

AUTOMATIC CHECKING OF FEATURE AND ATTRIBUTE CONSISTENCY OF A SPATIAL DATABASE

Dr.-Ing. Hayati TAŞTAN * , Prof.Dr. M.Orhan ALTAN **

* General Command of Mapping, Informations Systems Department, Cebeci TR-06100, Ankara, TURKEY
htastan@hgk.mil.tr

** Istanbul Technical University, Civil Engineering Faculty, Photogrammetry and Remote Sensing Division,
Maslak TR-80626 Istanbul, TURKEY
oaltan@srv.ins.itu.edu.tr

KEYWORDS: GIS, Spatial Data Quality, Quality Control, Consistency Check.

ABSTRACT

Since the quality of decisions made on geo-related issues depends on the quality of the spatial data upon which they are based, software tools are needed to automatically check the spatial data quality. Although positional & thematic & temporal accuracy checks and feature & attribute & spatial completeness checks need human interaction in some extent, format & topological & feature / attribute consistency checks are likely to be carried out automatically. Format and topology consistency checks depends on the data structure of the GIS software used to create the spatial database, whereas feature & attribute checks might be done by referencing any data dictionary for any spatial database. This study explains spatial data quality standards with the focus on the spatial data quality evaluation model and presents its application in a visual programming environment (VBASIC) to check feature & attribute consistency of a spatial database.

1. INTRODUCTION

The main goal of any Geographic Information System (GIS) is to produce information in support of decision making on geo-related issues (Taştan, 1991). The quality of such decisions is influenced by the quality of spatial data upon which they are based. In other words bad information produced from spatial data of poor quality may cause wrong decisions, while spatial data of good quality reduces the risk that wrong decisions will occur. Spatial data quality is may be defined as totality of indicators showing affirmative or negative distinguishing properties and characteristics of geographic data which describe confirmation to the specification and fitness for use, from the producers and users point of view, respectively. From this definition, spatial data quality seems to be subjective measure, which has been interpreted differently by users and producers. For instance, a set of spatial data may meet the specifications of a producer and therefore may be evaluated as being of good quality, whereas it might be of poor quality by not meeting the expectations of an other producer. Also, it is the same case for users as fitting or not fitting for their application fields (Taştan and Altan, 1999). Examples for quality evaluation may be found in various literatures, such as Fitzgerald et.al., 1996, Fatale et.al., 1996, Taştan, 1997, Veregin, 1998, Honkavaara et.al., 1998, Jakobsson, 1999 and Taştan, 1999.

There are three factors determining the spatial data quality, which are money (cost for producers, price for users), time (production period for producers, users time to access to data) and acceptable quality level (production goals of producers, application fields of users). These factors are not one fold. Producers decrease the quality in order to produce with low cost and in a short time period whereas users demands data with low price and in a short response time while accepting data with low quality. If users increase the quality they expect, then they are supposed to give much money and wait to get data. As for the acceptable quality level, it may be different for producers and users, according to the production goals of producers and application fields of the users as stated above.

2. SPATIAL DATA QUALITY STANDARDS

Standards on spatial data quality consist of a standard for a terminology on spatial data quality, a standard for a quality model comprising quality parameters & measures, a standard for acceptable (conformance) quality levels and a standard for a spatial data quality evaluation model.

2.1 Standard for a Terminology on Spatial Data Quality

This standard contains all the terms related with spatial data quality. A terminology for the spatial data quality is given in ISO 15046 Geographic Information Part 4 (ISO/TC 21, 1998a).

2.2 Standard for a Spatial Data Quality Model (Parameters & Measures)

Several spatial data quality models are described in various national (FGDC, 1995), regional (NATO, 1979; CEN/TC 287,1996) and international (ISO/TC 211, 1998b; ISO/TC 211; 1998c) standards on geographic information which are the same as those explained below to some extent. There are nine parameters defining the quality of the spatial data, which are data quality elements (positional accuracy, thematic accuracy, temporal accuracy, completeness, logical consistency) and data quality overview elements (currency, lineage, purpose, usage) (ISO/TC 211, 1998a; Aranof; 1991; Taştan, 1999).

2.2.1 Positional Accuracy: Accuracy is the degree of conformity with the standard. Positional accuracy is the expected deviance in the geographic location of a feature in the spatial data set from its true ground position. Positional accuracy as a data quality element consists of two data quality sub- elements, namely horizontal accuracy and vertical accuracy. Horizontal accuracy is the uncertainty in the horizontal position of a point with respect to the horizontal datum required by a product specification caused by random and any systematic errors. Quality measures for horizontal accuracy are Circular Standard Deviation (σ_c), Circular Probable Error (CPE), Circular Map Accuracy Standard (CMAS) and Circular Near Certainty Error (CNCE). Vertical accuracy is the uncertainty in the vertical position of a point with respect to the vertical datum required by a product specification caused by random and any systematic errors. Quality measures for vertical accuracy are Linear (Vertical) Standard Deviation (σ_v), Linear Probable Error (LPE), Linear (Vertical) Map Accuracy Standard (LMAS,VMAS) and Linear Near Certainty Error (LNCE). Positional accuracy of line and area features is calculated from the positional accuracy of the points that form the line or area. Extra information on positional accuracy may be found in Drummond, 1997 as well as in Taştan & Altan, 1999.

2.2.2 Thematic Accuracy: This data quality element has two sub-elements, which are classification (feature and attribute) correctness and quantitative attribute accuracy. Classification correctness is the probability level of feature classification in the data set. For example 90 % classification correctness for bridges means that 90 out of 100 bridges in the data set are bridges in the real world whereas the rest might be classified as any feature except bridge. The quality measure for classification correctness is the percentage of correctly classified (PCC). Classification correctness for an attribute describes the accuracy of data collection for an attribute. For example, "correctness for the attribute pavement of the road is 90%" means that 90 out of 100 bridges have been correctly attributed for their pavement. Quantitative attribute accuracy is the accuracy of a quantitative attribute. The quality measure for quantitative attribute accuracy is pass/fail or probability level in percentage, RMSE or standard deviation. Quantitative attributes might be height of bridge, width of road, etc (Tan and Altan, 1999).

2.2.3 Temporal Accuracy: The sub-elements of this data quality element are temporal accuracy, temporal consistency and temporal validity. The quality measures for this sub-elements are pass/fail (Boolean), error statistic (RMSE, probability level in percentage) and degree of inconsistency for temporal consistency and validity (number, percentage). Temporal accuracy is correctness of the temporal references of an item (reporting of error in time measurement). For example, accuracy of recorded information vs. actual information, accuracy of date of last change, accuracy of date of last review, and accuracy of period of review. A specific example of accuracy of recorded information vs. actual information is if the collection time is reported as occurring at 10:00 a.m. on a certain date when, in fact, the collection time actually occurred at 11.03 a.m. Temporal consistency is correctness of ordered events or sequences, if reported. For example, if the birth date of a feature proceeds its deletion date, then there is a temporal inconsistency. Temporal validity is the validity of data with respect to time. For example, temporal validity is false for a data set of 1994 census data containing data with collection dates of 1998.

2.2.4 Completeness: Completeness defines the relationship between the features in digital environment and features in the model world. There three sub-element of the data quality completeness are feature completeness, attribute completeness and spatial completeness. The quality measures for this data quality element are omission and commission. Commission shows excess data present in a data set, omission shows data absent from a data set. Commission and omission are expressed as the difference between the number of items in the model world and the number of items in the data set. Data quality value may be either a number or a ratio with a numerator equaling the number of encoded but non-existent items (for commission) or missing items (for omission) in the model world and a denominator equaling the total number of item in the model world. The item here might be feature instances, attribute value instances and spatial coverage. The ratio of the missing feature instances in the data set (i.e. digitized) to the number of feature instances in the model world (i.e. to be digitized) gives the omission in percentage for feature completeness. The same case is valid for attribute value instances, that is the ratio of the missing attribute value instances, which are not populated to the number of attribute values that require to be populated. For example, 5 % omission for the feature bridge means that 95 out of 100 bridges on the map (model world) are digitized (the rest might be missing). And 20 % omission for the attribute value bridge type means that the attribute bridge type is populated for 80 out of 100 feature instances while all the bridge types are defined on the map (in the model world). Omission for spatial completeness shows the percentage of the area, which is not digitized with respect to the whole coverage of the area in the model world that should be digitized. The same examples may be given for the

sub-element commission using the term *excess* instead of *missing*. (Commission for spatial completeness shows the percentage of the area, which shouldn't have been digitized ?!)

2.2.5 Logical Consistency: There are three sub-elements of the quality element logical consistency which are domain (data dictionary) consistency, format consistency and topological consistency. Domain consistency is the percentages of the features, attributes and attribute values in the data set with respect to those in the data dictionary's feature, attribute and attribute value domain sets which are called as feature consistency, attribute consistency and attribute value consistency, respectively. Format consistency indicates if the format of the product specification was followed. Topological consistency, the percentage of the topological rules followed that are defined by the data structure, such as *no overshoots and undershoots*, *all polygons should be closed*, etc. and user defined topological rules, such as *all bridges should be at the intersections of the roads and rivers*, *each road should end up with a residential area or point (for network analysis)*. The measure for the quality element consistency is pass/fail or degree of inconsistency. Data quality value for the measure *Pass/fail* is Boolean and data quality value for the measure *degree of inconsistency* is number or ratio (Kainz 1997; Taştan and Altan, 1999).

2.2.6 Currency: The currency represents the date at which the data was introduced to or modified in the data set. This parameter has not a measure like others for its not tested but used as metadata with quality indicators. In other words, old data may be of poor quality, since a number of changes in data set might have not been done, or of good quality for users who are seeking historical data for the production date of data set depending on the application field of the users.

2.2.7 Lineage: The lineage of a data set is its history. The lineage information consists of information on the source data, such as the name, scale, producer, edition number, edition date, accuracy, control points and information on the processing steps, such as date of processing, type of processing, operator, hardware & software used, algorithms, etc. Knowledge of lineage can be an important consideration in choosing a data set for a specific application (Clarke and Clark, 1997).

2.2.8 Purpose: Purpose records the purpose for creating a data set and contains information about a data set's intended use.

2.2.9 Usage: Usage documents the application(s) for which a data set has been used. Usage records uses of the data set by the data producer or by data users.

2.3 Standard for Acceptable Quality Levels for Spatial Data

The components of this standard are Acceptable Quality Levels (AQLs) for all spatial data quality parameters. Although there are AQLs for horizontal and vertical accuracy (U.S.BoB, 1947; NATO, 1979; ASPRS, 1985) there isn't any for the other quality parameters. AQLs for horizontal and vertical accuracy are described in United States National Map Accuracy Standards (U.S.BoB, 1947) are also accepted by NATO (NATO, 1979) and by ASPRS with some modifications later (ASPRS, 1985). This standard is as follow:

2.3.1 Horizontal accuracy : For maps on publication scales larger than 1 : 20 000, not more than 10 percent of the points tested shall be in error by more than 1/30 inch (0.508 mm), measured on the publication scale; for maps on publication scales of 1 : 20 000 or smaller, 1/50 inch (0.847 mm). These limits of accuracy shall apply in all cases to positions of well-defined points only. Well-defined points are those that are easily visible or recoverable on the ground, such as the following: monuments or markers, such as benchmarks, property boundary monuments; intersections of roads, railroads, etc.; corners of large buildings, or structures, (or center points of small buildings); etc. In general what is well defined will also be determined by what is plottable on the scale of the map within 1/100 inch. Thus while the intersection of two road or property lines meeting at the right angles would come within a sensible interpretation, identification of the intersection of such lines meeting at an acute angle would obviously not be practicable within 1/100 inch. Similarly, features not identifiable upon the ground within close limits are not to be considered as test points within the limits quoted, even though their positions may be scaled closely upon the map. In this class would come timberlines, soil boundaries, etc.

2.3.2 Vertical accuracy: Vertical Accuracy, as applied to contour maps on all publication scales, shall be such that not more than 10 percent of the elevations tested shall be in error more than one-half the contour interval. In checking elevations taken from the map, the apparent vertical error may be decreased by assuming a horizontal displacement within the permissible horizontal error for a map of that scale. The accuracy of any map may be tested by comparing the positions of points whose locations or elevations are shown upon it with corresponding positions as determined by surveys of a higher accuracy. Tests shall be made by the producing agency, which shall also determine which of its maps are to be tested, and the extent of such testing. Published maps meeting these accuracy requirements shall note this fact on their legends, as follows : This map complies with National Map Accuracy Standards. Published maps whose errors exceed those aforesaid shall omit from their legends all mention of standard accuracy. Having this AQLs for positional accuracy, some AQLs for definite scales for horizontal accuracy and vertical accuracy, are listed in Taştan and Altan, 1999.

2.4 Standard for a Spatial Data Quality Evaluation Model

The purpose of this standard is to express a consistent and standard manner to determine and report a data set quality information against product specifications or user requirements, for producers or users, respectively. This standard consists of the operations given as following:

2.4.1 Selecting Data Quality Element: Using the product specification, the data producer identifies the data quality element or data quality elements, which have to be evaluated to determine conformance to the specification. Using the user requirements, the data user selects the data quality element or data quality elements that have to be examined to determine how well the data set meets user requirements. Data quality elements are (as stated before) completeness, logical consistency, positional accuracy, temporal accuracy, thematic accuracy.

2.4.2 Selecting Data Quality Sub-element: Using the product specification, the data producer identifies the data quality sub-element or data quality sub-elements, which have to be evaluated to test for conformance to the specification. Using the user requirements, the data user selects the data quality element sub-element to be evaluated in the test for how well the data set's quality meets the user requirements. Data quality sub-elements are (as stated before) horizontal accuracy, vertical accuracy for positional accuracy; classification (feature & attribute) accuracy and quantitative attribute accuracy for thematic accuracy; temporal accuracy, temporal consistency and temporal validity for temporal accuracy, feature completeness, attribute completeness and spatial completeness for completeness; domain, format and topological consistency for logical consistency.

2.4.3 Selecting Data Quality Scope: The data producer shall determine the scope of the quality evaluation on each sub-element that is necessary to test for conformance to the specification and meet the intended product purpose. Each data quality sub-element may have a different scope or multiple scopes depending upon the product specification. Using the user requirements, the data user defines the scope of the quality evaluation necessary to test for how well the data set's quality meets the user requirements. One data quality scope is provided for each applicable data quality sub-element. A data quality scope can be a data set series to which a data set belongs, the data set, or an identified reporting group. The product specification and data quality overview elements are used to determine a data quality scope for each applicable data quality sub-element. If a data quality scope cannot be identified, the data quality scope shall be the data set. Quality can vary within a data set. Multiple data quality scopes may be provided for each applicable data quality sub-element to more completely report quality information. A data quality scope is adequately identified. The following can be used to identify a data quality scope: the level, such as a data set series to which a data set belongs, the data set, or a reporting group; the types of items (lists of feature types and feature attributes) or specific items (lists of feature instances and attribute values); the geographic extent; the temporal extent, including the time frame of reference and accuracy of the time frame.

2.4.4 Selecting Data Quality Measure: One data quality measure is provided for each data quality scope. A single data quality measure might be insufficient for fully evaluating a data quality scope and providing a measure of quality for all possible utilization of a data set. A combination of data quality measures can give useful information. Multiple data quality measures may be provided for a data quality scope.

2.4.5 Choosing Data Quality Evaluation Method: For each data quality measure selected, the producer or the user chooses the quality evaluation method to be used. There are two quality evaluation methods, direct and indirect. The direct method is accomplished by sampling or full inspection of the data in the data set; the indirect method is accomplished by evaluation of data quality information from sources other than the data in the data set. Choice of quality evaluation method is determined by the data producer and the data user.

Direct evaluation methods: There are two types of direct evaluation methods, full inspection evaluation method and sampling evaluation method. Full inspection evaluation method involves testing 100 percent of the items in a population to determine a quality result. Sampling evaluation method involves testing only a sample of all the items in a population to determine a quality result. Inspection by sampling and full inspection evaluation methods may be accomplished by either automated or non-automated means. For example some types of errors, such as feature consistency, attribute consistency and attribute value consistency can be detected by a computer program automatically (Taştan, 1999). For inspection by sampling, ISO 2859 (Inspection by attributes) and ISO 3951 (Inspection by variables) can be utilized. Steps for inspection by sampling are:

* Items are defined. An item is that which can be individually described or considered. An item may be any part of a data set, such as a feature, feature attribute, or combination of these.

* The data quality scope is divided into statistical homogeneous lots. Homogeneity may be evaluated based on the source data of production, production system (hardware, software, skill of operator), complexity and density of features. A lot is the minimum unit to which the result of quality evaluation is attached. If the lot does not pass inspection, the all items in the lot may be discarded or reproduced. In this sense, the definition of a lot is strongly related with the production process itself. For example, for a 1:25 topographic database populated from cartographic sources, a lot can be the coverage of a 1:250 K map sheet (i.e. 96 sheets per lot).

* Lots are divided into sampling units. Sampling unit area is a minimum geographical area in the model world for which the inspection is conducted for all items belonging to the geographical unit. Number of items in a lot is considered as lot size. For the example above, the coverage of a 1:25 K map sheet can be a sampling unit. By simple random sampling for inspection, sampling units are selected from each lot. The number of these units should be taken from the ISO 8422 (Sequential sampling plans for inspection by attributes) and ISO 8423 (Sequential sampling plans for inspection by variables). For the example above, 10 out of 96 map sheets may be selected.

* All items, which belong to the selected sampling units, are inspected. If the number of non-conforming items reaches a specified rejection number determined by AQL or LQ, the lot is not accepted. Since the inspection is by attribute, ISO 2859 applies. If the average and variance of inspected values do not satisfy limiting conditions determined by AQL, the lot is not accepted. Since the inspection is by variable, ISO 3951 applies. If all the lots are accepted, the data quality scope is accepted.

Indirect evaluation methods : Indirect evaluation methods are based on estimates of data quality measure values from sources other than the data items of the data set. The variety of sources includes, but is not limited to, metadata, knowledge of the data set's purpose, data set lineage documentation, history of uses made of the data set, and quality reports on the data used to produce the data set. Knowledge of the production process and errors that may have been introduced or detected during production is useful.

2.4.6 Specifying Conformance Quality Level: Producer or user specifies the conformance quality level for each data quality measure such that it establishes conformance or nonconformance to the product specification or user requirement.

2.4.7 Determining Data Quality Values: Data quality values are computed by applying the chosen quality evaluation method to the related sub-element scope.

2.4.8 Assessing Conformance to Product Specification: Data quality values are compared to the specified conformance quality levels for each selected data quality sub-element. Assessment results are either conforming or non-conforming (i.e. acceptable or non-acceptable).

2.4.9 Aggregating Quality Evaluation Results: The quality of a data set may be expressed by an aggregated quality result. This may require combining quality results from data quality evaluations based on differing data quality elements or data quality sub-elements, each result with perhaps different meaning than the others. The purpose of such an aggregation, even of dissimilar data quality results, is to provide a single measure of data set quality. A data set may be deemed to be of an acceptable aggregate quality even though one or more data quality elements or data quality sub-elements fail acceptance. The aggregate data set quality (ADQ) may be evaluated by several techniques (ISO TC/211,1999):

100% pass/fail method for Aggregate Data Quality (ADQ):

$$ADQ = v_1 * v_2 * v_3 * \dots * v_n$$

where v_i is the data quality evaluation result of each data quality sub-element scope (pass:1, fail :0) and n is the number of quality sub-element measured. If $ADQ = 1$, then the overall data set quality is deemed to be fully conforming. If $ADQ = 0$, then it is deemed non-conforming.

Weighted pass/fail method for ADQ :

$$ADQ = v_1 * w_1 + v_2 * w_2 + v_3 * w_3 + \dots + v_n * w_n$$

Where w_i is the weight (0.0-1.0) of the sub-element, which is based on the significance to the purpose of the product. The technique does provide a magnitude value indicating how close a data set is to full conformance as measured.

Minimum/Maximum Value method for ADQ:

$$ADQ = MAX(v_i, in = 1 \dots n)$$

$$ADQ = MIN(v_i, in = 1 \dots n)$$

The technique does provides a magnitude value indicating how close a data set is to full conformance as measured, but only in terms of the sub-element represented by the maximum or minimum.

2.4.10 Reporting Quality Evaluation Results: One data quality result is provided for each data quality measure. The data quality result shall be either the value or set of values obtained from applying a data quality measure to a data quality scope, or the outcome of evaluating the value or set of values obtained from applying a data quality measure to a data quality scope against a specified acceptable quality level. This type of data quality result is referred to in this part of the International Standard as pass-fail. A data quality result contains the following quality indicators: data quality scope, data quality measure, data quality evaluation procedure, data quality result, data quality value type and data quality date. These results are reported as metadata or additional data quality report. Identifiable data quality overview elements are also stated in the quality reports.

3. AUTOMATIC CHECKING OF FEATURE AND ATTRIBUTE CONSISTENCY OF A SPATIAL DATABASE

3.1 Base GIS Software Developed for Spatial Data Quality Check

In order to check the feature and attribute consistency, a base GIS software package, namely CBKK (Coğrafi Bilgi Kalite Kontrolü : Spatial Data Quality Check), has been developed via Visual Basic 6.0 (Enterprise Edition) and MapObjects2 (Evaluation Version). This software (Figure 1) has the basic GIS functions as shown in Table 1 and Table 2.

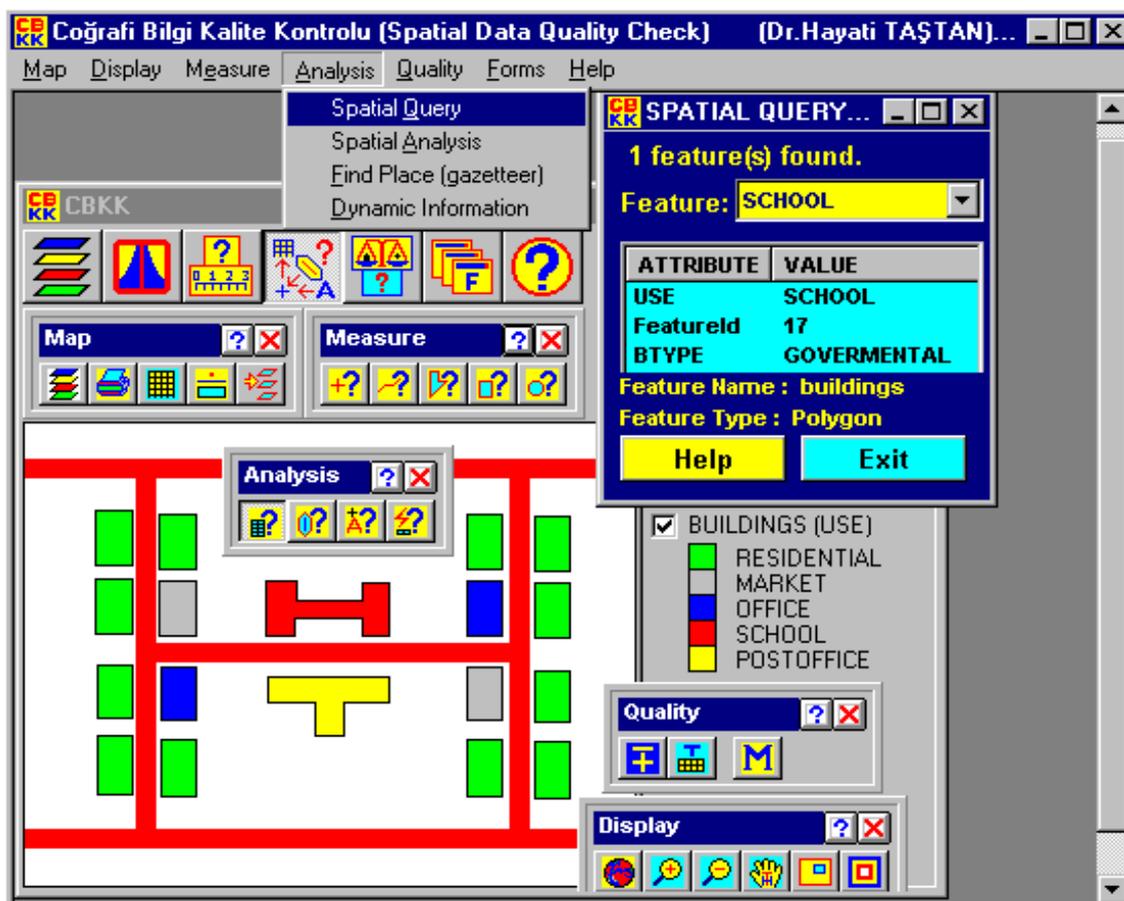


Figure 1. CBKK (Coğrafi Bilgi Kalite Kontrolü: Spatial Data Quality Check) Software

| Menu | Icon | Sub Menu | Icon | Function |
|---------|------|----------------------------|------|--|
| Map | | Layers | | Displaying and Symbolizing Raster & Vector Data |
| | | Output | | Printing and Saving Screen Displays |
| | | Edit Layer | | Editing Vector Data Attribute Values |
| | | New Layer | | Editing Vector Data Graphics in ESRI-Shape Format |
| | | Save New layer | | Saving Vector Data in ESRI-Shape Format |
| Display | | All | | Display features with Full Extent |
| | | Zoom in | | Zoom-in |
| | | Zoom Out | | Zoom-out |
| | | Pan | | Pan |
| | | Overview | | Creating and Overview Window |
| | | Properties | | Getting and setting the properties of the Display. |
| Measure | | Position | | Measuring the Position of the Cursor in Map Units |
| | | Distance | | Measuring the Distance in Map Units |
| | | Area/Perimeter (Polygon) | | Measuring the Area and Perimeter of a Polygon |
| | | Area/Perimeter (Rectangle) | | Measuring the Area and Perimeter of a Rectangle |
| | | Area/Perimeter (Circle) | | Measuring the Area and Perimeter of a Circle |

Table 1. Functions of the CBKK Software.

| Menu | Icon | Sub Menu | Icon | Function |
|----------|---|-------------------------------|---|--|
| Analysis |  | Spatial Query |  | Getting Descriptive Information of a Feature Selected |
| | | Spatial Analysis |  | Selecting a Feature(s) via Spatial Analysis |
| | | Find Place (Gazetteer) |  | Finding a Place (s) using Gazetteer Information |
| | | Dynamic Information |  | Getting Information of a Feature When Moving Over |
| Quality |  | Positional Accuracy |  | Determining positional Accuracy of a Feature Class in a a Vector Map using a Raster Map As a Reference. |
| | | Feature-Attribute Consistency |  | Determining Feature and Attribute Consistency of a Feature Class using a Predefined Data Dictionary A a Reference. |
| | | Meta Database |  | Defining a Data Dictionary as A Meta Database. |

Table 2. Functions of the CBKK Software (cont.)

3.2 Feature and Attribute Consistency Check via CBKK Software

In order check the feature and attribute consistency, the first thing to do is to create a Meta Database, using the built-in function of the CBKK software without any need to any other Data Base Management System (DBMS), such as Oracle, Access, Dbase, etc.

After the Meta Database created, a FEATURES table is created with the attributes FEATURE_NAME and FEATURE_TYPE and populated with the features names (such as buildings, roads, etc.) and feature types (such as point, line, polygon). Then an ATTRIBUTES table is created with the attributes FEATURE_NAME, ATTRIBUTE_NAME and ATTRIBUTE_TYPE and populated with the feature names (buildings, roads, etc), attribute names (such as BTYPE, RTYPE,USE, PAVEMENT, etc.) and attribute types (such as STRING, INTEGER, DATE, etc.). Then a ATTRIBUTE_DOMAINS table is created with the attributes ATTRIBUTE_NAME, VALID_VALUE and STATUS and populated with the attributes (such as BTYPE, RTYPE, etc.), valid values for each attribute (such as "GOVERNMENTAL" and "PRIVATE" for BTYPE, "PRIMARY" and "SECONDARY" for the RTYPE, and "REQUIRED" if the attribute value is required, "" if not required, etc.).

Having the data dictionary built, one displays the spatial data base (i.e. Vector Data Layers) and performs Feature and Attribute Consistency Check automatically using the related sub-menu under the menu "Quality". After the check, three tables are created in the same meta database. The INCONSISTENT_FEATURES table lists all features which are missing or abundant in the test data. The INCONSISTENT_ATTRIBUTES table lists all attributes which are missing or abundant in the test data. The INCONSISTENT_ATTRIBUTE_VALUES table lists all attribute values which are not in the domain list and also attribute values which are required but not populated. Reports from these tables can also been generated automatically and printed within this software.

4. CONCLUSIONS

Although spatial data quality check is valuable, its cost must be weighted against the benefits of quality information. Therefore, the expenditure of the quality evaluation must be matched to the consequences of errors. Also the willingness of the users to pay for the spatial data quality evaluation should be taken into consideration to justify for the level of evaluation.

Feature and attribute consistency check is the primary check of a spatial data base, since a spatial data base is a model of the real world and this check defines the quality of fitness of the spatial data to this model. Although this type of quality check can be performed programmatically, the other types of spatial data quality check are relatively complex and need more statistical procedures and hard study. Thus, international activities on spatial data quality management carried out by different bodies, such as ISO/TC 211, CEN/TC 287, OpenGIS Consortium, NATO, OEEPE-WG on Spatial Data Quality Management, ICA-WG on Data Quality and ISPRS works on this topic should be integrated in a cooperative manner and consolidated to one international standard.

REFERENCES

- Aranof, S., 1991. Geographic Information Systems: A Management Perspective, WDL Publications, Ottawa.
- Clarke, D.G., Clark, D.M., 1997. Lineage. In: Elements of Spatial Data Quality (Eds: Stephan C.Guptil, Joel.L.Morrison), Elsevier Science Ltd, ISBN 0 08 0424325, pp.13-30.
- Drummond, J., 1997. Positional Accuracy. In: Elements of Spatial Data Quality (Eds:Stephan C.Guptil, Joel.L.Morrison), Elsevier Science Ltd, ISBN 0 08 0424325, pp.31-58.

Fatale, L.A., Messmore, J.A., Ackeret, J.R., 1996. An Assessment of the Horizontal Accuracy of Interim Terrain Data. In: Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium (Eds: H. Todd Mowrer, Raymond L. Czaplewski, R.H. Hamre), May 21-23, Fort Collins, Colorado, pp. 301-308.

FGDC, 1995. Content Standards for Digital Geospatial Metadata Workbook, Federal Geographic Data Committee (FGDC), USA.

Fitzgerald, R.W., Ritman, K.T., Lewis, A., 1996. Attribute and Positional Accuracy Assessment of the Murray Darling Basin Project, Australia. In: Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium (Eds: H. Todd Mowrer, Raymond L. Czaplewski, R.H. Hamre), May 21-23, Fort Collins, Colorado, pp. 317-326.

Honkavaara, E., Kaartinen, H., Kuittinen, R., Huttunen, A., Jakkola, J., 1998. The Quality Control in the Finnish Land Parcel Identification System Orthophoto Production, In: International Archives of Photogrammetry and Remote Sensing, ISPRS Commission IV Symposium "GIS- Between Vision and Applications" (Eds: D. Fritsch, M. English, M. Sester), Stuttgart, Germany, p.252-259.

ISO/TC 211, 1998a. ISO 15046 : Geographic Information Part 4 : Terminology, ISO/TC 211 Doc.No.478.

ISO/TC 211, 1998b. ISO 15046 : Geographic Information Part 13 : Quality principles, ISO/TC 211 Doc.No.478.

ISO/TC 211, 1998c. ISO 15046 : Geographic Information Part 13 : Metadata, ISO/TC 211 Doc.No.538.

ISO/TC 211, 1999. ISO 15046 : Geographic Information Part 14 : Quality evaluation procedures, ISO/TC 211 Doc.No.665.

ISO/TC 211, 1999. Report of the ISO/TC 211 given at the 8th meeting of the ISO/TC 211, March-4, Austria.

Jacobsson, A., 1999. Experiences in Quality Management and Quality Control in Topographic Data Production at National Land Survey of Finland, In: 19th International Cartographic Conference Proceedings (CD-ROM), Ottawa.

Kainz, W., 1997. Logical Consistency. In: Elements of Spatial Data Quality (Eds: Stephan C. Guptil, oel.L. Morrison), Elsevier Science Ltd, ISBN 0 08 0424325, pp.109-137.

NATO, 1997. Digital Geographic Exchange Standard DIGEST (Standard Agreement [STANAG] 7074, NATO.

NATO, 1979. Evaluation of Land maps, Aeronautical Charts and Digital Topographic Data (Standard Agreement [STANAG] 2215, NATO.

Taştan, H., 1991. Geographic Information Systems Design & Implementation of a GIS, M.Sc. Thesis (Supervisor: Prof. Dr. M. Orhan ALTAN), Istanbul Technical University, Division of Geodesy and Photogrammetry Engineering, Istanbul.

Taştan, H., 1997. An Automatic Quality Control System for Digital Elevation Models, In: Second Turkish-German Joint Geodetic Days Proceedings (Eds: M. Orhan Altan, Lothar Gründig), Berlin, Germany, pp.677-686.

Taştan, H., 1999. Data Quality In Geographic Information Systems, Ph.D. Thesis, (Supervisor: Prof. Dr. M. Orhan ALTAN), Istanbul Technical University, Division of Geodesy and Photogrammetry Engineering, Istanbul.

Taştan, H., Altan, M.O., 1999. Spatial Data Quality. In: Third Turkish-German Joint Geodetic Days Proceedings (Eds: M. Orhan Altan, Lothar Gründig), Istanbul, Turkey, Vol. I, pp.15-30.
http://www.hgk.mil.tr/teknik_bildiriler/cbs/ozet/htas99.html (31 Mar 2000)

Thapa, K., and J. Bossler, 1992. Accuracy of Spatial Data Used in geographic Information Systems, PE&RS, Vol.58, No.6, pp.835-841.

U.S. Bureau of the Budget, 1947. United States National Map Accuracy Standards, U.S. Bureau of Budget, USA.

Veregin, H., 1998. Data Quality Measurement and Assessment. NCGIA Core Curriculum in GIScience.
http://www.ncgia.ucsb.edu/giscc/units/u100/u100_f.html (20 Oct. 1999)