Spatial technologies to evaluate vectorial samples quality in maps production

Abraham Cárdenas Tristán*
Eduardo Javier Treviño Garza**
Oscar Alberto Aguirre Calderón**
Javier Jiménez Pérez**
Marco Aurelio González Tagle**
Xanat Antonio Némiga***

Abstract. Despite significant progress in recent years, the methodological conceptualization for assessing quality of vectors that integrate digital mapping is still a complicated task. Due to the fact that there is not an official scheme in Mexico to evaluate vectorial cartographic quality, an alternative methodology is proposed for assessing vectorial quality through analysis of samples at various vectorial scales from the Mexican Republic coverage. The tests conducted with various spatial technologies are under the norm frame TC/211 (ISO19113 e ISO19114), and these have been developed with support of companies who are producers of new spatial technologies and supported by the official producing agency of vectorial information in the country. Of which it is intended searching for appropriate evidences and potential indicators to determine norms or specific models to evaluate quality, for potential benefit of cartographic production in natural resources use and others potential applications of them. The methodology described pursues current advances in research to establish an improvement in the assessment policies in vectorial editing and mapping carried out by international agencies, universities and research centers. To make such a proposal in vectorial assessment quality, recognition was made from different approaches of those who worked in the field.

Key words: Spatial data quality, cartography, maps production, spatial technologies, vectorial data.

Uso de tecnologías espaciales para evaluar la calidad de muestras vectoriales de la producción de cartografía

Resumen. A pesar de importantes progresos realizados en la materia en los últimos años, la conceptualización de la metodología para evaluar la calidad de vectores que integran la cartografía digital es aún una tarea complicada, no existiendo un esquema oficial de evaluación de la calidad de la producción cartográfica vectorial en el país. Se propone una metodología para evaluar la calidad de la producción cartográfica a través del análisis de muestras aplicadas a las

* Universidad Autónoma de San Luis Potosí (UASLP), Facultad de Ingeniería, Av. Dr. Manuel Nava # 8, Zona Universitaria, 78290, San Luis Potosí, S.L.P. México. E-mail: abraham.cardenas@uaslp.mx
** Universidad Autónoma de Nuevo León (UANL), Facultad de Ciencias Forestales, Carretera Nacional km. 145, AP 41, 67700, Linares, Nuevo León, México. E-mail: ejtrevin@gmail.com, oscar.aguirrecl@uanl.edu.mx, javier.jimenezp@uanl.edu.mx, marco.tagle@gmail.com
*** Universidad Autónoma del Estado de México (UAEM), Facultad de Geografía, Cerro de Coatepec s/n, Ciudad Universitaria, 50100, Toluca, Estado de México. E-mail: xanynemiga@rocketmail.com

A. Cárdenas Tristán, E. J. Treviño Garza, O. A. Aguirre Calderón, J. Jiménez Pérez, M. A. González Tagle and X. A. Némiga

diversas escalas vectoriales de la cobertura del territorio de la República Mexicana. Las pruebas realizadas con el uso de diversas tecnologías espaciales, se encuentran dentro la norma TC/211 (ISO19113 e ISO19114), éstas han sido desarrolladas con el apoyo de compañías productoras de nuevas tecnologías espaciales así como del organismo oficial, productor de información vectorial en el país. Se tiene como objetivo buscar justificaciones pertinentes e indicadores potenciales, para determinar normas o modelos específicos de evaluación de la calidad, beneficiando el potencial de la producción cartográfica en el aprovechamiento de los recursos naturales y las frecuentes aplicaciones potenciales de la misma. La metodología utilizada va a la par de los avances en la investigación para establecer una mejora en las políticas de evaluación y de edición de cartografía vectorial, llevada a cabo por organismos internacionales, universidades y centros de investigación.

Palabras clave: Calidad de datos espaciales, cartografía, producción de mapas, tecnologías espaciales, datos vectoriales.

INTRODUCTION

The increase and development of new spatial technologies in recent years, has allowed to address various problems in the geographical information field; from in the way in which it's generated, the methods to manipulate its primitive vectorial components (points, lines and the shaping of polygons), the adaptability of information to certain punctual users necessities, the opportunity to make the various formats in which information is presented interoperable, the adaptation to new spatial data infrastructures and the opportunity to use it from new knowledge perspectives. If in the past, the cartographic production was limited practically to the mass production of static maps, the progress of recent years has facilitated the increase of other kind of maps: on demand maps (Sabo, 2007). Maps on demand are a kind of cartography that is generated according to specific requirements of different users, contrary to the traditional mapping generation that has been produced in large quantities to meet general requirements.

Today, the cartography on demand has exceeded expectations, due to many informatics developments, both commercial software (licensed software) and the use of free software led by the Open Geospatial Consortium (OGC). Also, the possibility that Internet gives us through the use of Web services, using a variety CASE tools (Computer Aided Software Engineering) and new spatial technologies. Similarly, the access to different geographic information databases, with the opportunity to enter different online servers allowing free downloads of such information. In which it required to produce traditional mapping that needed the experts hand, on-demand mapping, in union with the democratization information concept, allowed users with new knowledge in geomatics and using the employment of the mentioned technologies (Geographic Information Systems (GIS)), use of satellite images, software for automatic cartographic generalization, the satellite positioning systems, the videogrammetry, LIDAR technology, among others) can simply and quickly produce desired mapping without going into lengthy traditional processes of training and acquisition of knowledge and experience through the years. On the other hand, untrained users and stakeholders, have the same rights arising from the freedom of spatial technologies use and their application in the geographical information management to manipulate and generate maps on demand. Goodchild (1995), described this situation as worrisome due to poor accountability of users without adequate knowledge wanting to conceive on demand mapping. "GIS is its own worst enemy: by inviting people to find new uses for data, it also invites them to be irresponsible in their use". With the use of geographic information democratization and its accessibility, mapping today must meet different specific needs, whether in scale terms, different topics, graphic semiology, and potentially in specific diversification elements for natural resources utilization that today are potential trends in order to regulate, control, measure, preserve, manage and to take advantage of new economic dependency perspectives of territories.

Considering recent technological advances and trying to answer one of the most important needs for the last 30 years, "The spatial data quality generation", has not been considered. It seeks to find a territory knowledge as faithful and appropriate as possible, through the cartographic production to

consider elements corresponding to their quality generation. Mapping production traditionally has been a long process that involves the acquisition and the validation of information, the cartographic databases development and cartographic generation at different scales. Procedures that have sought to optimize through automatic cartographic generalization (McMaster, 1991; Weibel and Dutton, 1999; Allouche and Moulin, 2001; Jabeur, 2006) and multiple representation (Rigaux, 1994; Vangenot, 1998; Müller *et al.*, 1995; Devogele *et al.*, 2002; Bédard *et al.*, 2002; Bernier, 2002; Cárdenas, 2004) facilitating the automatic generalization. These aspects of cartographic production optimization, has seeked, to become current standards, which on one hand, search to respond to common needs of precise geographic information and not leaving aside the intention to take care of producing this data in quality terms. However, for the mass users, the quality search in data production is uncommon. Recently, Kumi-Boateng and Yakubu (2010), raised awareness in the policies establishment to authenticate quality of spatial data production, "Is not only useful for in-house data development, but data customers and users are able to determine the validity of data by checking the sources and procedures used to create the data". The objective sought in this paper is to propose an alternative methodology for assessing vectorial quality through sample analysis at various vectorial scales from the Mexican Republic coverage.

BACKGROUND

In the last decade numerous alternatives have been proposed allowing the assessment cartographic quality produced in different countries. These alternatives have been the answer to growing needs in the spatial data quality determination of cartographic sources. Countries like the United Kingdom, France, Canada and Spain have established mechanisms that have shared with the international community, as the case of Ariza (2002, 2004) in Spain with two published books. Also, extensive research conducted in cartographic quality field has been developed. Such case Gago *et al.* (2006),

who worked on a methodology development for samples acquisition coordinates of planimetric and altimetric points from a specific area to certain scale, where a cartographic sheet was restituted photogrammetrically from the same study area, composed of aerial photographs to 1:25 000 scale with intention to compare the vectorial data obtained in order to standardize correspondences through formulas in a multicriteria analysis, looking for accuracy between elements representing same place. Ariza (2004) developed further work with the book publication "Casos prácticos de calidad en la producción cartográfica" ("Practical Quality Cases in the Cartographic Production"), such work is oriented to cartographic quality control. The authors present 31 cases about improving quality, sample sizes, process control, positional and thematic components, simulation and geographic databases. The practical cases presented were developed in cartography already implemented. Pavicic *et al.* (2004) collaborated on a quality system for new cartographic generation of Croatia at 1:25 000 scale, benefiting topographic features. In this system a production control model of topographic maps according to ISO (2008, 2010) specifications in the elements quality was implemented. The objective of this data production project was to generate databases considering the positional accuracy, in such a way that cartography production at smaller scales will benefit (1:50 000, 1:100 000 and 1:200 000). From this process emerged a manual for control quality processes, the process was automated with the spatial technology adoption FME (Safe Software Inc.) to detect anomalies, analyzing related files with cartographic product specifications, using FME Workbench operators. The system could detect an errors variety in the analized objects (in geometric correspondences, semantic, semiological, code classification in the polygons construction conforming buildings, areas, land use) that are controllable through a series of statistic reports for the quality control processes. Jobs and Twaroch (2006) presented an evaluation method, based on stochastic reasoning to support perceptible maps design, through a computational model that allows map designers to take appropriate parameters choice and interaction between

them, being supported by a decision process. Were used a bayesian network libraries provided by Microsoft Research. Bartoschek *et al.* (2006) conducted a study vectoring 153 maps of Alentejo National Ecological Reserve (REN), in southern Portugal, in order to measure spatial accuracy and ensure compliance with the original mapping, since each county produces its own maps; it was possible to generate diversity by implementing semiology on maps. The methodology consisted in the "sampling" algorithm implementation within of interface design, where the algorithm lists the class number and these combination in the original BD with respective areas for each county. The algorithm generates samples from original objects size, compares them with the REN digital classification and then reference with paper maps on the sampled points to calculate positional and thematic errors providing estimates and sampling error. Gui *et al.*, (2008) describe a methodology through algorithms to analyze a cadastral map scale 1:1 000 with 400 plots examined, of which various inconsistencies in spatial quality problems are derived. Sarmento *et al.* (2008), developed a methodology to assess maps thematic accuracy from land coverage that uses uncertain references in their conformation characteristics. This methodology involves the *fuzzy synthetic evaluation* (FSE), based on the linguistic fuzzy operators combination where specifically the land coverage magnitude from errors is evaluated by class, and measures their weight in the assessing map accuracy process. Stehman (2008), describes how to sample designs to assess maps accuracy, because of its demand, of the spatial data increase and to its use. Is specified that at this moment, the assessment elements are based beyond of a matrix error implementation, situation generated by the diversification detailed requirements to know about coverage characteristics of earth elements, on which new challenges appear.

Wu *et al.* (2010), introduce a new concept "the tetrahedron model" for analysis of cartographic quality control, detecting error in its production stages. This proposal is a quality analysis model that proposes references to people who provide data, who manipulates data and who verify it. The user is considered as a quality factor controller who would have the same status as a producer. However, authors believe that further exploration and analysis should be made to solve incertitude in cartographic evaluation errors. On the other hand, there has been significant research in the spatial data quality scope, such work has been embodied in Devillers and Jeansoulin's book, 2005 (Qualité de l'information géographique) and the book "Spatial data quality: from process to decisions", Devillers and Goodchild (2010). On such research and its methodological processes description, it has been shown that generation models quality is essential for required needs of a given territory. As described Shi (2008), "Quality control for spatial data Refers to Developing Method to Ensure the final spatial data are produced to meet the users requirements". Similarly, methodologies that have been generated recently and those arising are alternatives that are changing opportunity to enhance purposes for assessing the cartographic production quality.

DATA, USED MATERIALS AND INITIAL METHODOLOGY

To start with the analysis stages, at first it was important to become familiar with the vectorial cartographic databases from Mexico, as well as to obtain a used data description and how these would be analyzed. Used methodology consisted, in making a conceptual model of the analysis problem with UML (Unified Modeling Language), making a review of the country's cartography current state and familiarize with vectorial cartographic databases at different production scales (1:20K; 1:50K; 1:250K and 1:1,000K). The technologies that served these experimentations are FME Workbench, ArcGIS, DataViewer and Google Earth (GE). The data used to carry out such experimentations were officially requested to the National Institute of Stadistics and Geography (INEGI). To proceed to direct data representation in the study particular area, these were checked and at time were adapted to the technologies operators used. The following Table 1 it describes in detail the reference basis from coordinate systems used in analyzed cartographic scales.

Table 1. Reference systems description from cartographic scales used

Current coordinate system:			
1:20 000 scale		1:250 000 scale	1:1 000 000 scale
Geographic Projection UTM, Universal Transverse Mercator **Datum**: ITRF92 **Update methods:** Photogrametric-Field Classification (Overlay Raster/Vector) **Edition date**: 2005 (field classification), 2006 (digital update), vectorial data model, topographic data dictionay, standards for data capture), 2008 (start the editing process) **Contour intervals**: 10 meters **Photography scale used**: 1:40 000 **Spatial Resolution:** 15 meters	**Geographic Projection** UTM, Universal Transverse Mercator **Datum**: ITRF92 Edition date 1993 (vectorial data models), 1994 (altimetry and standards for data capture), 1996 (standards for data capture and topographic data dictionary), 1997 (digital update date), 2002classification of field and compilation), 2003 (digital update). **Contour intervals**: 20 meters **Editing Software**: Microstation **Spatial Resolution**: **Photography scale used**: 1:75 000	**Geographic projection**: Lambert Conformal Conic **Datum** ITRF92 **Edition date** 1996 (data dictionary), 2003 (standards for data capture and field classification), 2004 update digital) **Contour intervals**: 100 meters **Editing Software**: AutocadMap **Spatial Resolution**: 15 meters	**Reference system**: Planimetric: NAD27, Clarke ellipsoid 1866 Altimetric: Mean sea level North American vertical datum of 1929 **Geographic projection**: Lambert Conformal Conic **Updated datum**: ITRF92 **Edition date**: 1996 (data dictionary), 2000 (set compilation), 2003 standards for data capture and field classification), 2004 (digital update) **Editing methods**: manual cartographic generalization (mapping computer assisted) based on topographic vectorial cartography 1:250 000, series 2, updated from November 1995 to November 1997. **Contour intervals**: 200 meters **Editing Software**: AutocadMap

Once cartography was analyzed and referenced in the used technologies, we proceeded with some cuts analysis made in ArcGIS, which in turn served as elements, and file forms, they would be integrated into a FME Workbench platform. Subsequently, a platform operator series were used to start with layer integration from required information. This subsequently classified the following information layers that were integrated and their respective scales (Table 2).

Spatial data integration based on schemes for vectorial information analyzing

The initial stage for analyzing vectorial cartography samples at different scales was made based on a cartographic cuts series from selected spaces in which we had vectorial coverage. Due to the large amount of information which involved a lengthy process it was decided to analyze vectorial information in corresponding cuts from the state of San Luis Potosi. The cuts purpose was to select areas of analysis, trough several information samples, allowing to carry such samples through shapefiles into an assessment process with FME Workbench

Table 2. Integration of .shp layers in FME

File type	Operator	Geometric element
.shp 1:1m .shp 1:250k .shp 1:50k	NeighborFinder	Lines (contours)
.shp 1: 1m .shp 1:250	NeighborFinder	Lines (contours)
.shp 1:50k .shp 1:20k	AtributeFilter AtributeValueMapper AtributeCreator	Polygons (blocks, buildings)
.shp 1:20k .shp 1:50k	NeighborFinder	Polygons, Lines (blocks, communication routes)
.shp 1:20k .shp 1:50k	AtributeFilter AtributeValueMapper AtributeCreator	Polygons, (blocks, buildings)

technology. With this technology an integration process took place which implied the information analysis from different scales of same study area (integration samples were performed as described in Table 2). Pouliot (2002), describes the spatial data integration concept as a process (methodological or technological) that involves data space-time combination coming from different sources to extract a greater variety and better quality information. So, when performing an integration process, the combined data may include multi-temporal data, different spatial data resolutions, data from various sensors, data of diverse formats types, etc. Considering the different data sets that were used

for data integration indicated in Table 2, the spatial data inventory was left well structured, representing the urban San Luis Potosí zone. Subsequently, we proceeded to analyze the spatial data set included in the corresponding databases. To show anomalies of geometric, topological and semantic correspondences, the FME analysis consisted in data integrating from the same or different information sources and different scales from the same sites where objects should correspond to each other. From which, like in the example in Figure 1, a small extract was taken from a vectorial information sample, describing topological inconsistencies in the relationship between scales, as well as inconsistencies in geometric representation from the description of same objects.

In most cases, information integration from vectorial samples at different scales analyzed show in high percentage the same problematic in the indicated correspondences. Sometimes the semantic representation evidences changes, due to data sets produced temporality, by facts that scale production 1:50 000 was edited between 1968 and 1988 being integrated at 1:20 000 scales which has been edited recently. Such processes will be able to be detailed later. The integration purpose of data sets mentioned above was performed to analyze the geometric relationships, topological, semantic and positional accuracy between different information scales on the same territory. Since the 1:50 000 cartography has been essentially used in the country over the past 40 years, and having begun the editing mapping process to 1:20 000, we were interested in knowing the correspondences between the two information databases. For example, it should be
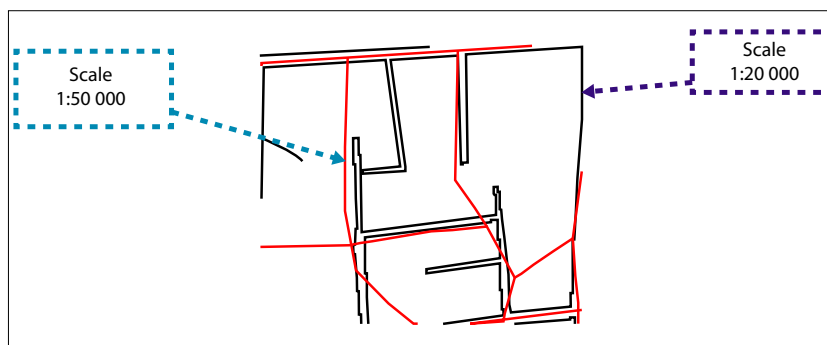

Figure 1. Evaluation example in FME of geometric, topological and semantic correspondences from a same site at two different scales.

common that the contours representation 1:50 000, that goes to each 10 meters, regarding the contours at 1:20 000 which also are set to the same equidistance, these should match respectively in geometry terms and positional accuracy. However, since in the editing process the cartography at 1:50 000 emerges from a photography scale to 1:75 000 and the 1:20 000 cartography emerges from a photography scale to 1:40 000. Such situation affects perhaps the resolution of both photographic scales and also considering the following circumstances, which have to do with alternative editing processes have been used over time, technological advances,

the different staff working in the editing process and the application of different regulations over time. To demonstrate these differences to resort migrate vectorial files in .shp format of cartographic sheets (f14a84_50k and f14a84d_20k) on both scales indicated in GE platform, according with procedures to be detailed further ahead. Such representations are outlined in Figures 2 and 3. According to geometric representation analysis, it is possible to describe vectorial cartography that best represents the orographic characteristics in GE images is the scale 1:20 000, equally it is possible to be verified that the editing processes
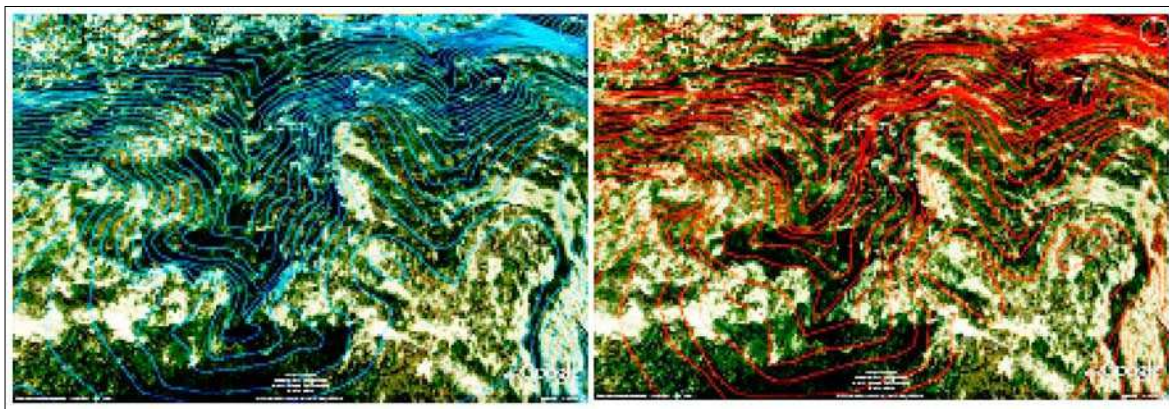
Figure 2. Cartographic layers representation in blue 1:20 000 scale and in red 1:50 000 scale of the same area (integration in Google Earth, 2011).
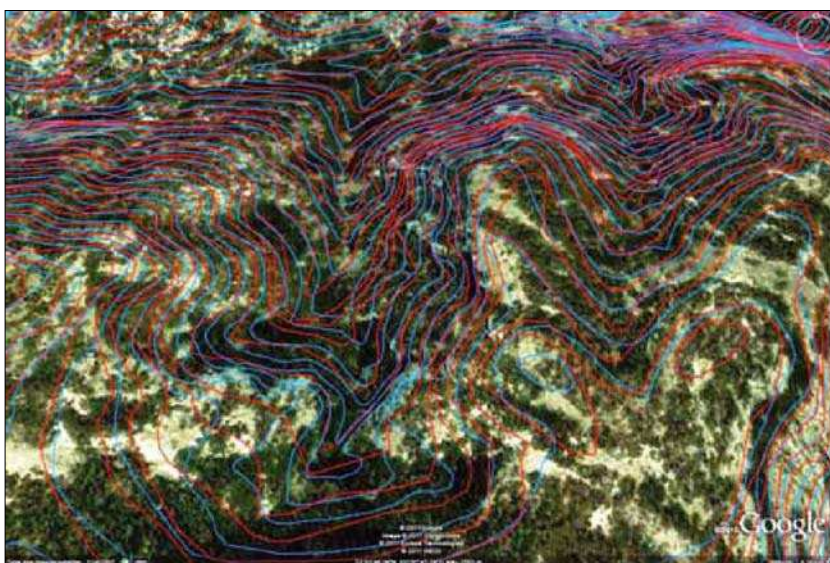
Figure 3. Representation of both cartographic layers, showing differences in geometric representation (integration in Google Earth, 2011).

have been of better quality than the processes carried out with the curves on scale 1:50 000. Since this vectorial cartography constitution is recent and in which have been used new editing technologies, it is natural that this happens; nevertheless, it does not have to be omitted that in the analysis process of both vectorial coverage, exist editing anomalies, that even imply important lack in the information representation on selected samples of the territory to be analyzed.

Information analysis in favor of measuring geometric, topological and positional accuracy metrics

Given previous processes to establish the methodological procedure, we resorted to use a geospatial technology that gave us the opportunity to enter in the steps to follow for the evaluation of specific aspects from cartographic quality. GIS Data Reviewer is an ArcGIS Desktop extension that provides a complete set of quality control tools (QC) to simplify many spatial quality control aspects with visual and automated procedures. Initially, we wanted to work with two tools, GIS Data Review and Geo Network; however we were not able to use the first, due to installation issues and technology incompatibility with systems that we currently have.

Subsequently, and being familiar with handling "Data ReViewer" technology, it was managed to conform the cartographic sheet integration of the selected territory coverage at a 1:50 000 scale, as well as the 6 corresponding cartographic sheets at a 1:20 000 scale representing the same study area. Table 3 provides a description coverage area of both scales and their different characteristics. Since the information amount is vast, we concentrated in analyzing the information layers that were more representative (contours, communication routes, streets, blocks, constructions, etc.), this allows the geometric integration analysis through their primitive ones (points, lines and the polygons conformation). It was planned to analyze, how information layers are represented in both cases and may have different information names they represent. Also, we focused to measure geometrical and topological representation differences and hence the positional accuracy degree. These indicators are described in section 4.

The selection of information layers to make its composition analysis of vectors, from geometric primitives, was executed according to existing set parameters of DataReviewer towards assessing vectorial quality. There are about 42 spatial operators, which are handled through direct operations, determining the appropriate parameters; according to data type that is being manipulated. Also, quality assessment can be done through SQL queries, having in existence the suitable database conformation that is build in ArcCatalog, its extension is .gdb and to read the ArcMap files, .shp or .mxd extensions are used. DataReviewer also allowed us an important way to check the representation current status from geometry and topology of certain objects that mostly did not correspond to the true territory features representation in the study

Table 3. Description and coverage characteristics of cartographic sheets at 20K and 50K scales (INEGI)

1:20 000 scale	1:50 000 scale
Features: Cartographic product that integrates information of: infrastructure, orography, hydrography and country population, compiled by photogrammetric techniques from aerial photographs, geodetic information and field verification	**Features**: Cartographic product that integrates information of: infrastructure, orography, hydrography and country population, compiled by photogrammetric techniques from aerial photographs, geodetic information and field verification
Coverage format: Seven minutes thirty seconds of latitude by six minutes forty seconds in length (Latitude: 07° 30' Length: 06° 40')	**Coverage format**: Fifteen minutes of latitude by twenty minutes in length (Latitude: 15° Length: 20°)
Area in km²: Approximated area of 160 km^2	**Area in km²**: Approximated area of 960 km^2

areas. In processes that are going to be described in the following section, it was managed to find anomalies in vectorial data integration at certain scale on the territory corresponding portions on GE platform. The properly edited vectorial representation, allowed to generate the statistical indicators of the editing process, which generated an alternative analysis of the vectorial cartography quality, resolving the anomaly within DataReviewer operators.

QUALITY ASSESSMENT RESULTS OF THE VARIOUS TECHNOLOGIES USED

Geometric correspondences assessment

Most of the representation conflicts of vectorial information have to do with the detail level in its geometry. And for such situations, several data sets were analyzed, which implicitly in their establishment, have abnormalities linked to way they are edited. Since vectorial structure is made up of points, lines and the polygons conformation, this has adapted to mathematical structures in plane geometry. Which divides the space in a discontinuous manner, being associated with the Cartesian metrics (X, Y, Z). This structure is well adapted to the easily identifiable border entities representation, such as administrative boundaries, property, engineering works, territorial boundaries, etc. The topology of this structure is not implicit and can be specified in different ways. This vectorial structure is codified in different ways (simple vector, connected vector and topological vector) the encoding conformation was examined for analysis purposes of such vectorial cartography with the technology operators used. Spatial operators of this technology are algorithms that execute various functions to facilitate requirements analysis to evaluate information quality, constituted through vectorial primitives. To begin with the first analysis, a set of spatial operators was chosen, those applied to the curves set, belonging to a cartographic sheet (f14a83), at 1:50 000 scales (Figure 4).

It was intended to check the regular expression of the vectors constitution (polylines seriation) conforming curves; a verification of their elements,

Figure 4. Verification report in group from vectorial curves set of the cartographic sheet (f14a83), at 50K scale, using several spatial operators.

a vertices conformation assessment, a verification in the polylines reduction, the non-linear segments, the lines length, the invalid geometry, the multipart polylines, and the trajectories polylines closure. The process was made with the application of these spatial operators, only taking effect, those in which there were anomalies in their analysis functions, for the assessing quality purpose. This report represents the percentage indicator of the evaluation that is performed with different spatial operators of technology. Since geometric evaluation of constitution vectors conforming the curves object, and by being represented as an interconnected polylines series, expressing the terrain shape elevation, their quality evaluation additionally has to evidence the geometric shape representation over a background image in the corresponding territory. This process usually can be done within DataReviewer, but this time, the orthophotos we had are from much earlier dates and have poor resolution, this situation led us to experiment with GE pro platform, which has a higher image resolution, it's updated and can interoperate

via shape files, with a procedure implementation series that has already been experienced.

To support consistency in the issue of representation geometric correlation between same sites information, we supported it with FME workbench technology, which allowed us to integrate different vector data layers with up to three different data sets at scales 1:20 000, 1:50 000 and 1:250 000, using different spatial technology operators. On the other hand, the graphical semiology parameters of Data Viewer did not allowed us to analyze the integration of the three different data sets with different textures and colors; however the tests were achieved by using a spatial operators group to the integrated set of vectorial scales in the objects denomination; curves, communication routes, streets, blocks and buildings. The following Figures 5 and 6 illustrate current correlation results in terms of curves.

By verifying the same curves corresponding at different scales, the geometric representation has to be similar; nevertheless, it differs greatly in the logic representation territory. Similarly, if you compare it with a background image, it can be seen that such inconsistencies are represented in the study area. The scales 1:250 000 and 1:1 000 000 have reportedly been subjected to cartographic generalization processes of manual type, due to lack of technology and perhaps to the little given importance of the quality issue adjudicated from these scales that should represent the territory.

Topological correspondences assessment

About topological correlation evaluation, the analysis in DataReviewer was particularly complex, because we use a recent trial in which, spatial operators that verify topology were disabled. However, there were several anomalies found related to the topology between corresponding objects in the same area at two different scales (1:20 000 and 1:50 000), these scales were the ones that interested us the most due to the representativeness of detailed objects in the information layers describing the Mexican territory. We tried to measure the related inconsistencies between the same objects assessed by analyzing sampled topology problems, as noted above. From such analysis, we found significant differences related to various problems in traditional editing processes. By analyzing the cartographic sheet f14a84_continuo curva_50_utm and reviewing their attributes detailing its ID, the polyline classification, the curve elevation, key, among other attributes, and comparing these integration with f14a84d_curva nivel_20_utm cartographic sheet, which corresponds to the same place, just as checking same attributes we could perceive differences in topological order indicated in the Figure 7.

What struck us was that while both scales curves edited every 10 meters, these have wide correspondence differences relating to representation of the terrain they describe. In blue continuous curves are described at 50K with poor quality in editing geometry with various and frequent peaks that if we analyze at a zoom, it could be seen that the edi-
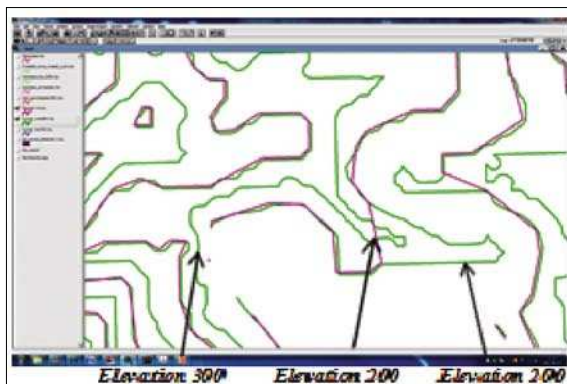

Figura 5. Vector layers integration 1:250 000 (green) and 1: 1 000 000 (pink).
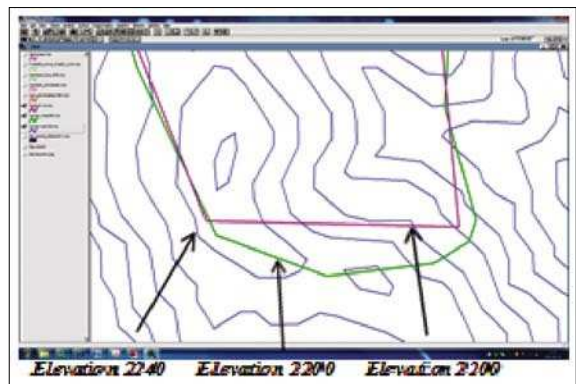

Figura 6. Layers integration 1:1 000 000, curves at 1:250 000 (green) and 1:50 000 (blue).

tion makes no sense on respect to elevation of the territory it represents. Likewise, the lack of quality problem can be found frequently in a large number of cartographic sheets covering the Mexican territory from the same scale. But then, green describes contours at 20k, which at the detail level enhances the editing possibility. Inconsistencies persist in certain cartographic sheets of the same scale. In both situations the issue would be the curves being represented in the same elevation to a given territory, these lack logic and the problem has wanted to evaluate from a topological point of view. In addition, we analyzed another data set corresponding to cartographic sheet *f14a84_calle_50_utm (*polylines in purple*)* and cartographic sheet *f14a84d_manzana_20_utm,* according to a same zone integration in DataReviewer technology (Figure 8).

About, the cartographic layer description in purple, corresponding to cartographic sheet f14a84_calle_50_utm, describes a field representation that differs from the map layer representation in blue which describes geometry corresponding to f14a84d_manzana_20_utm cartographic sheet, where it displays a problem in topological order between the information in both scales. Continuing with the analysis, another data set was added to the maps sheet f14a84d_calle_20_utm (purple) and f14a84d_carretera_20_utm (green), (Figure 9).

In this analysis it was found that equally in information layers of same scale, there are significant differences of geometric representation order, which result in topology problems about the territory they represent. This analysis situation, allowed us to state the following; if the vector mapping use at 1:50 000 scale, in which many projects have been developed over decades and these being based on map production, what reliability could have been of any user from vectorial cartography, which has generated developments? We should rather exercise certain level of distrust, which has been developed in many projects on the territory taken from the base to cartography 1:50 000. Being in process, the vectorial mapping development at 1:20 000 scales in which the information level detail must be accurate and must describe the territory with a clearer representation. For even on this scale ano-
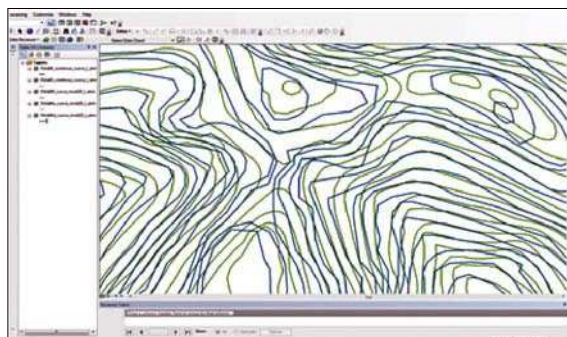

Figure 7. Representation and topological analysis of curve_50_utm f14a84_continuo cartographic sheet (blue), in relation to f14a84d_curva nivel_20_utm cartographic sheet (green).
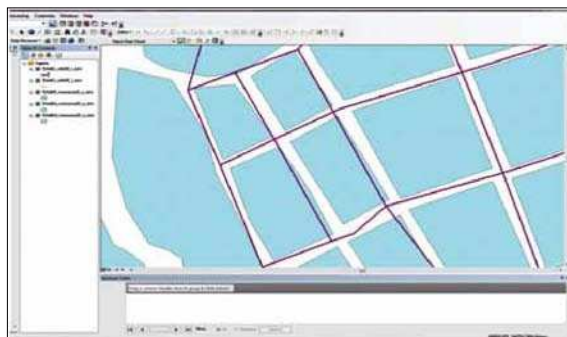

Figure 8. Representation and topological analysis of the *f14a84_calle_50_utm* cartographic sheet in connection with the *f14a84d_manzana_20_utm.,*cartographic sheet.
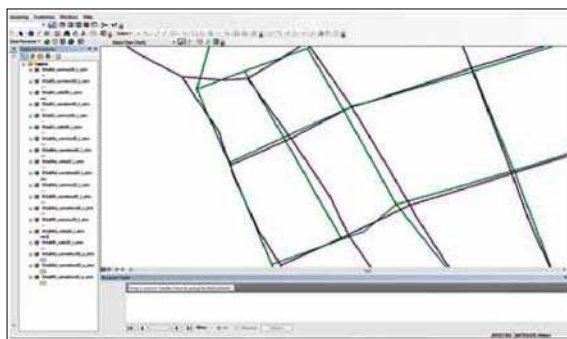

Figure 9. Representation and topological analysis of the cartographic sheet f14a84d_calle_20_utm in connection with f14a84d_carretera_20_utm cartographic sheet.

malies are manifested, which may correspond to the actual editing process or the lack of attention to regulatory process implementation for assessing the cartographic production quality.

Semantic correspondences assessment

In the technology used to evaluate vectorial quality, there are no spatial operators that can generate a semantic and thematic assessment of spatial datasets consistency, which have been selected in vector samples or cuts. Initially we supported it with FME. With the aim of ensuring a matching level of data meaning, related to identification and description. Semantic integration of certain spatial datasets were carried out, using data of two and three different scales (1:20 000, 1:50 000 and 1:250 000) on the same territory. Since in the country we do not have a spatial data infrastructure fully established, we obtained a semantic reference basis, seeking compatibility with datasets from scales databases used. Generally, geospatial data infrastructures being established to organize and manage the large amount of spatial information that integrates a country conformation can be taken as a reference infrastructure, comparable to the mass production of geographic information generated by the various agencies and institutions that produce it.

To ensure a geometric relationships combination between the same theme primitives in liaison to cartographic sheets for study areas across samples, our opportunity, was based on using existing databases and making comparisons between them. These INEGI map sheets were as follows: scale 1:50 000 (F14A84 and F14A83), scale 1:20 000 (F14A83C, F14A83F, F14A84A, F14A84D and F14A84E). In the analysis of such datasets it was found that there are no major semantic problems encountered. The analyzed datasets have the same attributes with the same structure, the object classes differ in name, but the value domains for object classes are comparable between both information representations.

Results of editing geometric anomalies supporting the quality process assessment in DataReviewer from Editor tool (reshape feature tool and edit vertices)

One advantage of the technology used to represent geographical objects formed by vectorial structures, was able to manipulate with editing tools, the shape and vectors position, so to make analysis of information quality. Opportunity, which became a mean to evaluate quality, that allows correct and instantly reissue the geometric representation anomalies, as well as topological problematic anomalies of position and shape. The technology used requires the database to be evaluated, being identified by the file format .gdb, which is read in ArcMap. Initially, using background referenced images and integrating cartographic layers or cuts from study areas, which overlaps vectorial data layer or cartographic cut on particular images and takes place at a current state identification from geometric representativeness of vectorial cartography in relation to the image information correspondence. Note that the parameters in images referencing in both DataReviewer platform like *Google* earth pro, should correspond to the relation of spatial parameters from vector data being analyzed. In the following Figures 10, 11, 12 and 13, the steps procedures are described to reedit vectors, in order to display a better geometry correspondence representing the object.

Within the integration processes of cartographic cut and the quality positional analysis in relation to study area images, there were 1843 occurrences from value domain in the object class known as "FID" determined in the technology used for quality evaluation. The problematic ones indicate a geometric mismatch that is a common display in the block occurrences that represent all housing in the analysis area (Figure 11). This geometric mismatch is manifested by an average of 3 to 4 meters from each vertex position. In other instances around the study area surroundings, manifests a varied geometric mismatch from 0 to 4 meters. The geometry analysis correspondence of vectors representing the objects variety from study area through the value domain "blocks" comes from a source editing vector maps at 1:20 000 scale, where noted above that the map scale integrates the following information: infrastructure, topography, hydrography, and population. Since the details of items mentioned are important in this scale, the editing process quality generated by photogrammetric means is not subject to a formal evaluation process that could correct anomalies and eventually verify the production quality.

To continue with the editing process instance 1454, in DataReviewer Editor tool that shows a

table of description attributes (Reshape Tool and Edit Feature vertices) was selected (Figure 12). Now, we proceeded to prepare the object to start the position analysis with reference to the base image that indicates geometric mismatch anomalies and proceeded to reedit. At this stage, you have to be very careful of what you are going to select as anomalies, because everything depends on the reissue purpose. Also, the criteria to take in base what is necessary to reedit, depends on the specific analysis needs, on which and according to certain purposes, it must generate vectorial quality.

Importantly for the vector analysis representing the home-room in the current geometry of cartographic sheet F14A8d at 1:20 000 scale, we would enter an exhaustive quality evaluation process. So to reissue the current geometric conformation of the objects that were indicated, the aim should be well determined, since there would be too many objects to reedit and the process would be long. To support this analysis methodology, we focus on showing the object in single editing, which exemplifies the processes to follow if you wish to reedit large amounts of objects. This methodology, supports current research in the interactive approach or amplified intelligence (Sabo, 2007), which uses commercial systems for interactive generalization; due to the difficulty of using fully automated solutions. The object has been selected, corresponds to a reality geometric representation through the vectors conformation. This randomly chosen object, described his position from UTM coordinates (Figure 13), which were analyzed with respect to image coordinates, used as a basis for the study area.

Once indicating the object vertices on the image position, the coordinates are reissued and are drawing a new and better form of geometric representation, corresponding to the object to be edited (Figure 14). You will need to decide on the vertices number of object to be reedited and this depends on the best geometric representation that is sought on the objects to analyze. About, seeking to correct a geometric mismatch from occurrence representation 1454 of cartographic sheet mentioned above. Also, the positional accuracy of object geometry enables compliance with a procedural way

Figure 10. Detection and analysis of instance 1454, representing a house-room in the object class manzana20_ utm of cartographic sheet F14A8d at 20k.

Figure 11. Study area sample, which describes at center the home-room cluster, representing a geometric mismatch with respect to position.

to assessing vectorial quality, reediting with tools more adapted for representing the reality objects. Having said earlier that the existing orthophotos of study area had poor resolution and dating from earlier dates, these did not allowed us to see the updated geometric correspondences of current edition from INEGI's vectorial cartography at 1:20 000 scale. This situation led us to experiment with Google Earth pro platform, which allowed us to explore its feasibility in relation to the shapefiles import via kml format, and the various processes to
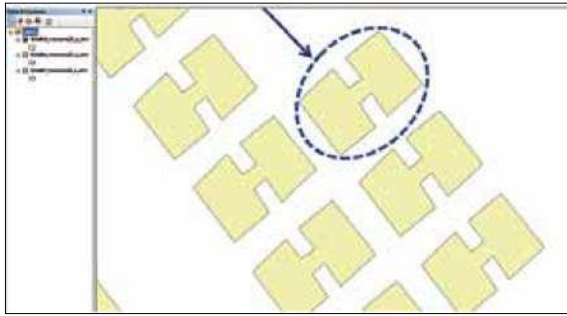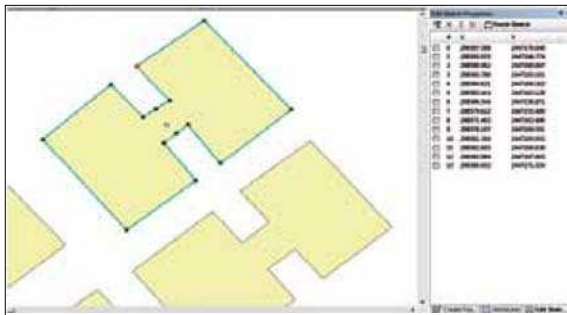
it, we used the tool Editor, Start Editing parameter, which allowed us to begin the process. After selecting the curve geometry, its current state can be appreciated, through a UTM coordinates series, which describes its current position. Once editing parameters were activated, we proceeded to edit the polyline with the Edit Vertices and Reshape Feature Tools. These tools allow to edit each part of the desired polyline, in such a way that it is based on its improved edition, its separation from other curves and its correct position according to the base image, set the new position curve which is composed of new coordinates series, describing its improved position (Figure 17).

Thus, by accepting the editing process, we save its new configuration and we proceed to check anomalies in the total of cartographic sheet cut in the evaluating quality process from geometric primitives. Each time it's necessary to evaluate quality, reissuing anomalies found in the information elements layers of the geographic objects (represented by geometric primitives); in the indicated edition tools there is no restriction to edit the primitive with new vertices numbers needed to improve the geometric representation quality and its object´s correct position.

Once the reissue process is completed, it is necessary to activate the evaluation parameters on the map sheet in analysis, to verify the anomalies existence on the continuous curves integration within the cartographic sheet. Next, Figure 18 describes an example report showing derived accuracy percentage from the evaluation process made with the parameter *Invalid geometry Check.*

The different reports generated from used analysis operators are processes in which, it must be clearly identified in the assessment technology, the analysis quality type to perform. There are
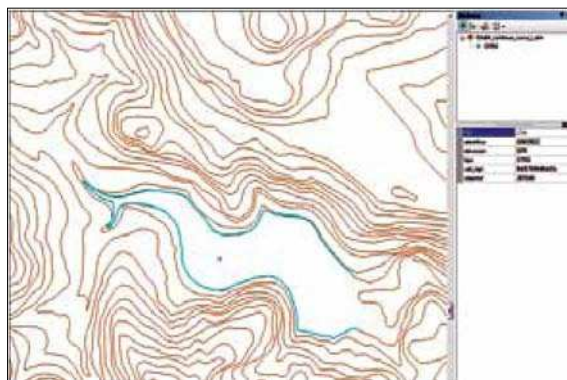

Figura 16. Selection of elevation curve 1970 from 1:50 000 scale for reissue process seeking to improve their geometric representation.
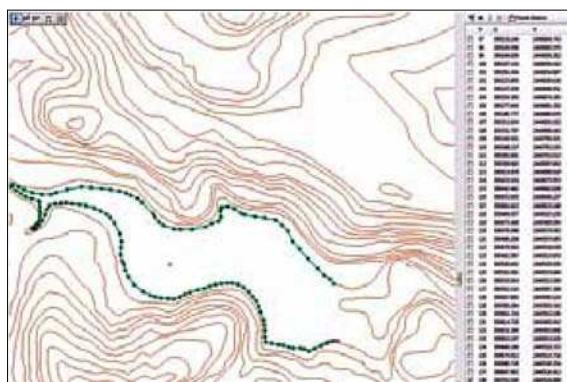

Figure 17. Reissue process of 1970 curve classified as FID "1744 showing its improved position.

procedures for quality information assessing that can measure anomalies depending on the parameters used. But it is complex to request assessment technology to be used, that promptly responds to specific analysis requirements. Since this tool has been designed to address specific situations of information quality evaluation, it does not meet all the needs of quality assessment.

Figure 18. Record evaluated report after editing anomalies on the cartographic sheet f14a84_continuo_curva.

Origin Table	Check Type	Check Title	Total Records*	Total Results	Total Percent Accuracy	Distinct Results	Distinct Percent Accuracy
f14a84_continuo_curva_l_utm			2369	1	99.96%	1	99.96%
	Invalid Geometry Check						
		Invalid Geometry Check (f14a84_continuo_curva_l_ut m)	2369	1	99.96%		
Total			2369	1	99.96%	1	99.96%

*This report is based on the total number of records in the Origin Table at the time this report is generated.

CONCLUSION

After analyzing diverse processes and methods carried out in various research works towards evaluating the information quality data and focused on vectorial constitution, we delve into *the interactive or amplified intelligence* research *approach*, experimenting with new spatial technologies to find a rapid and automated mechanism, being able to assess vectorial quality of large amounts of information. In this paper, it was described an alternative methodology for assessing the vectorial samples quality from cartographic production. This proposal emerges into a problematic that seeks to regulate the various organisms producing vectorial cartography, which over time have been emigrating in their production methods to each other, in order to optimize time and cost in these processes. Thus, by adapting to new methods of vectorial map production, a quality control of what these processes generate has not been carried. As a consequence this vectorial cartography edited decades ago with initial production methodologies, is coupled to recent cartographic production processes using new technologies and new methodological processes. In such situation, to couple and update mapping, there would be anomalies, generated of geometric, topological and positional accuracy. Little has been evaluated with rigorous regulated processes and most probably errors are propagated from existing production methods to others. For vectorial cartography production of the country, so far there is any instance or external review committee that trained in quality assessment standards or endorses these production processes and the information quality. The alternative proposal brings into question the need to revisit global regulations or in consequence adapt any rule or evaluation policy, which involve experts to audit on production processes and to suggest technologic mechanisms for review and quality analysis on the information generated.

The methodology used has involved organizing the vectorial information inventory, to make the quality assessment process through combine datasets from same study zones, formed by specific sites samples, because vectorial information is found in large amounts. It was necessary to work with cartographic sheets at different scales, in which, for tests optimization, some cuts have been made at different mapping information layers and in certain study areas. The different analysis geometry processes, topology, semantics, and positional accuracy have been the consideration subject in the vectorial quality evaluation, given to inconsistencies magnitude found in the information current representation at different edited scales. Such analysis concepts have been taken of the current ISO (2008, 2010) standard specifications for geographic information management. In the process of vectorial quality assessment, which has been effected using new spatial technologies, was carried out a methodology that integrates information layers and these being constituted by geometric primitives that allow adaptation with the use of spatial operators to evaluate vectorial quality. Such operators were simple in execution but were complicated in determine certain assessment functions. Nevertheless, it achieved to conclude with a clear result. Because there is not a unique geospatial technology to implement all analysis procedures for vectorial quality, several technologies were integrated for the process. From which, was achieved to measure with an indicator, the accuracy range of items evaluated on the vectorial information. However, these evaluation indicators are still generic, i.e, evaluation is facilitated by the spatial operator use of quality assessment, when the query is simple to evaluate. But evaluation becomes more complex when greater detail is requested on specific evaluation indicators. In general, evaluation is described through a generation report. This report specifies the evaluation parameter used within the type of spatial operator selected. Having evaluated a vectorial information file, the internal algorithms, that processed the requested evaluation type, throw a statistic percentage, describing total elements that make up the file and differences found. And also shows the accuracy percentage results from the analysis performed.

Now, in this research context, there have emerged a number of ideas for adapting interoperable communication between spatial technologies used, and other mechanisms adaptation, may go beyond in vectorial assessment quality, fortifying

our methodology. However, before delving into improving the methodology, it intends to generate a national committee composed of specialists in the field. Working at the same time on strengthening and establish the spatial data infrastructure on the country, trying to get back the adaptation regulations, for a constant process of quality evaluating in vectorial mapping production.

ACKNOWLEDGEMENTS

REFERENCES

Allouche, M. K. et B. Moulin (2001), "Reconnaissance de patterns par réseaux de neurones: application à la généralisation cartographique", *Revue Internationale de Géomatique*, vol. 11, no. 2, pp. 251-279.

Ariza Lopez, F. J. (2002), *Calidad en la producción cartográfica*, Ra-Ma, Jaen, España.

Ariza Lopez, F. J. (2004), *Casos prácticos de calidad en la produccion cartográfica*, Jaén, Universidad de Jaen, España.

Bartoschek, T., M. Painho, R. Henriques and C. A. C. Peixoto M. (2006), "RENalyzer: a tool to facilitate the spatial accuracy assessment of digital cartography", *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Lisbon, Portugal, pp. 379-385.

Bédard, Y., E. Bernier et R. Devillers (2002), "La métastructure VUEL et la gestion des représentations multiples", Rus, A. (ed.), *Généralisation et Représentation Multiple*, Paris, Hermes science publications, pp. 149-162.

Bernier, E. (2002), *Utilisation de la Représentation Multiple comme Support à la Génération de Vues de Bases de Données Géospatiales dans un Contexte SOLAP*, Departement des sciences géomatiques, Québec, Laval. M.Sc.: 89.

Cárdenas A. (2004), *Utilisation des Patrons Géométriques comme Support à la Généralisation Automatique*, Département des sciences géomatiques, Québec, Laval. M.Sc.: 110.

Devillers, R. and H. Goodchild (2010), *Spatial data quality: from process to decisions*, London, New York, CRC Press Taylor & Francis Group.

Devillers, R. e R. Jeansoulin´s book (2005), "Qualité et incertitude : présentation du problème. Introduction", *Qualité de l'information géographique*, H. S. Publications, Paris, Lavoisier: 343.

Devogele, T., T. Badard et T. Libourel (2002), "La problématique de la représentation multiple", Ruas, A. (ed.), *Généralisation et Représentation Multiple*, H. S. publications, Paris, pp. 55-74.

Gago Afonso, A. J., Ferreira Coelho Dias, R. A. and A. C. Costa (2006), "IGeoE: Positional quality control in the 1/25000 cartography", *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Lisbon, Portugal, pp. 835-839.

Goodchild, M. F. (1995), "Sharing Imperfect data", in Onsrud, H. J. and G. Rushton (eds.), *Sharing Geographic Information*, New Brunswick, Rutgers University Press, pp. 413-425.

Gui, D., G. Li, Ch. Li and Ch. Zhang (2008), "Quality check in urban and rural cadastral spatial data updating", *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, P. R. China, pp. 65-70.

Jabeur N., A. (2006), *Multi-agent system for on-the-fly web map generation and spatial conflict resolution*, Departement des sciences informatiques/Géomatique, Québec, Laval, Ph.D.

Jobst, M. and A. Twaroch Florian (2006), *An evaluation method for determining map-quality*, Institute of Geoinformation and Cartography, Vienna, Austria, pp. 293-304.

Kumi-Boateng, B. and I. Yakubu (2010), "Assessing the quality of spatial data", *European Journal of Scientific Research*, vol. 43, no. 4, pp. 507-515.

McMaster, R. (1991), "Conceptual framework for geographical knowledge", *Longman Scientific & Technical*, New York NY(Wiley), pp. 21-39.

Müller J. C., R. Weibel, J. P. Lagrange and F. Salgé (1995), *Generalization: state of the art and issues. GIS and Generalization: Methodology and Practice*, Taylor & Francis, Bristol.

OGC (2011), Open Geospatial Consortium, Welcome to the OGC Website, O. S. A. Specifications.

Pavičić, S., M. Rapaić and S. Lemajić (2004), *Topographic Data Production as Basis for NSDI - Croatian Example*, FIG Working Week 2004, Athens, Greece.

Pouliot, J. (2002), "Intégration des données spatiales, Concepts et Practiques", *Cours á option du programme de Maitrise en sciences géomatiques*, Québec, Qc. Canadá, GMT-66342.

Rigaux P. (1994), "La représentation multiple dans les systèmes d'information géographique", *Revue Internationale de Géomatique*, vol. 4, no. 2, pp. 137-164.

Sabo, M. N. (2007), *Intégration des algorithmes de généralisation et des patrons géométriques pour la création des objets auto-généralisants (sgo) afin d'améliorer la généralisation cartographique à la volée*, Faculté de Foresterie et Géomatique, Quebec, Lavla. Ph.D.

Sarmento, P., H. Carrão and M. Caetano (2008), "A fuzzy synthetic evaluation approach for land cover cartography accuracy assessment", *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, P. R. China, ACADEMIC, World Academic Union, pp. 348-355.

Shi, W. (2008), "From uncertainty description to spatial data quality control", *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, P. R. China, ACADEMIC, World Academic Union, pp. 412-417.

Stehman, S. V. (2008), "Sampling designs for assessing map accuracy", *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, P. R. China, June 25-27, ACADEMIC, World Academic Union, pp. 8-15.

Vangenot, C. (1998), "Représentation multi-résolutions, concepts pour la description des bases de données avec multi-représentations", *Revue Internationale de Géomatique*, vol. 8, no. 1-2, pp. 121-147.

Weibel, R. and G. Dutton (1999), "Generalising spatial data and dealing with multiple representations", in Longley, P. A., M. F. Goodchild, D. J. Maguire and D. W. Rhind (eds.), *Geographic Information Systems-Principles and Technical Issues*, John Wiley & Son, pp. 125-155.

Wu D., H Hu, X. M. Yang, Y. D. Zheng and L. H. Zhang (2010), "Digital chart cartography: error and quality control", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, Part II, pp. 255-260.