ON THE ACCURACY OF THE GIS MAPPING BASE OF VLORA

Adelajda HALILI[1] and **Pal NIKOLLI**[2]

UDC: 528.4:004.6]:528.9(496.5)

SUMMARY

Maps in general, and those topographic in specific, serve as GIS data source, and are also its products. The maps, as tools destined to collect information with a certain purpose, should be trustworthy, not only for their thematic information, but also for their geometric accuracy with which each geographic object is identified.
Despite the care and the methods chosen, these maps contain several errors that influence the accuracy of the GIS database. These errors are multiplied due to data entry errors: during scanning, Geo-referencing, digitization etc. so, because of the conceptual and technical limitations, the geographic positioning attributes, described within spatial data contain errors and uncertainty.
This article develops the theoretical concepts on the accuracy of GIS database, and especially the mapping base of GIS, and assesses the geometric accuracy of the topographic map of Vlora at the scale of 1:50000, published by the Military Geographic Institute of the Albanian Ministry of Defense in cooperation with the US and NATO Defense Departments.

Key words: mapping, GIS, errors, qualitative attributes, accuracy, test, topology, accurate assessment.

1. INTRODUCTION

In order to have accurate interpretations and have the right decisions made, it is necessary that the data used should meet an acceptable level of quality. However, the spatial data are prone to a range of errors and uncertainties during data collection and processing.

[1] **Msc. Adelajda Halili**, adelahalili@yahoo.com; Universiteti Politeknik i Tiranës, Fakulteti i Gjeologjisë dhe Minierave, Departamenti i GjeoInformatikës, Tiranë, Shqipëri.
[2] **Prof.Dr. Pal Nikolli**, pal.nikolli@unitir.edu.al; Universiteti Tiranës, Fakulteti I Historisë dhe i Filologjisë, Departamenti i Gjeografisë, Tiranë, Shqipëri.

The quality of spatial data includes the accuracy, precision, reliability and validity, where each represents a special attribute of the database.
Until recently, the GIS usage and development experts were very careful to these errors, inaccuracies and lack of precision of spatial data. Of course, a general awareness had prevailed on the errors of all data available, with regard to inaccuracies and lack of precision, but the magnitude of the problem and its GIS solutions were not entirely on the focus. This situation has changed considerably during the last years. It is nowadays well-accepted that the error, inaccuracy and lack of precision can lead many GIS projects to fail.

2. THE SPATIAL DATA QUALITY AND ITS ATTRIBUTES

The data quality is an important factor in the process of conveying the information accurately and efficiently. It can vary from person to person, from organization to organization or application to application. It is the responsibility of the user to decide whether a data network complies with quality requirements or not; the standards varying depending on the application or use. A certain data quality could be satisfactory for one project but not necessarily for another.
The accuracy is the first consideration when data is assessed and analyzed. The accuracy, precision and the degree of error have their own meaning and importance. The degree of error is the difference between the real and estimated values. The accuracy, as relates to the error, is the degree on which an estimated value approximates a real one. The precision is the distribution of the estimated values as compared to the real ones. It is assessed based on the values of the standard deviation of all observations to the mean. The precision also refers to the ability for displaying values up to a certain value after the decimal point. The high precision or reliability does not equal high accuracy, and neither does high accuracy equal high reliability. Sometimes the two terms are used interchangeably in a wrong way.
The error is ubiquitous during each phase: on the conceptual one (perception), in the field (collection of false data due to the limitations of the theodolite, GPS satellites etc), and in measuring the attributes (due to changes in the environment, observer bias, poor data handling and manipulation, misinterpretation of the obtained data etc).
The spatial data transfer standard lists 5 components deemed adequate to assess the utility of the obtained data:

- data origin
- positioning accuracy
- attributive accuracy

- logical sustainability
- completeness

all these elements provide qualitative information in relation to the data network.

2.1 Geometric accuracy (or positioning or spatial accuracy)

The geometric accuracy determines the deviation from the values of actual positions between the GIS and the nominal terrain (Servigne, Lesage and Libourel, 2010). The nominal terrain is in principle an abstract of the real world. The position accuracy can be divided in absolute or relative accuracy. It can also be sub-divided into horizontal or vertical accuracy. The assessment methods are generally based on the assessment of the sources, as compared to standards of higher accuracy or in the empirical assessments. The accuracy variations can be reported as quality covers or of the additional attributes (Vereign, 1999).

One should be careful on the risks of false accuracy and precision, for all information that is received beyond the the level of accuracy and precision of the map. This is a potential mistake for computer systems that allow the user to move or magnify the map into larger scales. The accuracy and the precision is linked to the original scale of the map and do not change is the use magnifies the map. Zooming in or out however can be misleading to the user in leading them to believe that the accuracy and precision has increased.

The position and the accuracy, as they represent the trustworthiness, the belief and the danger, define the way that the geographic database can eventually be used. The spatial analysis required sustainable accuracy within each geographic strata (Ries 1993).

Spatial error is an important concern in GIS as the error might influence several attributes of the information saved in a database. The standards and the specifications are crucial for facilitating the development and exchange of the data and geospatial products.

2.2 Semantic accuracy (or the non-spatial attributes accuracy)

The accuracy attributes or thematic one is defined as the accuracy of data attributes. If, for example, in a rice field, is marked on a map as a wheat field, the result is a thematic or attributes error. The attributes tests or thematic ones can result from deductive assessments or with samples independent from the polygon coverage. The quantitative data could be collected by using the accuracy of the data collection tool. The attributive quantitative accuracy is missing, for example, when in a region with an average of rainfall of 100 cm the measurements are collected using an instrument that collects accurately at 0.1 cm, meaning the measurement error is 0.1 cm. For the qualitative or categorical errors, the resolution is determined in the aspect of the cleanliness of the categorical definition. For example, a pine tree region is marked as a maple tree region is an example of an attributive qualitative error. One must differentiate between the original and obtained attributes.

2.3 Completeness

The completeness describes the relationship between the objects in a group of data and the relationship of the same group of objects with the real world. The completeness of the data is useful in detecting commission errors (additional erroneous characteristics) and the inactivity of certain objects (lost characteristics). In a spatial model, the model completeness, expresses the adequacy of representation in relation to user requirements (Servigne, Lesage dhe Libourel, 2010). It can also include such information as selection criteria, definition and other mapping rules that are used to create the database etc, (Veregin, 1999).

2.4 Origins

The origins provides the necessary information for rebuilding the history of a database so that it can be analyzed eventually. The origin provide the information as follows: data source, including information on the organization that provides it; coordination system, projection systems, and related corrections; data collection methods, data computing or derivation; conversion methods during digitization or vectorization of raster data; transformations (i.e. transformation of coordinates, re-classification etc (Servigne, Lesage dhe Libourel 2010).
It is important to be aware of the data source as many of the errors result from the mistakes made in this phase. To validate the previous results it is necessary to create a group of control points on data collection or description. These should be documented at such detail to allow for the improvement of data accuracy. Also, all the transformation process should be properly documented so that confusion is avoided and so is the approximation error in later stages.

2.5 Logical sustainability

The logical sustainability aims to explain the accuracy or fidelity of relationship coded in the database structure in relation to the limitations imposed by the specifications of the inputed data (Servigne, Lesage dhe

Libourel, 2010). A data network is considered sustainable is it meets the structural attributes of the object it represents. For example, a contour must be closed, construction buildings should be closed polygons, roadway nodes should be connected etc. This includes the tests on the validity of attribute values, and the description of the topology mismatches based on graphic tests or topology specifications (Veregin , 1999).
Lack of logical sustainability or the mismatch stems also from the incorrect determination of the GIS process. All of the process, from the conceptual phase, to data collection, data analysis, data manipulation and interpretation, should be standardized and accompanied with a detailed description.

2.6 Temporal accuracy

The temporal accuracy relates to the temporal attributes of the data as the time of data collection, type and frequency of data collection, as well as the period for which the data are valid (Servigne, Lesage dhe Libourel, 2010). There are 3 temporal concepts: the actual time, that is the exact time when the event took place, i.e. the date of areal photography, time of observation and time of the transaction, time of data entry etc. The temporal accuracy is defined as the part of the data error that stems from the temporary nature of the data.

2.7 Semantic sustainability

It describes the importance of the meaning of the objects in relation to the model chosen; it describes the number of objects, relationships, and correctly coded attributes in compliance to the universe of rules and specifications (Servigne, Lesage dhe Libourel, 2010). The semantic sustainability is linked to the quality of the description of the geographic objects. For example, is someone has created a hospital database, should the clinics be included in this database? Solutions to such issues require profound geographic and consolidated reasoning.

2.8 Time-Frames

Time-frames provide information on how recent are the data, which consists of a starting and ending date (Servigne, Lesage dhe Libourel, 2010). Even though some studies considers this to be a parameter on its own, Harding (2010) suggests that usage can be considered as a form of the semantic accuracy, and that the completeness and the semantic sustainability could be included under the semantic sustainability.

Due to errors linked to the data during data collection or transformation, inaccuracies (deviations from the true value) in increased and leads to data quality deterioration. Even though the errors are always found in GIS data, special care should be taken during each of the steps, so that data errors are minimized and the data quality is preserved at the highest standard. To minimize data errors, we should analyze the origin and characteristics of data errors. There are 3 kinds of errors in GIS:

1. Source errors, which are present at data sources. They take place before GIS implementation and are linked to errors of instruments, human errors, changes of objects and events on which data is being collected on.
2. Data processing errors, that are errors that happen during GIS implementation. These are errors happening during data entry (digitization – human errors, the width of a line, turns, nodes, attribute entries; solitary nodes, pseudo-nodes, entry of erroneous projections etc); data manipulation (interpolation of point data in lines and areas; overlapping of layers digitized separately – i.e. land and vegetations; mixing the effects of processing and the analysis of multi layers, i.e. if each of the layers has an accuracy of 90% , the resulting layer accuracy will be 81% and the observations density), and
3. Errors linked to results generated (scale change – the graphical and numerical scale; color templates – using the wrong colors on monitor and in print etc).

In addition to the above, other errors linked to data transfer and conversion can take place. These errors can lead to data loss. The idea that GIS data are of high quality is erroneous. The quality of GIS information depends on the quality of the data used as input.

3. ASSESSING THE DATA QUALITY FOR THE TOPOGRAPHIC MAP (SCALE 1:50000) OF VLORA PREPARED IN COMPLIANCE WITH NATO STANDARDS

The basis for making decisions in spatial projecting and planning is the spatial data, and first of all the mapping data, such as cadastral, topographic and thematic maps, and later all statistical and other available documentation. During the last decade of the 20th century, the spatial data were adapted to the digital format, revolutionizing their use. The advances in the information technology, and the revolution of the GIS data collection technology (GPS, digital photogrametry, laser scanning etc), played a leading role in the management of spatial databases.

The quality of classic topographic maps includes the geographic accuracy as well as the thematic loyalty of the map content. The positioning accuracy of the geographic information is an important quantitative element of data quality, no matter whether they are in analog or digital form. For this data analysis we took under consideration the topographic map of Vlora (scale 1:50000, sheet 2975 I and 2975 IV) developed by Albanian Military Geographic Institute, in compliance with NATO standards.

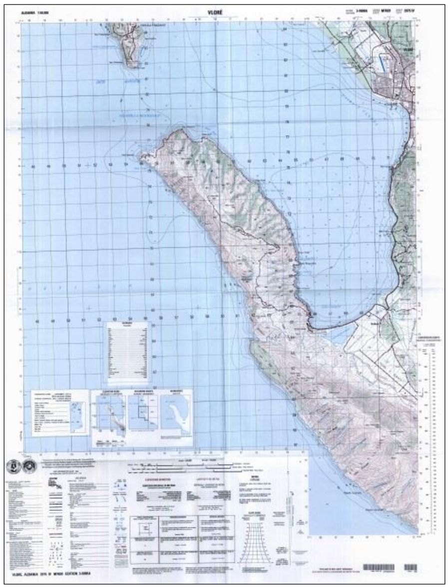

Fig. 1. The topographic map of Vlora (sheet 2975 IV)

The quality control for this map consists in establishing a series of GPS control points, clearly identified in the field and on the map (geodesic points, buildings, line intercepts, vertices etc). For these points, terrain coordinates were identified via GPS and on the map via GIS. Its is important to state that both types of coordinates were transformed on the same system, UTM evaluated for the Albanian territory. By calculating the differences in the two types of coordinates on the 'field' and 'map' we developed an index that was analyzed statistically.

The errors of digitization of current maps includes errors in the collection and processing of current data (field measurement and mapping errors) and transformation errors (scanning, geo-referencing and vectorization).

The accuracy of source maps used as input is defined by the Gauss mean quadratic error and the maximal acceptable error (twice the value of the mean quadratic error). The expression of the Gauss mean quadratic error is:

$$\sigma = \sqrt{\frac{1}{n}} \sum \Delta^2$$

where:
σ – standard deviation (Gauss mean quadratic error) of the data source;
Δ – deviation from the real values;
n – the number of observations

The statistical analysis of the above data shows that the standard deviation of the geodesic control points is σ = ± 3.2 m. The standard deviation of the detail points is less than ± 10 m.
Hydrography is better represented and more accurate as compared to other elements of the topographic maps.
The thematic accuracy includes includes the fidelity of the data of topographic maps. It is impossible to identify a mathematical expression to express thematic accuracy. In the environment displayed in the above mentioned maps there is a multitude of changes due to human intervention (buildings, trenches etc) and the influence of nature factors (landslides, erosion etc). In addition, essential differences are identified is specific methods of surveying and topography are applied (topographic and photogrametric surveys, basic height, generalization of types of terrain etc).
There are visible differences on land cover, with the most visible ones near human households. Populated areas and new infrastructure developed recently makes up most of these differences. New households are identified and old ones that have been changes their appearance. All these differences have an impact on the fidelity of old maps. Big differences are also notices on transportation network, dwelling and geographic names. The thematic accuracy for maps of 1:50000 scale, constructed in compliance to NATO standards (1994-1995), especially in their content, is higher than that of maps constructed through direct field measurements or through photogrametry. For all the other kinds of maps taken into consideration from other mapping sources, the quality is poorer due to the low accuracy of thematic content and neglect of mapping sources.
The conversion of data from analog to digital, involves other errors, but their magnitude does not pass the maximal permitted error for analog data sources. Therefore, the accuracy of digital data after processing may be estimated as a function of the maximal error of source data (m_{mb}) and the maximal error of data processing ($m_{përp}$) through this formula:

$$m^2_{max} = m^2_{mb} + m^2_{përp}$$

where: m_{max} – digital data error, m_{mb} = σ (standard deviation of source data)

If we accept that $m_{mb} = m_{përp}$, then:

$m^2_{max} = m^2_{mb} + m^2_{përp} = 2\ m^2_{mb}$, while:

$m_{max} = \pm\ 1.4\ m_{mb}$

Based on the expressions above, the mean quadtratic error was calculatedand the quality of digital data for the geodesic base points was estimated: $m_{max} = 1.4\ m_{mb} = \pm\ 1.4 * 3.2\ m = \pm\ 4.48\ m$.
The accuracy of the mathematical base (map projection, reference points and geodesic points) of topographic digital maps is listed in the following table:

Base mathematical elements	*Mean quadratic error at map scale*
Map projection (lines)	± 0.10
Reference points (marked)	± 0.10
Geodesic points	± 0.18

Tab. 1

Forest areas and/or mountains or sand hills, objects are 1.5 times less frequent with a maximal error of 0.75 mm. Where the symbols of topographic maps are not deviated, the absolute value of horizontal line errors identified strictly in relation to the nearest control point or network does not exceed 0.5 mm.
The errors due to deforming of map paper by scanning at large can be minimized, but only under certain conditions.

CONCLUSIONS

Being aware of the quality and accuracy of data, especially those in mapping, is one of the key aspects in GIS implementation. Preliminary assessment of topographic maps of Vlora at scale 1:50000 is based on the analysis of the accuracy of its plan and content. The mean quadratic error of the geodesic control points is ± 3.2 m. Positioning accuracy of the other points of the topographic content is within the map graphic accuracy. The mean quadratic error of detail points is less then ± 10 m.
The accuracy of digital data for the geodesic points obtained from topographic maps mentioned above is around ± 4.48 m, which speaks in favor of a very good quality of the spatial data of such maps.
The initial thematic data at the time maps were constructed were very reliable. However, the the quality of the general information has deteriorated

through time, and it is necessary to update the content of these thematic maps periodically.

REFERENCES

1. Harding, J. 2010. Vector Data Quality: A Data Provider's Perspective in Fundamentals of spatial data quality edited by Devillers R. and Jeansoulin R. London, UK: Wiley.
2. M.F. Goodchild (1995) Attribute accuracy. In Elements of Spatial Data Quality, edited by S.C. Guptill and J.L. Morrison. New York: Elsevier
3. Menezes, L.P.M., Miscel, B.M., Godoy, V.F., 2011, Quality Evaluation Criterium of Historical Maps, CO-363, ICC, Germania, Dresden.
4. Ries, T., "Design Requirements for Location as a Foundation for Transportation Information Systems," GIS-T '93 Geographic Information Systems for Transportation Symposium, Albuquerque, New Mexico, pp.48, March 29-31, 1993.
5. Servigne S., Lesage, N., Libourel, T. 2010. Quality components, standards, and metadata in Fundamentals of spatial data quality edited by Devillers R. and Jeansoulin R. London, UK: Wiley.
6. Veregin, H. 1999. Data quality parameters. In Geographical information systems, 2nd Ed edited by Longley P.A.,. Goodchild M. F, Maguire D.J., and Rhind D.Ë., eds., Vol. 1, Chapter 12. New York, NY: Wiley.