic geovisualization in crises management' supported by the Czech Ministry of Education (grant MSM0021622418).

BIBLIOGRAPHY

DGIWG-500. (2010). *Implementation Guide to the DGIWG Feature Data Dictionary (DFDD)* (2.2.2 - 19 July 2010. vyd.). DGIWG.

Miles, L. D. (1989). *Techniques Of Value Analysis Engeneering* (3rd ed.). USA: Eleanor Miles Walker.

MoD-GeoS. (2010). *Catalogue of the Topographic Objects DMU25* (7.3 ed.). Dobruska: Ministry of Defence of the Czech Republic, Geographic Service.

MTI. (2005). *Catalogue of the Topographic Objects DLM25* (4 ed.). Dobruska: Military Topographic Institute.

Rybansky, M., & Vala, M. (2010). RELIEF IMPACT ON TRANSPORT. *ICMT'09: INTERNATIONAL CONFERENCE ON MILITARY TECHNOLOGIES* (pp. 551-559). Brno: UNIV DEFENCE, BRNO.

Talhofer, V., Hošková, Š., Kratochvíl, V., & Hofmann, A. (2009). Geospatial Data Quality. *ICMT'09 - International conference on military technologies 2009* (pp. 570-578). Brno: Univerzita obrany.

Tatra, a. (2010). *Tatra is the solution*. Retrieved 5 17, 2010, from TATRA: http://partners.tatra.cz/exter_pr/vp/new/typovy_listprospekt.asp?kod=341&jazyk=CZ