

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

Phan Quốc Yên¹, Trần Anh Tuấn², Bùi Thị Xuân Hồng²

⁽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

Tóm tắt

Chất lượng dữ liệu cho phép đo lường sự phù hợp của cơ sở dữ liệu vào lĩnh vực ứng dụng. Nếu chất lượng không đảm bảo, kết quả phân tích và hỗ trợ ra quyết định không chính xác, gây ra tổn thất to lớn trong nhiều hoạt động như quy hoạch, hỗ trợ các hoạt động quân sự, logistic, v.v. Kiểm tra tính nhất quán không gian xác định sự không thống nhất logic theo quy tắc không gian đã quy định trước, bao gồm dư thừa, chồng đè dữ liệu, kết nối, v.v. 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 tỷ lệ 1:5.000 khu vực thị xã Biên Hòa. Nghiên cứu thực hiện xây dựng các luật kiểm tra dựa vào toán tử không gian mối quan hệ giữa hai đối tượng hình học trong GIS. Qua thực nghiệm 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 tìm 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 tìm đường bộ 0,97%.

Từ khóa: *Geographic information, Geospatial Quality Data, Quality model, Quality Test, Quality Evaluation*

1. Giới thiệu

Chất lượng là toàn bộ các tính năng và đặc điểm của một sản phẩm hoặc dịch vụ có khả năng đáp ứng các nhu cầu nhất định. Chất lượng dữ liệu không gian là tất cả các tiêu chí về kỹ thuật, tính năng, và đặc điểm đáp ứng được quy định đặt ra [25]. Với nhu cầu sử dụng thông tin địa không gian ngày càng tăng, việc tạo ra dữ liệu có chất lượng tốt ngày càng trở nên quan trọng. Đồng thời, sự cạnh tranh trên thị trường ngày càng nhiều, thúc đẩy các nhà sản xuất dữ liệu cần phải nâng cao chất lượng dữ liệu của họ [11]. Để làm được điều này, chúng ta cần kiểm soát chất lượng dữ liệu GIS theo một quy chuẩn nhất định. Trong đó, thông tin về chất lượng dữ liệu phải được lưu trữ trong tệp siêu dữ liệu nhằm cung cấp cho người

dùng khi cần thiết [5], và nó là chìa khóa cho việc tái sử dụng dữ liệu và đặc biệt trong đánh giá dữ liệu có phù hợp với lĩnh vực trước khi áp dụng [15], [16].

Khái niệm về tính thống nhất logic ban đầu được giới thiệu cho mục đích tích hợp cơ sở dữ liệu, liên quan đến sự tồn tại mâu thuẫn logic trong một bộ dữ liệu [9], [12]. Các ràng buộc đó là các quy tắc điều chỉnh dữ liệu với cấu trúc đã chọn và cung cấp cơ hội để tối ưu hóa tốc độ lưu trữ/truy xuất. Khái niệm này sau đó được cộng đồng địa không gian sử dụng để giải quyết các loại mâu thuẫn khác nhau phát sinh trong quá trình nhập và phân tích dữ liệu trong GIS. Tính thống nhất logic có mối tương quan cao với các lỗi vị trí. Bởi vì có rất nhiều loại lỗi vị trí trong các dự án GIS có thể có tác dụng phụ trong việc phân tích và sử dụng thông tin [9].

Trên thế giới đã có một số phân loại chất lượng dữ liệu không gian từ những năm 1980. Trong hầu hết tất cả các nghiên cứu, các quy chuẩn, tiêu chuẩn quốc tế, năm yếu tố chính mô tả độ chính dữ liệu nền địa lý là tính đầy đủ, tính nhất quán, độ chính xác thời gian, độ chính xác chủ đề và độ chính xác vị trí [1], [2], [4], [8], [14]. Dựa trên tiêu chuẩn ISO 19157:2013, tính thống nhất logic có thể được định nghĩa là “mức độ tuân thủ các quy tắc logic về cấu trúc dữ liệu, phân bố và các mối quan hệ”. Bốn yếu tố phụ chính được xem xét để đảm bảo tính nhất quán logic trong tiêu chuẩn này là tính nhất quán về khái niệm, tính nhất quán về miền, tính nhất quán về định dạng và tính nhất quán về cấu trúc liên kết. Tính nhất quán về khái niệm giám sát việc tuân thủ các quy tắc của lược đồ khái niệm. Tính nhất quán của miền kiểm tra các phạm vi giá trị, phải nằm trong các miền giá trị nhất định trong khi tính nhất quán về định dạng kiểm soát định dạng được lưu trữ theo cấu trúc dữ liệu vật lý. Tính nhất quán cấu trúc liên kết kiểm tra tính chính xác của các đặc điểm cấu trúc liên kết được mã hóa [18]. Như vậy, tính thống nhất không gian được xem như tính thống nhất về cấu trúc Topo (liên kết). Nó gồm nhiều quy định khác nhau như bắt điểm quá mức (overshoots) hoặc dưới mức (undershoots), vùng không đóng, vùng tự cắt, đường tồn tại điểm treo, đường tự cắt, đường tự chồng đè, điểm giao lộ không nằm trên đường, v.v.

Tại Việt Nam, người dùng sản phẩm và các nhà khoa học về lĩnh vực GIS chưa quan tâm nhiều đến chất lượng dữ liệu nền địa lý. Tuy nhiên, các tiêu chuẩn, quy chuẩn cơ bản theo kịp được các tiêu chuẩn của quốc tế. Năm 2020, Bộ Tài nguyên và Môi trường đã có Thông tư số 06/2020/TT-BTNMT ngày 31 tháng 8 năm 2020 Quy định Quy chuẩn kỹ thuật quốc gia về chuẩn thông tin địa lý cơ sở với mã số QCVN 42:2020/BTNMT [22]. Quy chuẩn này thay thế quy chuẩn 42:2012/BTNMT [20]. Trên cơ sở quy chuẩn trên và luật đo đạc bản đồ, Bộ Tài

nguyên và Môi trường cũng đã xây dựng các quy định kỹ thuật dữ liệu nền địa lý cho các loại tỷ lệ lớn [21], trung bình [23] và tỷ lệ nhỏ [24]. Đây là các căn cứ để kiểm soát chất lượng dữ liệu nền địa lý một cách tốt và hiệu quả nhất.

Theo các quy chuẩn, hiện nay có các giải pháp kiểm tra chất lượng dữ liệu là kiểm tra thủ công, kiểm tra tự động và bán tự động [22]. Dữ liệu GIS ngày càng tăng lên với số lượng lớn, việc kiểm tra chất lượng dữ liệu với hàng trăm lớp dữ liệu và hàng chục tham số chất lượng gây ra khó khăn cho phương pháp thủ công. Phương pháp này đòi hỏi nhiều thời gian và tốn kém về chi phí. Với sự bùng nổ công nghệ thông tin, công nghệ GIS ngày càng phát triển vượt bậc. Người dùng có thể sử dụng các chức năng phân tích không gian trong các phần mềm GIS máy tính để bàn để kiểm tra chất lượng dữ liệu GIS như ArcGIS, MapInfor, FME, Microstation, QGIS, v.v. Ngoài ra, có các công cụ chuyên dụng để kiểm tra chất lượng dữ liệu như ArcGIS Data Reviewer, SQUAD for QGIS, Check for topological errors; check validity; DSG Tool, FME [13], [17], [19]. Trong đó, ArcGIS Data Reviewer là công cụ được sử dụng phổ biến nhất hiện nay [3], [10], [19].

Trong bối cảnh công nghệ 4.0, công tác chuyển đổi số đang triển khai sâu rộng ở nước ta. Cơ sở dữ liệu nền địa lý là dữ liệu nền tảng, thành phần quan trọng của chuyển đổi số quốc gia. Dữ liệu không gian có đầy đủ các thông số về chất lượng dữ liệu cần được quan tâm đúng mức. Mục tiêu của bài báo này là thực nghiệm quy trình kiểm tra tự động chất lượng dữ liệu không gian được quy định bởi QCVN 42:2020/BTNMT. Bài báo tập trung vào áp dụng các kỹ thuật GIS để phân tích mối quan hệ không gian, kiểm tra tính thống nhất logic của dữ liệu không gian trong một đối tượng và giữa các đối tượng khác nhau, đáp ứng được các yêu cầu kỹ thuật quốc gia.

2. Phương pháp

2.1. Quy trình kiểm tra chất lượng dữ liệu theo quy chuẩn Việt Nam

Theo quy chuẩn QCVN 42:2020/BTNMT, thực hiện quy trình đánh giá chất lượng dữ liệu nền địa lý có 5 bước (Hình 1) [22]. Áp dụng quy trình này để đánh giá tính thống nhất không gian theo quy chuẩn kỹ thuật của dữ liệu GIS cụ thể như sau:

Bước 1. Xác định các tiêu chí chất lượng và phạm vi áp dụng cho từng tiêu chí: Tiêu chí chính là Mức độ phù hợp của dữ liệu với mô hình cấu trúc dữ liệu (Logical consistency), tiêu chí phụ là các quan hệ không gian (thống nhất về Topo). Phạm vi không gian áp dụng cho tiêu chí này là toàn bộ khu vực của tập dữ

liệu. Giới hạn về đối tượng các lớp dữ liệu và quan hệ không gian theo quy định chất lượng dữ liệu của từng loại tỷ lệ dữ liệu.

Bước 2. Xác định phép đo chất lượng dữ liệu: Có một số phép đo cơ bản như xác định lỗi, tức là đánh dấu và hiển thị những đối tượng lỗi cho người dùng; đếm số lỗi và tính % bị lỗi. Đây là những phép đo cơ bản cho người dùng biết lỗi để sửa chữa và đánh giá xem tập dữ liệu có đạt hay không. Đối với yếu tố chất lượng về tính tuân thủ quan hệ không gian, áp dụng phép đo xác định lỗi và đếm lỗi cho các tiêu chí phụ như: Lỗi tự chồng đè của cung; lỗi cung tự cắt; lỗi đỉnh treo của cung; lỗi chồng xếp bề mặt; lỗi khoảng hở giữa các bề mặt; lỗi bề mặt tự giao không hợp lệ; lỗi cung không trùng với cung; lỗi điểm không nằm tại vị trí đầu, cuối cung; và lỗi quan hệ giữa cung và biên của bề mặt.

Bước 3. Xác định và áp dụng phương pháp đánh giá chất lượng dữ liệu: Đối với mỗi quan hệ không gian, việc xác định thủ công bằng mắt thường là rất khó khăn, tốn kém thời gian và kinh phí. Tận dụng sức mạnh các kỹ thuật GIS để xác định bằng phương pháp tự động sẽ đem lại hiệu quả cao.

Bước 4. Phân tích kết quả kiểm tra: Phân tích các kết quả đánh giá chất lượng dữ liệu, nêu cụ thể các trường hợp lỗi và viết báo cáo.

Bước 5. Kết luận về chất lượng dữ liệu: Đánh giá ở mức đạt hoặc không đạt trên cơ sở so sánh các kết quả kiểm tra với các yêu cầu cho từng tiêu chí đánh giá. Đối với trường hợp chất lượng về mỗi quan hệ không gian, yêu cầu đưa ra là phải đạt 100% tập dữ liệu không có lỗi. Có nghĩa là, còn tồn tại lớn hơn hoặc bằng 01 lỗi là tập dữ liệu không đạt.

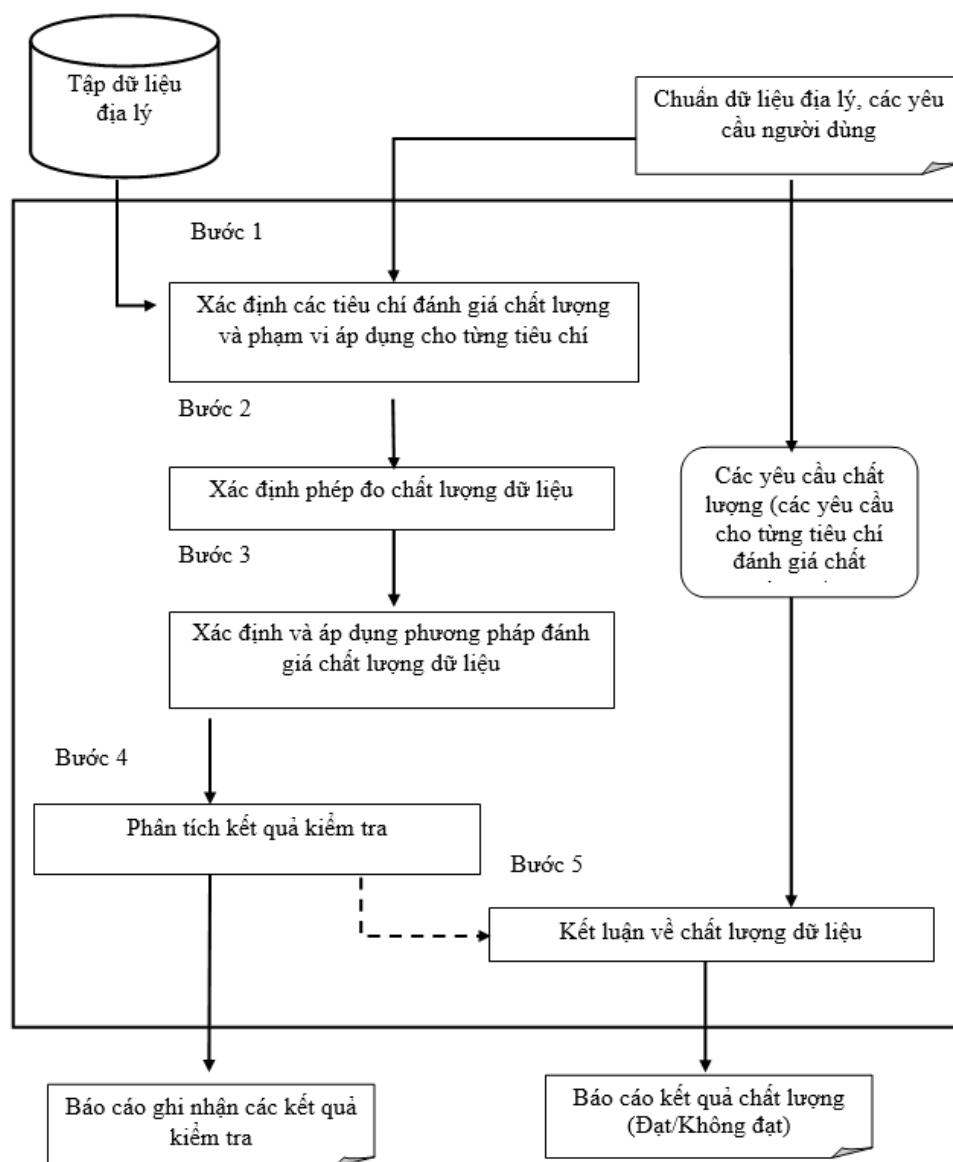


Fig 1. Quy trình kiểm tra chất lượng dữ liệu nền địa lý theo QCVN 42:2020 [22]

2.2. Phương pháp kiểm tra tính thống nhất Topo bằng công cụ GIS

2.2.1. Một số toán tử quan hệ không gian trong GIS

Để thực hiện kiểm tra chất lượng không gian của dữ liệu, sử dụng các toán tử không gian là cần thiết. Có 6 toán tử không gian phổ biến là: Equals, là phần bên trong và kiểu hình học giống hệt nhau; Intersects, là hai đối tượng có quan hệ chứa, nằm bên trong, cắt ngang, chạm hoặc chồng đè; Contains, là một trong hai đối tượng là tập con của đối tượng còn lại mà có phần bên trong giao nhau; Within, toán tử này ngược lại với toán tử Contains; Crosses, là một trường hợp con của Intersects, chỉ quy định cho mối quan hệ giữa đường với đường, đường với vùng; Touches, hai đối tượng chạm nhau là có đỉnh giao nhau nhưng các đối tượng không trùng nhau; và Overlaps quy định cho hai đối tượng hình học cùng loại và

giao điểm của chúng cũng có cùng loại hình học nhưng giao điểm khác với cả hai hình học ban đầu [7].

Bảng sau liệt kê các mối quan hệ không gian được hỗ trợ để xác định mối quan hệ không gian giữa hai đối tượng hình học. Ký hiệu đối tượng hình học là điểm P (Point), đường gấp khúc C (Curve) và vùng là S (Surface).

Bảng 1. Các toán tử không gian hỗ trợ mối quan hệ giữa hai đối tượng hình học [7]

	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. Áp dụng các kỹ thuật phân tích GIS trong kiểm tra tính nhất quán không gian

Để kiểm tra tự động cho toàn bộ tập dữ liệu và cho các yếu tố chất lượng không gian khác nhau. Thực hiện 5 bước quy trình ở trên dưới sự hỗ trợ của công cụ ArcGIS Data Reviewer (Hình 2) [6].

Trong các bước thực hiện, bước quan trọng nhất là tạo bộ luật kiểm tra bằng công cụ Reviewer Batch Job Manager. Bộ luật này chính là mô hình hóa các luật về chất lượng dữ liệu theo quy định trong QCVN 42:2020/BTNMT. Để thực xây dựng bộ luật kiểm tra chất lượng, cần xác định một số tham số:

- Tham số các lớp đối tượng dữ liệu đầu vào cần kiểm tra. Có thể thực hiện kiểm tra cho dữ liệu toàn bộ các lớp hoặc giới hạn tập dữ liệu bằng các câu lệnh SQL;

- Tham số về toán tử không gian. Các toán tử này đã được liệt kê ở Bảng 1, người dùng lựa chọn kiểu toán tử cho phù hợp với từng luật không gian khác nhau (Hình 2).

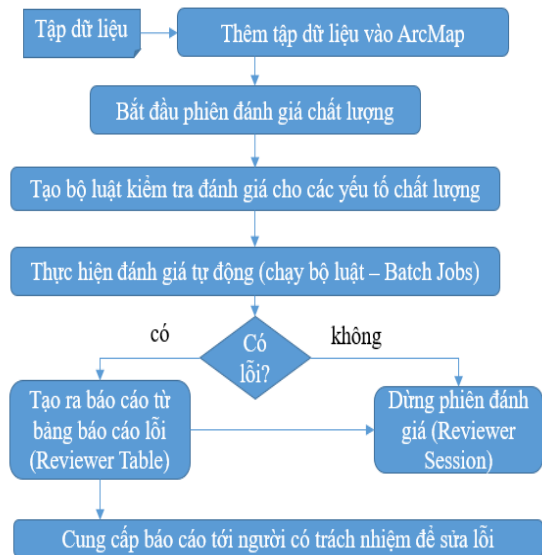
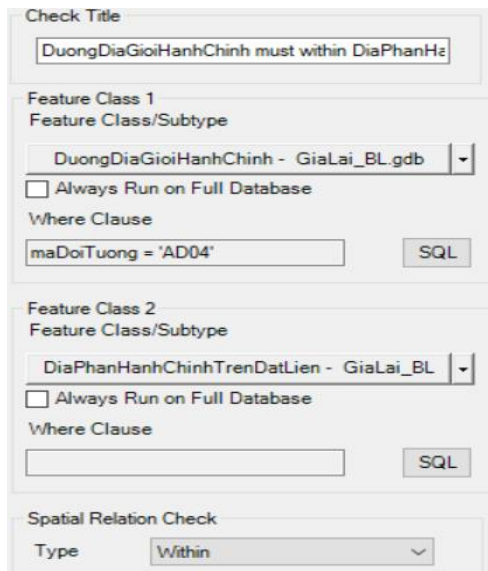


Fig 2. Quy trình kiểm tra tự động bằng ArcGIS Data Reviewer [6]

- Còn một số tham số khác về mức độ quan trọng của phép kiểm tra, dung sai để xác định đối tượng lỗi, v.v. Các tùy chọn này phụ thuộc vào từng yếu tố kiểm tra cụ thể.

Sau khi đã lập bộ luật kiểm tra, thực hiện xác định lỗi tự động bằng phần mềm, xác định và phân loại tập lỗi để tính toán phần trăm lỗi và lập báo cáo. Công thức tính phần trăm phần tử lỗi như sau:

$$\text{Phần trăm lỗi} = \frac{\text{Tổng số phần tử lỗi của lớp đối tượng}}{\text{Tổng số phần tử thuộc phạm vi kiểm tra của lớp đối tượng}} * 100\%$$

[22]

3. Thực nghiệm và đánh giá

3.1. Dữ liệu và khu vực thực nghiệm

Dữ liệu thực nghiệm là nhóm lớp dữ liệu giao thông của tập dữ liệu cơ sở nền địa lý tỷ lệ 1:5.000 thị xã Biên Hòa, gồm các phường xã: Phường Long Bình, Phường Tân Hoà, Phường Tân Biên, Phường Hố Nai, Phường An Bình, Phường Tân Hiệp, Phường Tam Hoà, v.v. Dữ liệu được xây dựng đảm bảo theo mô hình, cấu trúc nội dung quy chuẩn Thông tư số 23:2019/BTNMT và QCVN 42:2020/BTNMT, gồm 7 nhóm dữ liệu, tương đương với 7 Feature Dataset: Biên giới địa giới (BienGioiDiaGioi), cơ sở đo đạc (CoSoDoDac), Thủy văn (ThuyVan), Giao thông (GiaoThong), Dân cư (DanCu), Địa hình (DiaHinh), và Phủ bề mặt (PhuBeMat). Trong mỗi Feature Dataset có các lớp dữ liệu gọi là Feature Class. Nghiên cứu thực nghiệm trên một số Feature Class điển hình của nhóm Giao Thông như: Nút mạng đường bộ có 16512 đối tượng, đoạn tim đường bộ 8265 đối tượng, cầu giao thông 204 đối tượng, mặt đường bộ 2668 đối tượng,

ranh giới đường bộ 18163 đối tượng (Hình 3). Đây là những lớp dữ liệu điển hình cho các phân tích không gian để xác định tính nhất quán chất lượng dữ liệu không gian.

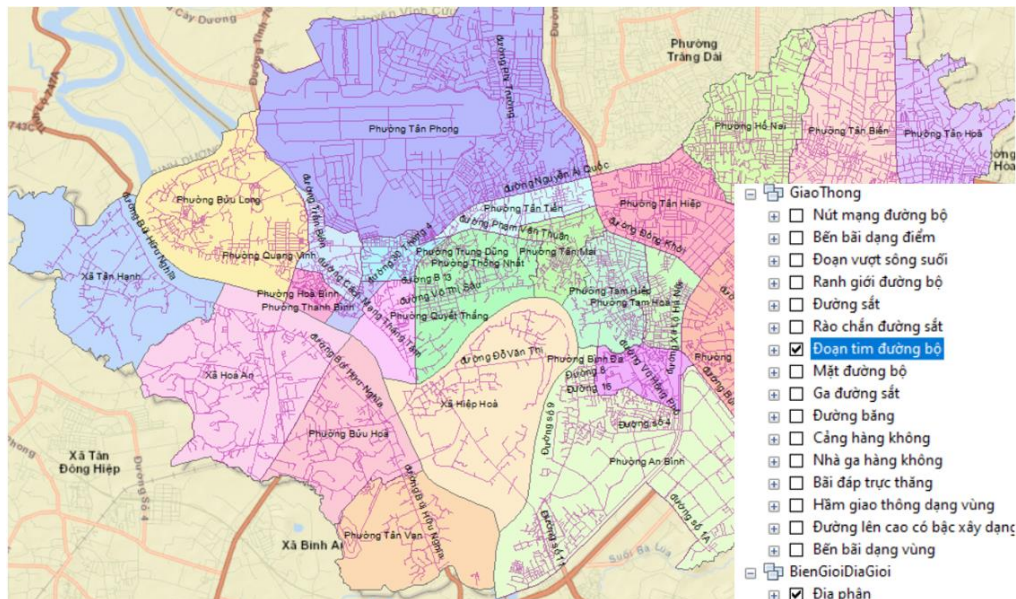


Fig 3. Dữ liệu Geodatabase và khu vực thực nghiệm

3.2. Kết quả thực nghiệm

3.2.1. Xây dựng bộ luật cho nhóm dữ liệu giao thông

Để thực hiện nội dung kiểm tra, việc xác định các đối tượng và phép đo chất lượng được thực hiện đầu tiên. Phép đo chất lượng chung được thực hiện cho toàn tập dữ liệu là “Xác định số đối tượng trùng lặp”. Phép đo này kiểm tra yếu tố dư thừa dữ liệu, mà cụ thể là dữ liệu không gian, được áp dụng quy tắc “Duplicate Geometry”. Ngoài ra, các lớp đối tượng được xác định theo từng tiêu chí riêng theo quy chuẩn như Bảng 2:

Bảng 2. Phép đo chất lượng và luật không gian áp dụng cho từng lớp dữ liệu

Kiểu đối tượng	Phép đo chất lượng	Đối tượng quan hệ	Luật áp dụng
Tim đường bộ	Xác định lỗi tự chồng đè của cung		Must Not Self-Overlap
	Xác định lỗi tự cắt nhau của cung		Must Not Self-Intersect
	Xác định lỗi đỉnh treo của cung		Must Not Have Dangles
	Xác định đầu, cuối cung không trùng với vị trí của điểm	Nút mạng đường bộ	Endpoint Must Be Covered By
Nút mạng đường	Xác định điểm không nằm tại vị trí đầu, cuối cung	Tim đường bộ	Point Must Be Covered By Line, Must Be Covered By Endpoint

bộ			Of
Cầu giao thông	Xác định cung không trùng với cung thuộc kiểu đối tượng khác	Tim đường bộ	DoanTimDuongBo Must Overlaps CauGiaoThong
Mặt đường bộ	Xác định lỗi vùng nhỏ		area Check <25m ²
	Xác định lỗi chồng xếp/lỗ hổng bề mặt		Polygon Overlap/Gap
	Xác định số lỗi bề mặt tự giao không hợp lệ		MultiPart Ploygon Check
	Xác định số lỗi quan hệ giữa bề mặt và đường biên	Ranh giới đường bộ	RanhGioiDuongBo Must be Covered by Boudary Of MatDuongBo

3.2.2. Kết quả kiểm tra chất lượng tính nhất quán dữ liệu không gian

Sau khi đã xác định các yếu tố và phép đo chất lượng, thực hiện xây dựng bộ luật và đánh giá chất lượng tự động bằng phần mềm. Nghiên cứu ứng dụng phần mềm bản quyền của hãng ESRI, bộ phần mềm ArcGIS gồm: ArcMap, ArCatalog và công cụ Data Reviewer. Thực hiện quy trình kiểm tra chất lượng dữ liệu theo bộ luật và kết quả như bảng tổng hợp sau:

Bảng 3. Tổng hợp lỗi tính nhất quán không gian

TT	Lớp đối tượng	Tên lỗi	Tổng số lỗi (đối tượng)	Phần trăm lỗi (%)
1	Tim đường bộ	Tự chồng đè của cung	0	0
2	Tim đường bộ	Tự cắt nhau của cung	0	0
3	Tim đường bộ	Đỉnh treo của cung	80	0,97
4	Tim đường bộ, Nút mạng đường bộ	Đầu, cuối cung không trùng với vị trí của điểm	1	0,01
5	Nút mạng đường bộ	Trùng lặp đối tượng	2000	12,11
6	Cầu giao thông, Tim đường bộ	Không trùng với cung thuộc kiểu đối tượng khác	29	14,22
7	Mặt đường bộ	Lỗi vùng nhỏ	19	0,71
8	Mặt đường bộ	Chồng đè/lỗ hổng bề mặt	12	0,45
9	Mặt đường bộ	Bề mặt tự giao không hợp lệ	0	0,00
10	Mặt đường bộ, Ranh giới đường bộ	Xác định số lỗi quan hệ giữa bề mặt và đường biên	94	3,52

11	Ranh giới đường bộ	Trùng lặp đối tượng	172	0,95
----	--------------------	---------------------	-----	------

Các thành phần được kiểm tra theo luật đã xây dựng mà không có trong bảng này là những thành phần có 0% lỗi, tức là đạt yêu cầu theo quy chuẩn đặt ra.

Lỗi trùng lặp dữ liệu: Đây là lỗi thường xuyên xảy ra do quá trình chỉnh sửa, biên tập dữ liệu (Fig 4). Lớp dữ liệu Ranh giới đường bộ với 2000 đối tượng (12,1%) và Nút mạng đường bộ với 172 đối tượng (0,95%). Qua kiểm tra chi tiết lớp Nút mạng đường bộ, chỉ một số ít điểm nút mạng là không bị lỗi trùng lặp, số lượng lỗi chiếm cao nhất, gây ra dư thừa dữ liệu và tốn bộ nhớ.

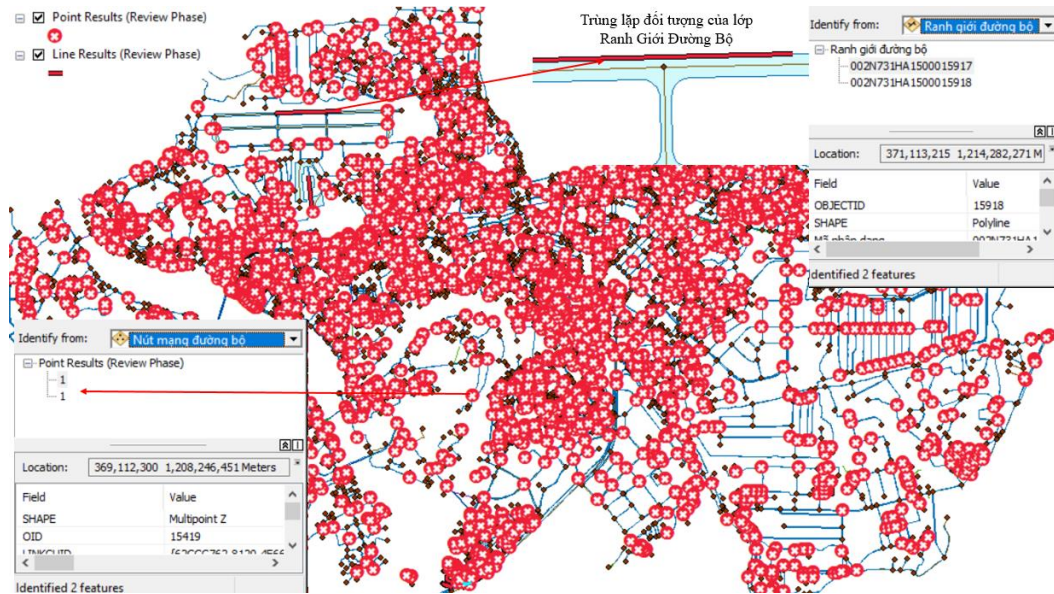
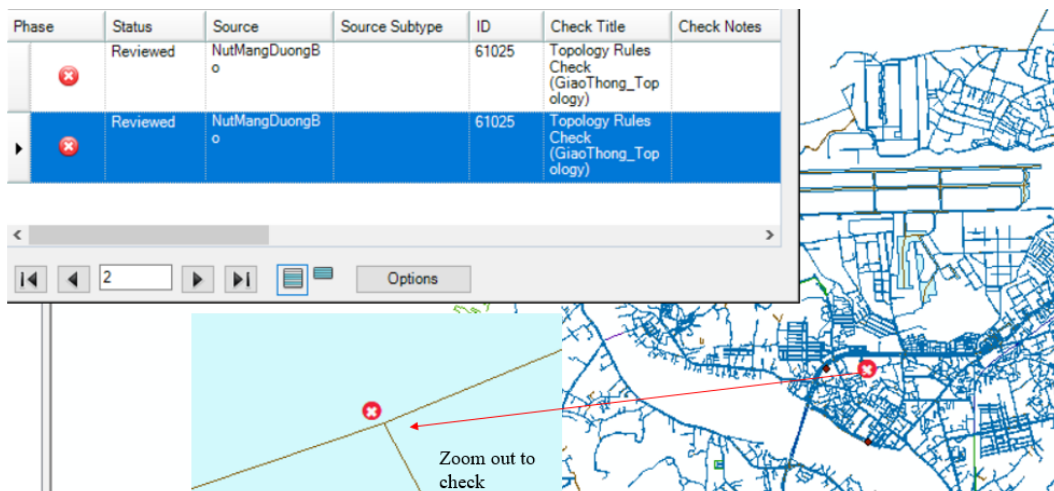


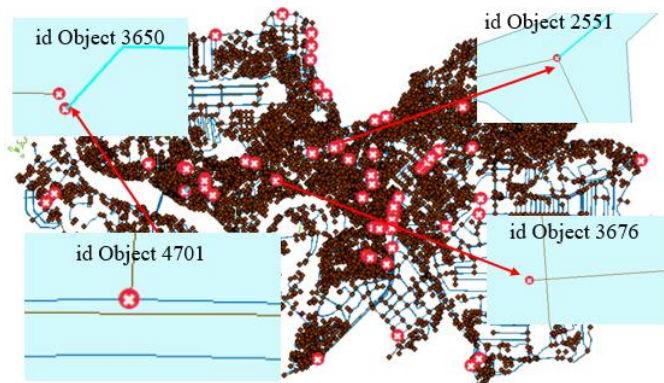
Fig 4. Lỗi trùng lặp dữ liệu với lớp Nút mạng đường bộ và Ranh giới đường bộ

Lỗi đầu, cuối không trùng với vị trí của điểm: Lỗi chỉ ra 02 điểm có mã giống nhau, cùng một vị trí. Thực ra đây chỉ là 01 điểm duy nhất, do lỗi trùng lặp dữ liệu Nút mạng đường bộ ở trên gây ra. Phóng to vị trí lỗi để kiểm tra cho thấy, điểm này không nằm trên vị trí ngã ba đường (Fig 5)



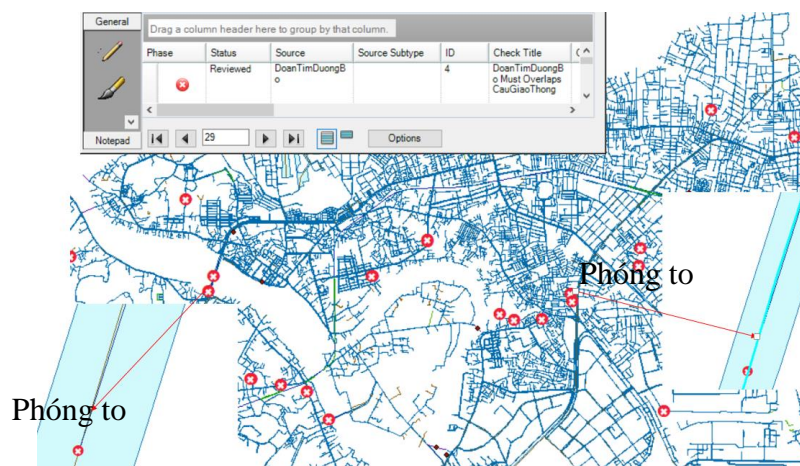
Hình 5. Lỗi đầu, cuối không trùng với vị trí của điểm

Lỗi điểm treo của lớp Đoạn tim đường bộ: Tổng cộng có 80 lỗi về điểm treo, các điểm treo này cơ bản đều vi phạm luật không gian kết nối, có thể chưa tới (ID 3650), vượt quá (ID 3676) hoặc lệch trái, lệch phải so với tim đường kết nối (ID 2551) (Hình 6). Chưa tới hoặc vượt quá có thể bắt nhầm đối tượng trong quá trình chuẩn hóa dữ liệu, ví dụ đối tượng 4701 là do người dùng bắt điểm nhầm vào biên của đường chứ không phải tim đường. Với dữ liệu lỗi về điểm treo này sẽ gây ra kết quả sai nghiêm trọng trong các bài toán phân tích mạng giao thông để hỗ trợ ra quyết định trong định tuyến đường đi.



Hình 6. Lỗi điểm treo của lớp Đoạn tim đường bộ

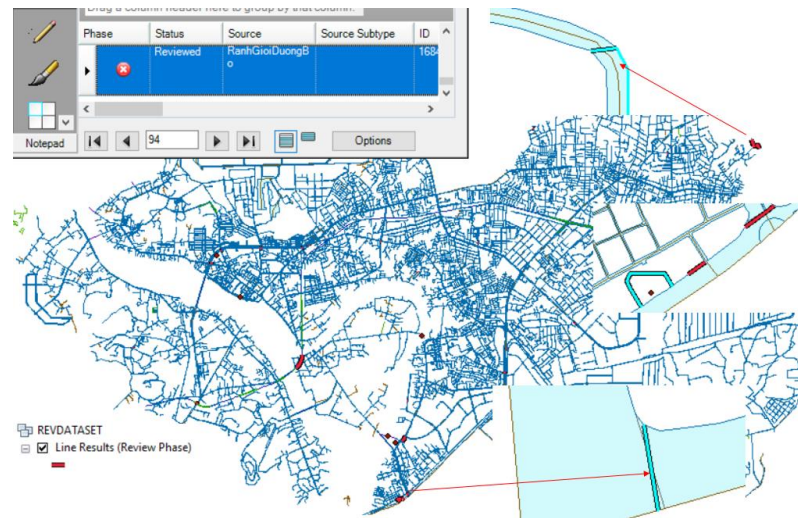
Đoạn tim đường bộ phải bao phủ cầu vượt sông: Đây là một luật quan hệ không gian giữa tim đường bộ và cầu vượt sông. Tổng có 29 lỗi, chiếm tỷ lệ lỗi cao thứ 2 với 14,22%. Các lỗi này do công tác xây dựng và chuẩn hóa dữ liệu nội nghiệp. Sử dụng quan hệ không gian để phát hiện lỗi và phóng to đối tượng mới nhìn thấy rõ về lỗi, việc kiểm tra bằng mắt thường sẽ rất khó nhận biết được (Hình 7).



Hình 7. Lỗi đoạn tim đường bộ phải bao phủ cầu vượt sông

Mặt đường bộ và ranh giới đường bộ: Đây là lớp dữ liệu rất quan trọng trong hỗ trợ quy hoạch giao thông. Nếu ranh giới và mặt đường không khớp nhau, kết quả phân tích không gian sẽ bị ảnh hưởng. Qua kiểm tra cho thấy có 94 đối tượng

(3,52%) lỗi được đánh dấu màu đỏ. Thực hiện phóng to một số đối tượng để kiểm tra cho thấy cơ bản đường ranh giới đều cắt chéo vào khu vực mặt đường bộ (Hình 8).



Hình 8. Lỗi chồng xếp bề mặt đường bộ và ranh giới đường

4. Kết luận

Nghiên cứu đã ứng dụng kỹ thuật GIS, thực nghiệm kiểm tra thành phần tính thống nhất logic không gian của chất lượng tập dữ liệu cơ sở nền địa lý tỷ lệ 1:5.000 thị xã Biên Hòa theo quy chuẩn QCVN 42:2020/BTNMT và quy định kỹ thuật về mô hình, cấu trúc nội dung cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:2.000 và 1:5.000. Thực nghiệm cho thấy rằng: Áp dụng kỹ thuật phân tích mối quan hệ không gian trong GIS để thực hiện kiểm tra chất lượng không gian một cách tự động đem lại hiệu quả cao. Phương pháp có thể xác định được nhanh các lỗi về nhất quán không gian theo quy tắc đã định như quan hệ giữa điểm-đường, đường-vùng, điểm và vùng.

Nghiên cứu thực hiện trên một phần nhỏ là nhóm dữ liệu giao thông của bộ cơ sở dữ liệu đầy đủ gồm 7 nhóm theo quy định. Dữ liệu thử nghiệm là điển hình, đại diện đầy đủ cho các bài toán về phân tích không gian. Tổng số các đối tượng và các phép so sánh lên đến hàng trăm ngàn phép toán đối sánh khác nhau. Nếu sử dụng các phương pháp kiểm tra thủ công bằng mắt thường thì rất tốn thời gian và kinh phí. Vì vậy, sử dụng quy trình kiểm tra chất lượng dữ liệu nền địa lý một cách tự động là cần thiết. Đặc biệt, ứng dụng kỹ thuật phân tích GIS với dữ liệu không gian là rất hiệu quả và chính xác.

Acknowledgements

Nghiên cứu này được hỗ trợ bởi Đề tài cấp Bộ Tài nguyên và Môi trường, Nước Cộng hòa Xã hội Chủ nghĩa Việt Nam, “Nghiên cứu giải pháp công nghệ

kiểm tra, thẩm định chất lượng cơ sở dữ liệu nền địa lý quốc gia”, mã số TNMT.2023.04.08.

References

1. SIST EN ISO 19113:2005 (2002), Geographic information – Quality principles (ISO 19113), editor.
2. SIST EN ISO 19114:2005 (2003), Geographic information – Quality evaluation procedures (ISO 19114), editor.
3. Z Ahmed (2020), "Quality Control and Quality Assurance in GIS Database and Applications", *GIS Consultant–Ain Shams University*.
4. Peter A Burrough, Rachael A McDonnell và Christopher D Lloyd (2015), *Principles of geographical information systems*, Oxford University Press, USA.
5. Rodolphe Devillers và các cộng sự (2010), "Thirty years of research on spatial data quality: achievements, failures, and opportunities", *Transactions in GIS*. 14(4), pp. 387-400.
6. ESRI (2019), *Get Started with ArcGIS Data Reviewer*, USA.
7. ESRI (2023), *Spatial relationships in ArcGIS GeoAnalytics Server*, truy cập ngày, tại trang web <https://pro.arcgis.com/en/pro-app/latest/tool-reference/big-data-analytics/spatial-relationships-with-big-data.htm#>.
8. Michael F Goodchild (1988), "Stepping over the line: technological constraints and the new cartography", *The American Cartographer*. 15(3), pp. 311-319.
9. Peyman Hashemi và Rahim Ali Abbaspour (2015), "Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept", *OpenStreetMap in GIScience: experiences, research, and applications*, pp. 19-36.
10. Mike Hendricks, Jen Athey và Amy Macpherson (2022), "QA/QC of GeMS DataCase Study: The Alaska DGGs Geologic Mapping System".
11. RM Henriksson và Tomi Kauppinen (2007), "An ontologydriven approach for spatial data quality evaluation", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 34.
12. Wolfgang Kainz (1995), "Logical consistency", *Elements of spatial data quality*. 202, pp. 109-137.
13. Anita Kukulska và các cộng sự (2018), "Methodology of evaluation and correction of geometric data topology in QGIS software", *Acta Scientiarum Polonorum: Formatio Circumiectus*. 17(1).
14. H. Moellering (1987), "draft proposed standard for digital cartographic data", *National Committee for Digital Cartographic Data Standards*.
15. Joanne Nightingale và các cộng sự (2018), "Quality assurance framework development based on six new ECV data products to enhance user confidence for climate applications", *Remote Sensing*. 10(8), pp. 1254.
16. Ge Peng và các cộng sự (2021), "International community guidelines for sharing and reusing quality information of individual earth science datasets", Available online (Accessed on 9/25/2023): <https://osf.io/xsu4p/download>.

17. Sukhjit Singh Sehra, Jaiteg Singh và Hardeep Singh Rai (2017), "Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox", *Future Internet*. 9(2), pp. 15.
18. International Organization for Standardization (2013), "ISO 19157:2013 Geographic Information—Data Quality", *ISO: Geneva, Switzerland: ISO/TC211*.
19. Danang Budi Susetyo và Aji Putra Perdana (2016), Automation of Indonesian Topographic Data Quality Control Using Data Reviewer, *International Conference on Technology, Innovation, and Society (ICTIS)*.
20. Bộ Tài nguyên và Môi trường (2012), Thông tư quy định Quy chuẩn kỹ thuật quốc gia về chuẩn thông tin địa lý cơ sở QCVN 42:2012/BTNMT, editor.
21. Bộ Tài nguyên và Môi trường (2014), Thông tư số 55 /2014/TT-BTNMT ngày 12 tháng 9 năm 2014 Quy định kỹ thuật về mô hình cấu trúc, nội dung cơ sở dữ liệu nền địa lý tỷ lệ 1:2000 và 1:5000 editor.
22. Bộ Tài nguyên và Môi trường (2020), Thông tư Ban hành Quy chuẩn kỹ thuật quốc gia về chuẩn thông tin địa lý cơ sở QCVN 42:2020/BTNMT, editor.
23. Bộ Tài nguyên và Môi trường (2020), Thông tư Số 15/2020/TT-BTNMT ngày 30 tháng 11 năm 2020 Quy định kỹ thuật về mô hình cấu trúc, nội dung cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:10.000, 1:25.000, editor.
24. Bộ Tài nguyên và Môi trường (2022), Thông tư số 07/2022/TT-BTNMT ngày 30 tháng 6 năm 2022 Ban hành Quy chuẩn kỹ thuật quốc gia về về cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:50.000, 1:100.000, QCVN 71:2022/BTNMT, editor.
25. Aleš Veršič (2009), Spatial data quality control process based on ISO 19113 and ISO 19114, *24th International Cartographic Conference, Santiago, Chile*.