RESEARCH ON SPATIAL CONSISTENCY CHECKING ACCORDING TO NATIONAL TECHNICAL REGULATION ON BASIC GEOGRAPHIC INFORMATION

Quoc Yen Phan[1,*], Anh Tuan Tran[2], Thi Xuan Hong Bui[2]

[1]*Le Quy Don Technical University, Hanoi, Vietnam*

[2]*Department of Survey, Mapping and Geographic Information Viet, Hanoi, Vietnam*

Abstract

Spatial consistency checking identifies logical inconsistencies according to pre-specified spatial rules, including data omission or commission, data overlap, and spatial connection. This research experimentally checks the consistency of these components according to national technical regulations on basic geographic information and the construction of checking rules based on GIS analysis methods to determine the relationship between geometric objects within a layer and between two different layers. In experiments using the ArcGIS software suite on some typical feature classes of the transportation group of the uncalibrated and uncorrected geographic database of the Bien Hoa town area at a scale of 1:5,000, it was shown that: the largest percentage of errors is the relationship between the traffic bridge (CauGiaoThong) layer and the road centerline (DoanTimDuongBo), 14.22%; next is the error of duplicate data of the road network node feature class (NutMangDuongBo), at 12.11%; then the relationship between the road surface and the road boundary, at 3.52; and the error of dangles of the road centerline, at 0.97%.

Keywords: *GIS; geospatial quality data; spatial consistency; quality check; spatial relationship rules.*

1. Introduction

Quality is the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs. Spatial data quality is all the technical criteria, features, and characteristics that meet specific spatial rules [1]. With the increasing demand for geospatial information, generating good-quality data is becoming increasingly important. At the same time, competition in the market is increasing, pushing data producers to improve the quality of their data [2]. Data quality information must be stored in metadata files to be made available to users when needed [3], and it is key for data reuse, especially in evaluating data suitability for the field before applying it [4, 5].

* Email: phanquocyen@lqdtu.edu.vn

DOI: 10.56651/lqdtu.jst.v6.n02.742.sce

The concept of logical consistency was originally introduced for the purpose of database integration, relating to the existence of logical contradictions within a set of data [6, 7]. Those constraints align the data with the chosen structure and provide opportunities to optimize data storage and access. This concept was later used by the geospatial community to resolve various types of conflicts that arose during data input and analysis in GIS. Logical consistency is highly correlated with positional errors. Because many types of location errors in GIS projects can have adverse effects on the analysis and use of information [6].

There have been a number of spatial data quality classifications in the world since the 1980s. In most research and international standards, five main elements describe geospatial data quality: completeness, consistency, temporal accuracy, thematic accuracy, and position accuracy [7-10]. Based on the ISO 19157:2013 standard, logical consistency can be defined as "the degree of compliance with logical rules of data structure, distribution, and relationships". The four main sub-elements considered to ensure logical consistency in this standard are conceptual consistency, domain consistency, format consistency, and topological consistency wherein topology consistency checks the correctness of the encoded topology features [11]. Thus, spatial consistency is considered topological consistency. It includes many different rules, such as overshoots or undershoots, non-closed areas, lines with dangles, self-intersect lines, self-overlapping lines, polygon overlap/gap, etc.

In Vietnam, product users and scientists in the field of GIS do not pay much attention to the quality of geographic databases. However, standards and regulations have been updated to international standards. In 2020, the Ministry of Natural Resources and Environment (MONRE, Vietnam) issued the national technical regulation on basic geographic information, QCVN 42:2020/BTNMT [2], replacing standard 42:2012/BTNMT [12]. Based on that standard and the law on surveying and mapping, the MONRE has also developed national technical regulations on national fundamental geographic databases at large scales [7], medium scales [11], and small scales [8]. These are the bases for controlling the quality of geographic data in the best and most effective way.

According to standards, data quality checking solutions are manual, automatic, and semi-automatic checking [7]. GIS data is increasing in large quantities, and checking data quality with hundreds of data layers and dozens of quality parameters makes manual methods difficult, time-consuming, and costly. With the rapid development of GIS technology, users can use spatial analysis functions in desktop GIS software to check data quality. In addition, there are specialized tools to check data quality, such as ArcGIS Data Reviewer, SQUAD for QGIS, Check for Topological Errors, Check Validity, DSG Tool, and FME [6, 10, 12]. Among them, ArcGIS Data Reviewer combined with ArcGIS is the most commonly used toolkit today [8, 12, 13]. The research focuses on applying GIS analysis techniques to check spatial consistency within a feature class and between different feature classes, meeting national standard requirements.

2. Methodology

2.1. Data quality checking process according to Vietnamese standards

According to QCVN 42:2020/BTNMT standard, implementing the process of assessing the quality of basic geographic information data has five steps (Fig. 1) [3].

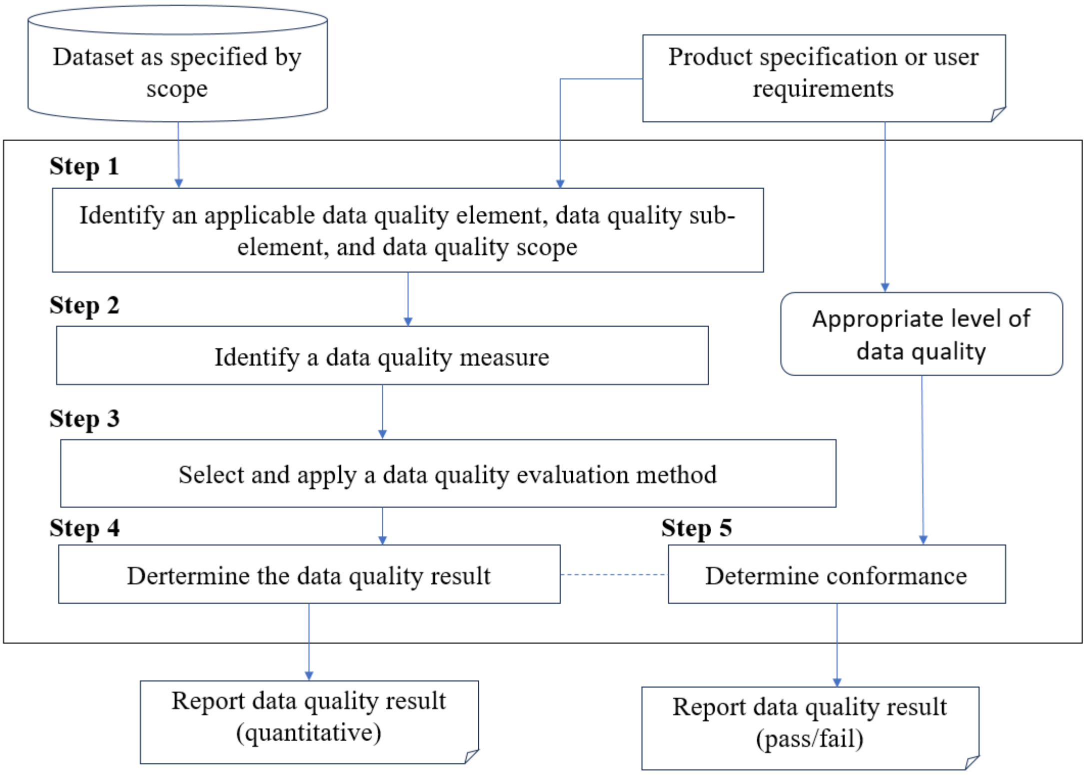

Fig. 1. Data quality checking process according to Vietnamese standard QCVN 42:2020 [3].

Step 1. Identify the quality elements and scope of application for each element: The main quality element is the degree of conformity of the data with the data structure model (logical consistency); the sub-elements are spatial relationships (topology consistency). The spatial scope applicable to this element is the entire area of the data set. Limits on data class objects and spatial relationships according to the data quality regulations of each type of data scale.

Step 2. Determine data quality measurements: There are some basic measurements such as error identification (highlighting and displaying error objects), counting the number of error elements, and calculating the percentage error. These are basic measurements to help users identify errors to fix and evaluate whether the data set passes or not.

Step 3. Identify and apply data quality assessment methods: For spatial relationships, determining them manually with common observation is very difficult, time-consuming, and costly. Taking advantage of the power of GIS techniques to determine using automatic methods will bring high efficiency.

Step 4. Analyze checking results: Analyze data quality assessment results, specify error cases, and write a report.

Step 5. Conclusion about data quality: Evaluated at a pass or fail level based on a comparison of checking results with the requirements for each assessed quality element. In the case of quality in terms of spatial consistency, the data set requirement is that there are no errors in the checking results to meet the requirements.

2.2. Method for checking topology consistency using GIS

2.2.1. Spatial relationship operators in GIS

To perform spatial quality checks on data, the use of spatial operators is necessary. There are 6 common spatial operators: *Equals operator*, interiors are identical, and the geometry types are the same; *Intersects operator*, a feature shares any portion of its geometry with the other feature. If a target feature contains, is within, crosses, touches, or overlaps a join feature, it intersects; *Contains operator*, a feature is within if the intersection of the interiors is nonempty and the target feature is a subset of the join features. If there is no intersecting interior, the join feature is not within the target feature; *Within the operator*, a target feature is within a join feature if it is completely inside the join feature. A feature is within if the intersection of the interiors is nonempty and the target feature is a subset of the join features. If there is no intersecting interior, the join feature is not within the target feature; *Crosses operator*, two polylines cross if their intersection contains only points, and at least one of the points of intersection is internal to both polylines. A polyline and polygon cross if a connected part of the polyline is partly inside and partly outside the polygon. A polyline and polygon cross if they share a polyline in common on the interior of the polygon, which is not equal to the entire polyline. The target and join features must be either polylines or polygons; *Touches operator*, a target feature touches a join feature if they have an intersecting vertex, but the features do not overlap; and overlaps, Touches, a target feature touches a join feature if they have an intersecting vertex, but the features do not overlap [14].

The following table lists the supported spatial relationships for defining the spatial relationship between two geometric objects. Geometric object symbols are point P (Point), Polyline is C (Curve) and Polygon is S (Surface).

Table 1. Spatial operators support relationships between two geometric objects [14]

	Equals	**Intersects**	**Contains**	**Within**	**Crosses**	**Touches**	**Overlaps**
P-P	✓	✓	✓	✓			
P-C		✓		✓		✓	
P-S		✓		✓			
C-P		✓	✓			✓	
C-C	✓	✓	✓	✓	✓	✓	✓
C-S		✓		✓	✓	✓	
S-P		✓	✓			✓	
S-C		✓	✓		✓	✓	
S-S	✓	✓	✓	✓		✓	✓

2.2.2. Apply GIS analysis techniques in checking spatial consistency

For automatic testing of the entire dataset and different spatial quality elements. Perform the 5 process steps above with the support of ArcGIS software (Fig. 2) [15].

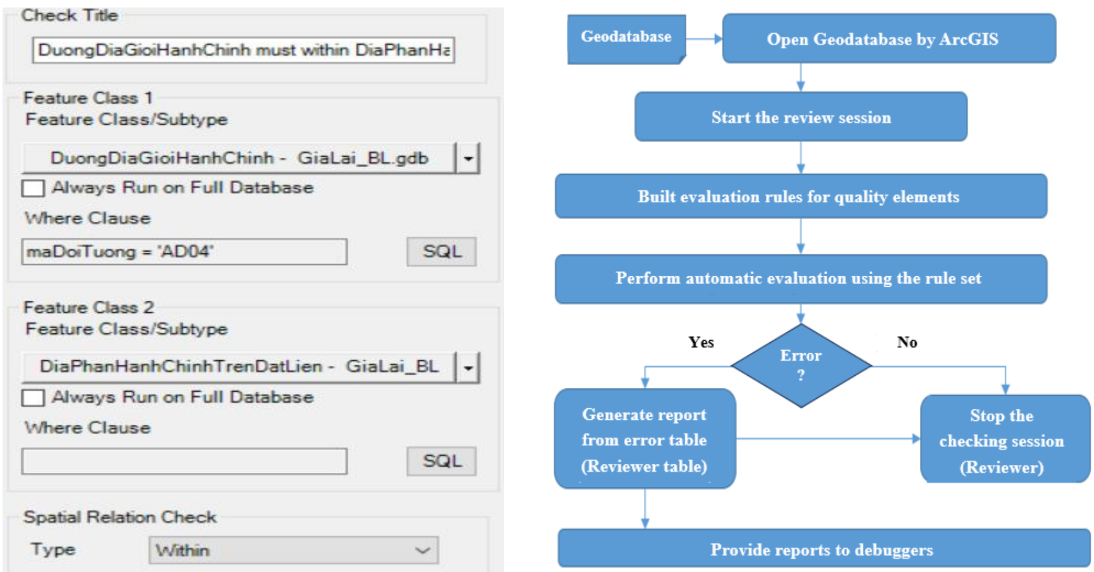

Fig. 2. Automated checking process using ArcGIS Data Reviewer [15].

Among the implementation steps, the most important step is to create a checking rule set using the Reviewer Batch Job Manager tool. This rule is modeled from the regulations on the content model, structure, and data collection according to the provisions of QCVN 42:2020/BTNMT and the corresponding data scale technical regulations. To build a quality-checking rule, it is necessary to define a number of parameters:

- Parameters of input data object classes need to be checked. Testing can be performed for entire classes or limited data sets using SQL statements;

- Spatial operator parameters. These operators are listed in Table 1, and users can choose the operator type to suit each different spatial law (Fig. 2).

- Some other parameters about test criticality, tolerance for identifying defect objects, etc. These options depend on each specific test element.

Finally, perform automatic error identification with software, identify and classify error sets to calculate error percentages, and prepare reports. The formula to calculate the percentage of error elements is as follows:

$$Error\ percentage = \frac{Total\ number\ of\ error\ elements\ of\ the\ object\ class}{Total\ number\ of\ test\ elements\ of\ the\ object\ class} * 100\% \, [3]$$

3. Result and discussion

3.1. Data and experimental area

The experimental data is a group of traffic data layers of the basic uncalibrated and uncorrected geographic information data set at a scale of 1:5,000 of Bien Hoa town, including wards and communes: Long Binh, Tan Hoa, Tan Bien, Ho Nai, An Binh, Tan Hiep, Tam Hoa, etc. The data is built according to the model and content structure of QCVN 42:2020/BTNMT and regulations on the geographic database at scale of 1:2,000, 1:5,000 according to Circular No. 23/2019/TT-BTNMT, including 7 data groups, equivalent to 7 Feature Dataset. In each Feature Dataset, there are data classes called Feature Classes. Experimental research on some typical Feature Classes of the GiaoThong group, such as Road network node (NutMangDuongBo) with 16512 objects, the centerline section (DoanTimDuongBo) with 8265 objects, traffic bridge (CauGiaoThong) with 204 objects, road surface (MatDuongBo) with 2668 objects, road boundary (RanhGioiDuongBo) with 18163 objects (Fig. 3). These are typical data layers for spatial analysis to determine spatial data quality consistency.

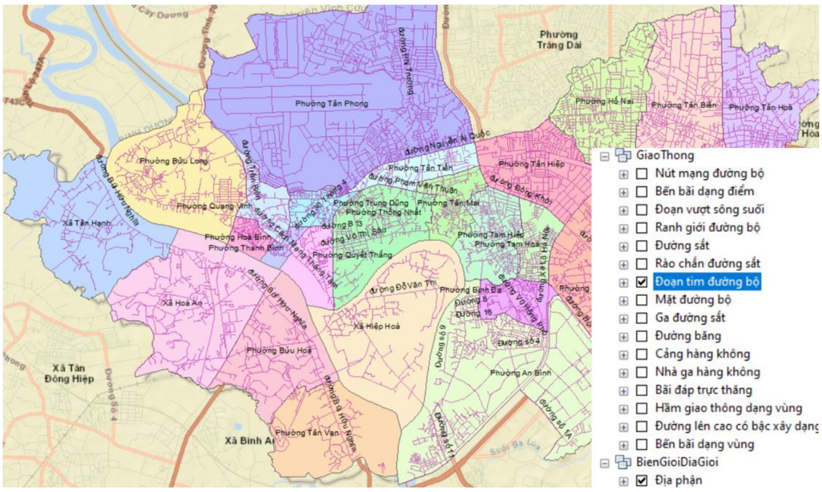

Fig. 3. Geodatabase data and experimental area.

3.2. Experimental results

3.2.1. Develop a set of rules to check spatial consistency for traffic data groups

To perform the inspection content, the identification of objects and quality measurements are performed first. The common quality measure performed for the entire data set is "determining the number of duplicate objects". This measurement checks the omission or commission of data, specifically spatial data, to which the rule is applied "Duplicate Geometry". In addition, object classes are determined according to each quality element according to standards as shown in Table 2.

Table 2. Quality measures and spatial laws applied to each data layer

Feature Class	Quality measurement	Relationship object	Applicable Ruler
Centerline section (DoanTimDuongBo)	Determine the self-overlap error		Must Not Self-Overlap
	Determine the self-intersection error		Must Not Self-Intersect
	Determine the dangles error		Must Not Have Dangles
	Determine that the beginning and end of the arc do not coincide with the location of the point	Road network node (NutMangDuongBo)	Endpoint Must Be Covered By NutMangDuongBo
Road network node (NutMangDuongBo)	Determine the zero point located at the beginning and end of the arc	Centerline section (DoanTimDuongBo)	Point Must Be Covered By Line, Must Be Covered By Endpoint of DoanTimDuongBo

Feature Class	Quality measurement	Relationship object	Applicable Ruler
Traffic bridge (CauGiaoThong)	Determines an arc that does not overlap with an arc of another object type	Centerline section (DoanTimDuongBo)	DoanTimDuongBo Must Overlap CauGiaoThong
Road surface (MatDuongBo)	Identify small area errors		Area Check < 25 m^2
	Identify overlap/gap errors of polygon		Polygon Overlap/Gap
	Determine self-intersection surface errors		MultiPart Plolygon Check
	Determine the number of errors in the relationship between the surface and the boundary	Road boundary (RanhGioiDuongBo)	RanhGioiDuongBo Must be Covered by Boundary of MatDuongBo

3.2.2. Results of checking the quality of spatial data consistency

After determining the quality elements and quality measurements, build a set of rules and evaluate quality automatically using the software. Research and apply copyrighted software from ESRI, ArcGIS 10.8.2 software suite, including ArcMap, ArCatalog, and the Data Reviewer tool. Check the data quality according to the rulers, and the results are as shown in the Table 3.

Quality elements not included in Table 3 are those with 0% errors, meaning they meet the requirements according to set standards.

Duplicate data error: This is an error that often occurs due to a duplicate data error. This is an error that often occurs due to the data editing process (Fig. 4). Road RanhGioiDuongBo data layer with 2000 objects (12.1%) and NutMangDuongBo with 172 objects (0.95%). Through detailed inspection of the NutMangDuongBo layer, only a few network nodes do not have duplicate errors; the number of errors is the highest, causing data redundancy and memory waste.

The error that the beginning and end of the arc do not coincide with the position of the point: The error indicates two points with the same code and the same location. This is only a single point, caused by the error of duplicating the NutMangDuongBo data above. Zoom in on the error location to check that this point is not located at the intersection (Fig. 5).

Table 3. Results spatial consistency errors

No.	Feature Class	Error	Total errors (object)	Total object Class	Error percentage (%)
1	DoanTimDuongBo	Self-overlaps of the arc	0	8265	0
2	DoanTimDuongBo	Self-intersection of the arc	0	8265	0
3	DoanTimDuongBo	Dangles of the arc	80	8265	0.97
4	DoanTimDuongBo, NutMangDuongBo	The beginning and end of the arc do not coincide with the position of the point	1	8265	0.01
5	NutMangDuongBo	Duplicate objects	2000	8265	12.11
6	CauGiaoThong, DoanTimDuongBo	Do not overlap with arcs of other object types	29	16512	14.22
7	MatDuongBo	Small area error	19	204	0.71
8	MatDuongBo	Polygon overlap/gap	12	2668	0.45
9	MatDuongBo	Self-intersecting polygon	0	2668	0.00
10	MatDuongBo, RanhGioiDuongBo	Relationship between surface and boundary	94	2668	3.52
11	RanhGioiDuongBo	Duplicate objects	172	2668	0.95

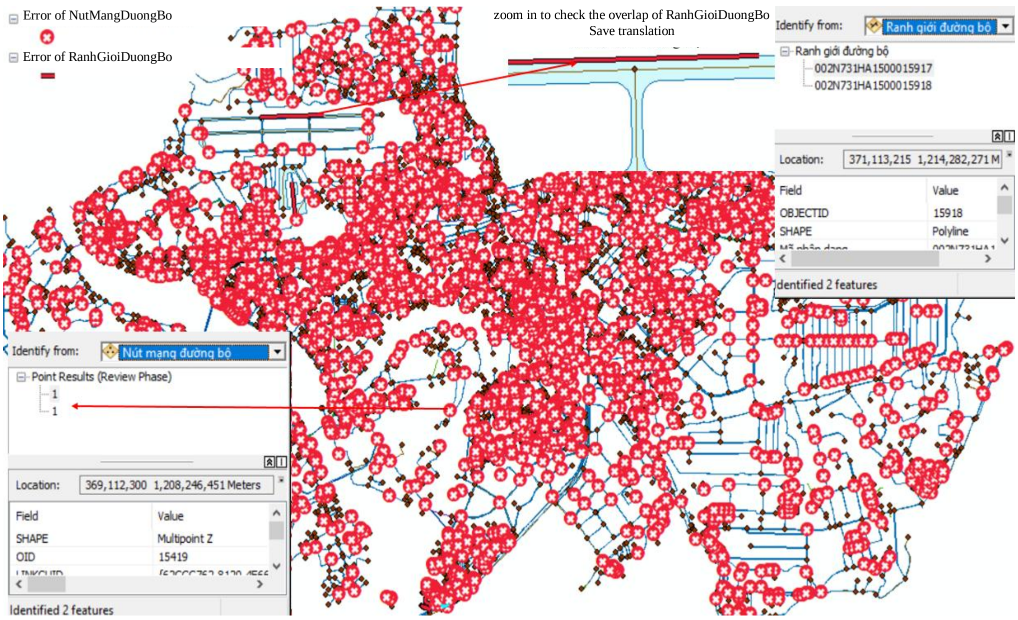

Fig. 4. Duplicate data error with NutMangDuongBo and RanhGioiDuongBo.

Dangles Error of DoanTimDuongBo: A total of 80 errors in dangles violate the law of space connection; they can be under (ID 3650), exceeded (ID 3676), or deviated left or right compared to the center of the connection line (ID 2551) (Fig. 6). Not reaching or exceeding can cause the wrong object to be snapped during the data normalization process; for example, object 4701 is caused by the user mistakenly snapping a point at the edge of the road, not the center of the road. These dangle errors will cause seriously incorrect results in traffic network analysis problems to support decision-making in route routing.

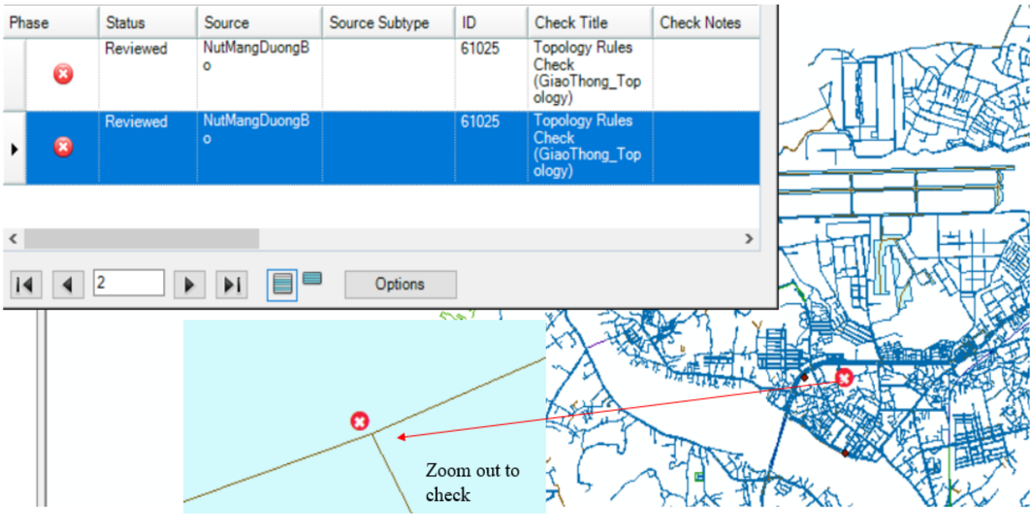

Fig. 5. The beginning and end of the arc do not coincide with the position of the point.

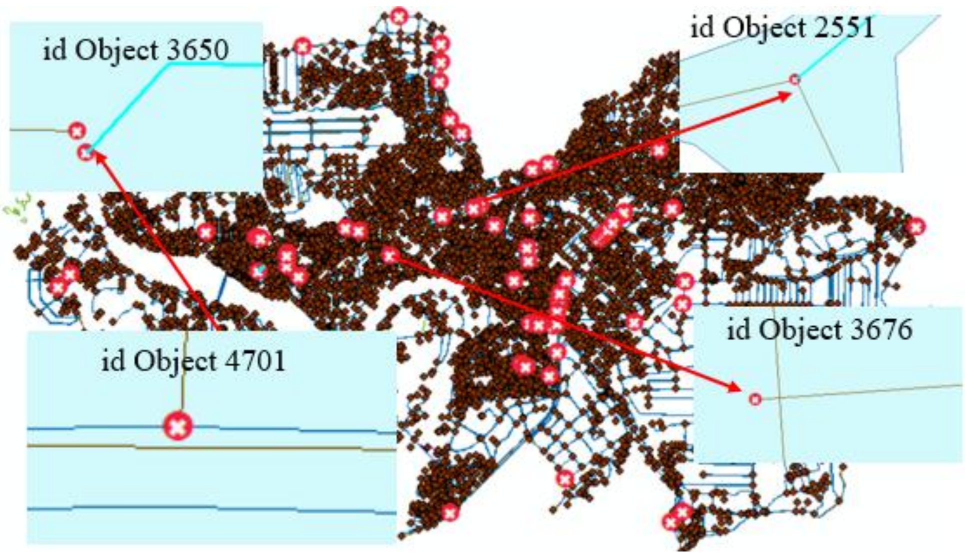

Fig. 6. Dangles error of DoanTimDuongBo.

DoanTimDuongBo must overlap with CauGiaoThong: This is a ruler of the spatial relationship between the DoanTimDuongBo and the CauGiaoThong. There were 29 errors in total, accounting for the second-highest error rate at 14.22%. These errors are due to the construction and standardization of internal data. Using spatial relationships to detect errors and zooming in on the object will show the error, which will be very difficult to detect with the common observation (Fig. 7).

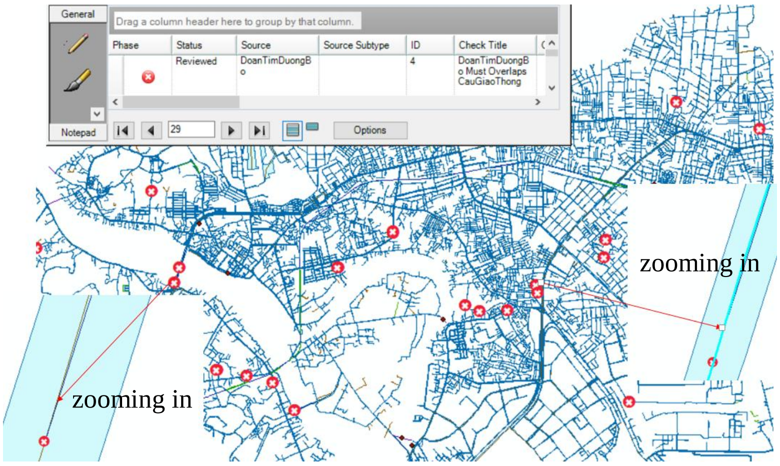

Fig. 7. The error of the DoanTimDuongBo must overlap with CauVuotSong.

MatDuongBo and RanhGioiDuongBo: This is a very important data layer for supporting traffic planning. If the boundary and road surface do not match, the spatial analysis results will be affected. The checking showed that there were 94 subjects (3.52%) with errors marked in red. Zooming in on some objects to check shows that the boundary line cuts diagonally into the road surface area (Fig. 8).

Fig. 8. Overlay error of BeMatDuongBo and RanhGioiDuong.

4. Conclusion

The research applied GIS techniques and experimentally checked the quality components of the spatial logical consistency of the geographic database set at a scale of 1:5,000 in Bien Hoa town. Testing standards include QCVN 42:2020/BTNMT and technical regulations on the model and content structure of the national technical regulation on the national fundamental geographic database at scales of 1:2,000 and 1:5,000. Experiments show that applying spatial relationship analysis techniques in GIS to automatically check spatial quality brings high efficiency. The method can quickly identify errors in spatial consistency according to defined rules such as relationships between point-line, line-polygon, point, and polygon.

The research was conducted on a small part of the traffic data group of the full database, including seven groups according to regulations. The experimental data is typical and fully representative of spatial analysis problems. There are many causes of inconsistency, such as data collection methods, data formats, models, standards, transformations, integration, and updates. The total number of objects and comparisons is up to hundreds of thousands of different comparison operations. Using manual inspection methods with the naked eye is very time-consuming and costly. Therefore, using an automated process to check the quality of geographic background data is necessary. In particular, applying GIS analysis techniques to spatial data is very effective and accurate.

Acknowledgments

This research is supported by a research project at the Ministry of Natural Resources and Environment, Vietnam: “Research technology solutions to check the quality of the basic national geographic database”, code TNMT.2023.04.08.

References

[1] Aleš Veršič, *Spatial data quality control process based on ISO 19113 and ISO 19114*, 24th International Cartographic Conference, Santiago, Chile, 2009.

[2] R. M. Henriksson and T. Kauppinen, "An ontology-driven approach for spatial data quality evaluation", in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 34, 2007.

[3] Rodolphe Devillers et al., "Thirty years of research on spatial data quality: achievements, failures, and opportunities", *Transactions in GIS*, Vol. 14(4), pp. 387-400, 2010.

[4] Joanne Nightingale and et al., "Quality assurance framework development based on six new ECV data products to enhance user confidence for climate applications", *Remote Sensing*, Vol. 10(8), 2018, 1254.

[5] Ge Peng et al., "International community guidelines for sharing and reusing quality information of individual earth science datasets", Open Science Framework, 2021. [Online]. Available: https://osf.io/xsu4p/download. [Accessed Sept. 25, 2023].

[6] Michael F. Goodchild, "Stepping over the line: technological constraints and the new cartography", *The American Cartographer*, Vol. 15(3), pp. 311-319, 1988.

[7] Mike Hendricks, Jen Athey and Amy Macpherson, *QA/QC of GeMS DataCase Study: The Alaska DGGS Geologic Mapping System*, U.S. Department of Commerce NOAA, 2022.

[8] *Geographic information - Quality principles (ISO 19113)*, SIST EN ISO 19113:2005, 2002.

[9] *Geographic information - Quality evaluation procedures (ISO 19114)*, SIST EN ISO 19114:2005, 2003.

[10] ESRI, Spatial relationships in ArcGIS GeoAnalytics Server, arcgis.com, 2023. [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/big-data-analytics/spatial-relationships-with-big-data.htm#. [Accessed 10.2023].

[11] Wolfgang Kainz, "Logical consistency", *Elements of spatial data quality*, Vol. 202, pp. 109-137, 1995.

[12] Peyman Hashemi and Rahim Ali Abbaspour, "Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept", *OpenStreetMap in GIScience: experiences, research, and applications*, 2015, pp. 19-36.

[13] ESRI, *Get Started with ArcGIS Data Reviewer*, USA, 2019.

[14] Peter A. Burrough, Rachael A. McDonnell and Christopher D. Lloyd, *Principles of geographical information systems*, Oxford University Press, USA, 2015.

[15] Z. Y. Ahmed, "Quality Control and Quality Assurance in GIS Database and Applications", GIS Consultant - Ain Shams University, 2020.

NGHIÊN CỨU KIỂM TRA TÍNH THỐNG NHẤT KHÔNG GIAN THEO QUY CHUẨN KỸ THUẬT CƠ SỞ DỮ LIỆU NỀN ĐỊA LÝ QUỐC GIA

Phan Quốc Yên[1], Trần Anh Tuấn[2], Bùi Thị Xuân Hồng[2]

[1]*Trường Đại học Kỹ thuật Lê Quý Đôn, Hà Nội, Việt Nam*

[2]*Cục Đo đạc, Bản đồ và Thông tin địa lý Việt Nam, Hà Nội, Việt Nam*

Tóm tắt: Kiểm tra tính nhất quán không gian là kiểm tra sự không thống nhất logic theo quy tắc không gian đã quy định trước, bao gồm dư thừa, chồng đè dữ liệu, kết nối… Bài báo này thực nghiệm kiểm tra tính thống nhất của thành phần này theo quy chuẩn kỹ thuật về mô hình cấu trúc, nội dung cơ sở dữ liệu nền địa lý quốc gia tỉ lệ 1:5.000 khu vực thị xã Biên Hòa. Sử dụng phương pháp phân tích GIS để xác định mối quan hệ của các đối tượng hình học trong một lớp và giữa hai lớp khác nhau. Nghiên cứu được tiến hành sử dụng bộ phần mềm ArcGIS trên một số Feature Class điển hình của nhóm Giao Thông cho thấy rằng: Phần trăm lỗi lớn nhất là quan hệ giữa lớp cầu giao thông và tim đường bộ 14,22%; tiếp theo là lỗi về trùng lặp dữ liệu của lớp nút mạng đường bộ 12,11%; sau đó là quan hệ giữa mặt đường bộ và ranh giới đường bộ 3,52; lỗi đỉnh treo của tim đường bộ 0,97%.

Từ khóa: *GIS; chất lượng dữ liệu địa không gian; tính thống nhất không gian; kiểm tra chất lượng; các tiêu chuẩn quan hệ không gian.*

Received: 03/10/2023; Revised: 25/12/2023; Accepted for publication: 27/12/2023

❑