Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Topological Consistency of Geospatial Data

Consistência Topológica de Dados Geoespaciais

Leandro Luiz Silva de França [1] and José Luiz Portugal [2]

1 Federal University of Pernambuco, Graduate Program in Geodetic Sciences and Geoinformation Technologies, Recife, Brazil. geoleandro.franca@gmail.com. ORCID: https://orcid.org/0000-0003-0863-1926
2 Federal University of Pernambuco, Graduate Program in Geodetic Sciences and Geoinformation Technologies, Recife, Brazil. joseluiz.portugal@gmail.com. ORCID: https://orcid.org/0000-0002-6508-1687

Received: 06.2021 | Accepted: 01.2022

Abstract: For correct analysis and decision making, based on geospatial data, it is essential to attest the data's reliability and integrity. Regarding Logical Consistency, there are still few standards and research that address the subject in a systematic way, and there is no consensus on the proper procedures for quality control in this category. Consequently, geospatial data producers have several difficulties to ensure the integrity and consistency of their cartographic products, due to inadequate verification of logical relationships, causing rework with repeated and inefficient verifications. This article, therefore, aimed to document concepts inherent to the assessment of topological consistency and present practical examples of validation applied to real data from cartographic production. The application of this research took place in a case study for a geospatial vector dataset (*Conjunto de Dados Geoespaciais Vetoriais* – CDGV) of the Mapping Project of the State of Bahia, detailing the entire methodological procedure for the topological validation of these data. From the results, there was a considerable number of topological inconsistencies, indicating the need for modernization of validation procedures by geospatial data producers. Regarding the contributions of this work, in addition to the compilation of a theoretical basis on the topological consistency of geospatial data, besides the development of solutions for the verification of inconsistencies with SQL queries based on PostgreSQL and its spatial extension PostGIS.
Keywords: Topological Consistency. Quality control. Topology. Validation. PostGIS.

Resumo: Para corretas análises e tomadas de decisões, baseadas em dados geoespaciais, é essencial que esses dados tenham sua confiabilidade e integridade atestadas. No que se refere à Consistência Lógica, ainda são poucas as normas e pesquisas que tratem do assunto de forma sistemática, não havendo um consenso sobre os procedimentos adequados para o controle de qualidade desta categoria. Consequentemente, alguns produtores de dados geoespaciais apresentam diversas dificuldades para garantir a integridade e consistência dos seus produtos cartográficos, devido à verificação inadequada dos relacionamentos lógicos, ocasionando retrabalhos com verificações repetidas e pouco eficientes. Este artigo, portanto, teve como finalidade documentar conceitos inerentes à avaliação da consistência topológica e apresentar exemplos práticos de validação aplicados em dados reais de produção cartográfica. A aplicação desta pesquisa ocorreu em um estudo de caso para um Conjunto de Dados Geoespaciais Vetoriais (CDGV) do Projeto de Mapeamento do estado da Bahia, detalhando-se todo o procedimento metodológico para a validação topológica desses dados. Dos resultados, verificou-se uma quantidade considerável de inconsistências topológicas, indicando a necessidade de modernização dos procedimentos de validação pelos produtores de dados geoespaciais. No que se refere às contribuições deste trabalho, além da compilação de uma base teórica sobre a consistência topológica de dados geoespaciais, destaca-se também o desenvolvimento de soluções para a verificação das inconsistências com consultas SQL baseadas no software livre PostgreSQL e sua extensão espacial PostGIS.
Palavras-chave: Consistência Topológica. Controle de Qualidade. Topologia. Validação. PostGIS.

1 INTRODUCTION

For advanced users of Geographic Information System (GIS), who need to perform activities that go beyond the simple visual exploration of data such as geoprocessing analysis, coding, routing, susceptibility maps, etc., it is necessary to ensure the geometric properties and maintenance of relationships between features (WADEMBERE; OGAO, 2014).

The complete detection and correction of errors is indispensable to prove that all procedures have been employed, so that geospatial data adequately represent the real world, allowing studies, queries and analyses to be more reliable.

In this context, topological consistency refers to the coherence of topological characteristics explicitly established to a dataset (ISO, 2013).

Researches on the implementation of integrity constraint has already been developed for the process of acquisition and insertion in the database, where it can be cited Borges et al. (2002), Stempliuc (2008) and Lizardo and Davis Jr. (2017), however, it is observed that these integrity constraints need be systematically listed as requirements (or rules) for the final evaluation of the dataset (FRANÇA et al., 2020).

Ariza-López et al. (2019) state that logical rules can be verified by automatic computational routines. However, for this validation to be performed properly, it is necessary to detail the evaluated quality elements.

The purpose of this article is to be a guidance for the development of quality control procedures with regard to the validation of topological consistency. The methodology developed here arose in the context of mapping the State of Bahia, a project sponsored by the Superintendence of Economic and Social Studies of Bahia and executed by the Brazilian Army.

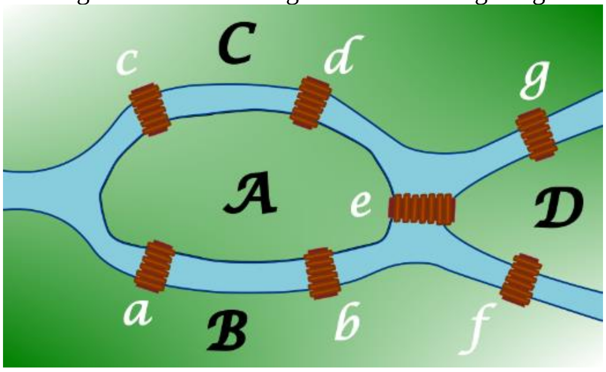
Thus, this research and application of topological validation were carried out considering the main classes of the hydrography category of this project, on a database in the modeling ET-EDGV 2.1.3. From the results, it was verified that Gothic, a software for validation and correction of vector data used in the mapping project of the Bahia by the Geographic Service Bureau (PASSOS et al., 2017), has been insufficient to ensure the complete integrity of the data.

In addition to the theoretical contribution, this work also demonstrates in practice how to perform topological validation, using the free DBMS software PostgreSQL with its spatial extension PostGIS, from SQL queries, which are openly available to be adapted and reused in other mapping projects.

2 TOPOLOGY FOR VECTOR GEOSPATIAL DATA

The term topology dates to the eighteenth century, more precisely 1736, where Leonhard Euler, the founder of the bases of topology and graph theory, solved the problem of the seven bridges of Königsberg (Figure 1), present-day Kalingrad, Russian exclave in the Baltic Sea. Euler proved, based on the topology, that it was not possible to cross all the bridges without having to cross one of them at least twice (SHIELDS, 2012; ROMANHOLI; QUEIROZ FILHO, 2018).

Figure 1 - Seven Bridges of Euler Königsberg.



Source: Adapted from Shields (2012).

In general, topology can be defined as a study of the properties of objects, such as adjacency, connectivity and continence (ROMANHOLI; QUEIROZ FILHO, 2018). Topological relationships are fundamental in the definition of spatial integrity rules, which specify the geometric behavior between objects (BORGES et al., 2001).

According to Maraş et al. (2010), vector data has been used to represent geographic features due to the optimization of spatial analyses. Vector data consists of a structure of pairs of coordinates, and can be divided into two situations:

a) Non-topological structure (vector data without associated topology);
b) Topological structure (vector data with associated topology).

2.1 Non-topological structure

The set of geographical features in non-topological spatial structure, also called *spaghetti data structure* (MARAŞ et al., 2010), represented the features through three geometric shapes (point, line and polygon), according to the mapping scale.

Features that are represented by point geometry are zero-dimensional (0D), each point is defined by a pair of coordinates (x, y). The features of the line type are one-dimensional (1D), each element being defined by a sequence of coordinates (x, y). Polygon-type features are two-dimensional (2D), defined by closed shapes and composed of lines that start and end at the same point.

Despite the ease of creating and editing vector data without topology in the various GIS software, non-topological structures may contain several problems that hinder spatial data analysis (MARAŞ et al., 2010):

a) point-type feature may not be at the intersection point of lines, e.g. bridge at the intersection of river and highway;

b) Neighborhood relationship between features may not be clear;

c) Points of contact do not coincide, for example, river that does not coincide with the lake edge;

d) As the neighborhood between two polygons is represented twice, there may be no total coincidence in all vertices, causing *overlaps* or *gaps*;

e) Navigation is not possible when the concept of network and direction is not present in the features of the line type.

Siejka et al. (2013) point out that the simple (non-topological) vector model has two basic disadvantages. One is data redundancy in data duplication at borders, where a point belongs to two or more objects. In these cases, the coordinates of the points are saved in each of these objects. The other significant disadvantage is that spatial relationships between objects can only be detected through more complex methods of analytical geometry, causing a higher computational cost in queries.

2.2 Topological structure

The topology makes it possible to examine characteristics that go beyond the geometric information of the features. The objective of topological knowledge in GIS is to increase the opportunities for spatial analysis, in order to represent spatial relationships such as neighborhood, coincidence, direction and connections.

In topology, a **node** corresponds to the point; the **arc** (or **edge**) is an element corresponding to the line; and the **polygon** (or **face**). The arc is a set of pairs of coordinates that begins with a node and ends with a node. The polygon, in turn, is a two-dimensional space delimited by arcs.

There are two approaches: graph-based arc-node topology, which is widely used to represent networks; and the arc-node-polygon, which considers the area of polygons (CASANOVA et al., 2005). In both the arc-node and arc-node-polygon approach, topology can be understood as the relationship between its elements (nodes, arcs and polygons), being a means of defining spatial relationships, in addition to geometrically identifiable metric relationships (MARAŞ et al., 2010).

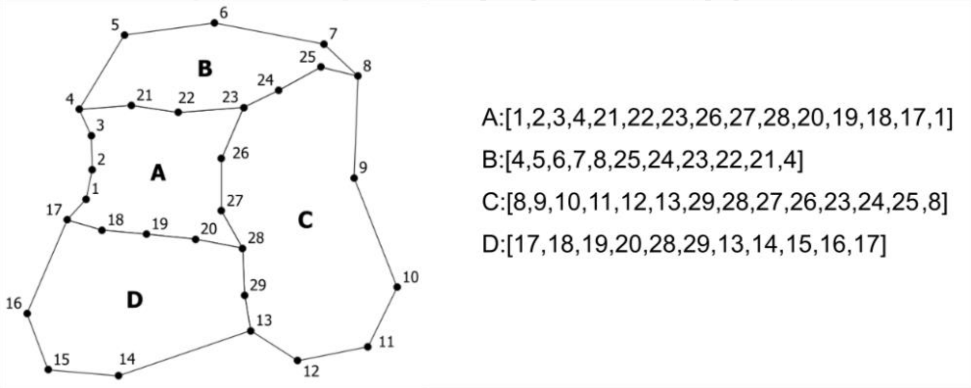
For the arc-node-polygon topology, a geospatial database has a topological structure when the following elements are determined and stored:

a) Arcs that define the boundaries of each polygon (in the polygon topology table);
b) Neighborhood relationship between polygons (in the arc topology table);
c) Connection at intersection points (in the node topology table);
d) Start and end points of the arcs (in the arc-coordinates table).

2.3 Comparison between non-topological and topological structures

Figure 2 presents an example of non-topological structure (spaghetti), verifying, on the left, polygons A, B, C and D, and on the right the identifier of the points used for the construction of each polygon. It is observed that each polygon is represented by a “loop”, represented by a sequence of points.

Figure 2 - Example of non-topological structure (spaghetti).



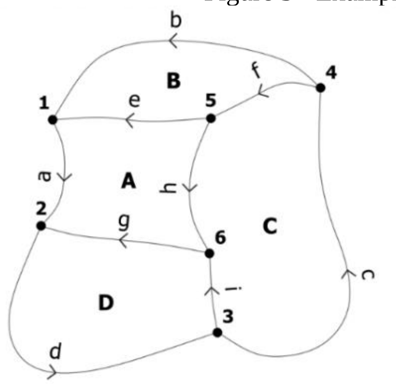
Source: Adapted from Cao (2014).

In the non-topological structure, there is no explicit representation of topological relationships between features, such as adjacency. In addition, data redundancy is verified for points shared by two or more polygons.

Figure 3 is an example of the topological structure of features A, B, C, and D, with "X" being the polygon "universe". From this figure, the following properties are observed:

a) each arc has exactly one start and end node;
b) each node can be initial or end (or both) of at least one directional arc;
c) each polygon is surrounded by one or more arcs;
d) Arcs may be intercepted only in their nodes; and
e) Each directed arc must have only one right or left face.

Figure 3 - Example of topological structure.



Arc	Start	End	Left	Right
a	1	2	A	X
b	4	1	B	X
c	3	4	C	X
d	2	3	D	X
e	5	1	A	B
f	4	5	C	B
g	6	2	D	A
h	5	6	C	A
i	3	6	D	C

Source: Adapted from Cao (2014).

In databases, topological properties can be stored and used for several queries such as connecting lines, network points, and features that are to the right or left of linear features.

However, the existence of topological data, in addition to non-topological data and its attributes, can cause a considerable increase in the volume of the geographic database, besides strongly impairing performance in operations of inserting, altering, and deleting data, due to the need to maintain updated the *winged-edge topological structures* (BAUMGART, 1975). Furthermore, common operations, which depended on the use of the complete geometry of objects, would also be hampered by the need of "join" between tables of nodes, arcs and faces.

When the vectorization of non-topological data is performed according to the topological construction for the common arcs (FRANÇA et al., 2018), it is possible to guarantee quality, within other advantages:

a) Time gain (avoiding duplicate vectorization and common borders);
b) ensure that polygons share the same border;
c) borders will be represented without repeated points;
d) fast data processing; and
e) cost reduction.

Currently, the representation of geographic data as non-topological structures becomes the most viable option for assessment of topological consistency by following the standards of the Open Geospatial Consortium (OGC) (STOLZE, 2003, HERRING, 2011), which establishes the Dimensionally-Extended Nine-Intersection Model (DE-9IM) as the basis for verifying integrity constraints in the spatial relationship between features (LIZARDO; DAVIS JR., 2017).

3 TOPOLOGICAL RELATIONSHIP WITH THE USE OF THE DE-9IM MATRIX

In the context of GIS, topology is understood as the relative positioning between features, where topological consistency is related to the verification of the adequacy of spatial relationships of geometries of features to situations identified in the real world.

Regardless how vector data is structured, the understanding of the relationships between geometries is essential for the execution of spatial analyses. These relationships are defined through topological operators (EGENHOFER et al., 1993).

The types of spatial relationships between features can be determined through the Dimensionally-Extended Nine-Intersection Model (DE-9IM)(CLEMENTINI; DI FELICE, 1995).

DE-9IM is applied to feature geometries stored in non-topological vector structures. The DE-9IM result was established by the Open Geospatial Consortium (OGC) as a basis for the implementation of spatial relationship functions in DBMS (LIZARDO; DAVIS JR., 2017).

The DE-9IM matrix is the result of the intersections of the interior, boundary and exterior of two analyzed geometries. The interior (I), boundary (B) and exterior (E) for the point, line and polygon geometries are:

a) I (point) = point;
b) I (line) = all points of the line, excluding the start and end points in the case of open line;
c) I (polygon) = the points within the boundary;
d) B (point) = ∅
e) B (line) = the start and end points of an open line or ∅ for a closed line;
f) B (polygon) = rings (lines) that make up the polygon;
g) E (points, lines or polygons) = what is not an interior or boundary.

Figure 4 shows the configuration of the M matrix resulting from the intersections between interior, boundary and exterior of geometry A in relation to geometry B.

Figure 4 - DE-9IM matrix.

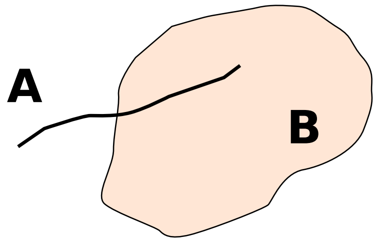
$$M = \begin{bmatrix} I(A) \cap I(B) & I(A) \cap B(B) & I(A) \cap E(B) \\ B(A) \cap I(B) & B(A) \cap B(B) & B(A) \cap E(B) \\ E(A) \cap I(B) & E(A) \cap B(B) & E(A) \cap E(B) \end{bmatrix}$$

Source: Adapted from Clementini and Di Felice (1995).

DE-9IM can be applied to spatial objects of different dimensions (BORRMANN; RANK, 2009). In case a geometry A intersects another geometry B, the value 0, 1 or 2 is returned, according to the intersection dimension (point, line and polygon, respectively).

Figure 5 is an example of the result of the DE-9IM matrix for a configuration between a line-type Geometry A and another polygon-type B geometry. When there is no intersection, the result is "F" (for *false*); in the otherwise (*true)* case, the dimension of the intersection is returned.

Figure 5 - Result example of the DE-9IM matrix.



$$M = \begin{bmatrix} 1 & 0 & 1 \\ 0 & F & 0 \\ 2 & 1 & 2 \end{bmatrix}$$

Elaboration: The authors (2021).

Topological operators, also called topological predicates (BORRMANN; RANK, 2009), return the Boolean value “True” if the result of M is classified in the possible spatial relations schemes, verifying situations such as: intercepts, touches, crosses, contains, within, disjoint, equal, overlap, among other spatial predicates.

Table 1 presents the pattern of accepted values of the DE-9IM matrix for the main topological predicates. When the character “T” is used, it means that any dimension (0, 1 or 2) can be accepted. When the asterisk character (*) is used, it means that it can be any “T” or “F” value.

Table 1 - DE-9IM matrix pattern for the main topological predicates.

Topological Predicate	**Matrix Pattern**
A.Equals(B)	$\begin{bmatrix} T & * & F \\ * & * & F \\ F & F & * \end{bmatrix}$
A.Disjoint(B)	$\begin{bmatrix} F & F & * \\ F & F & * \\ * & * & * \end{bmatrix}$
A.Intersects(B)	$\begin{bmatrix} T & * & * \\ * & * & * \\ * & * & * \end{bmatrix}$ or $\begin{bmatrix} * & T & * \\ * & * & * \\ * & * & * \end{bmatrix}$ or $\begin{bmatrix} * & * & * \\ T & * & * \\ * & * & * \end{bmatrix}$ or $\begin{bmatrix} * & * & * \\ * & T & * \\ * & * & * \end{bmatrix}$
A.Touches(B)	$\begin{bmatrix} F & T & * \\ * & * & * \\ * & * & * \end{bmatrix}$ or $\begin{bmatrix} F & * & * \\ * & T & * \\ * & * & * \end{bmatrix}$ or $\begin{bmatrix} F & * & * \\ T & * & * \\ * & * & * \end{bmatrix}$
A.Crosses(B)	$\begin{bmatrix} T & * & T \\ * & * & * \\ * & * & * \end{bmatrix}$ or $\begin{bmatrix} T & * & * \\ * & * & * \\ T & * & * \end{bmatrix}$ or $\begin{bmatrix} 0 & * & * \\ * & * & * \\ * & * & * \end{bmatrix}$
A.Overlaps(B)	$\begin{bmatrix} T & * & T \\ * & * & * \\ T & * & * \end{bmatrix}$ or $\begin{bmatrix} 1 & * & T \\ * & * & * \\ T & * & * \end{bmatrix}$
A.Within(B)	$\begin{bmatrix} T & * & F \\ * & * & F \\ * & * & * \end{bmatrix}$
A.Contains(B)	$\begin{bmatrix} T & * & * \\ * & * & * \\ F & F & * \end{bmatrix}$

Source: Strobl (2017).

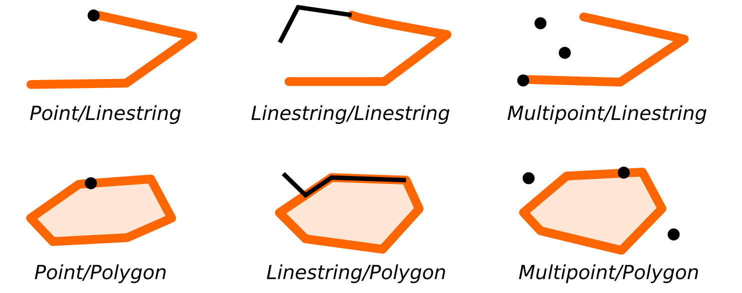
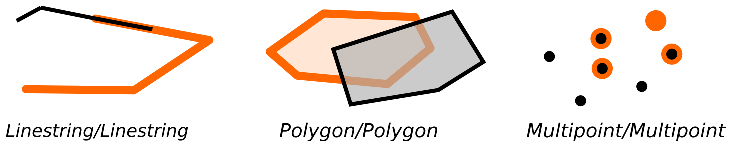
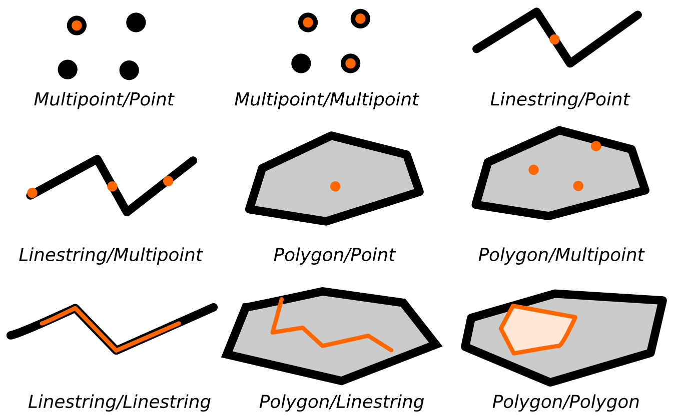
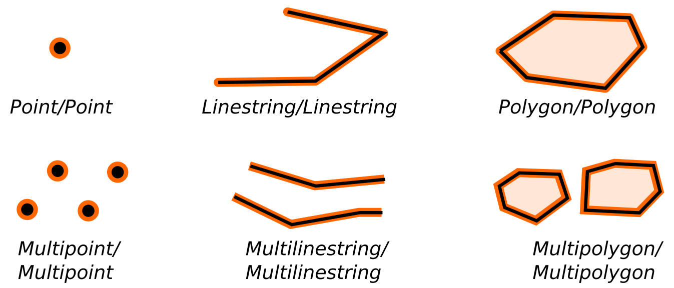
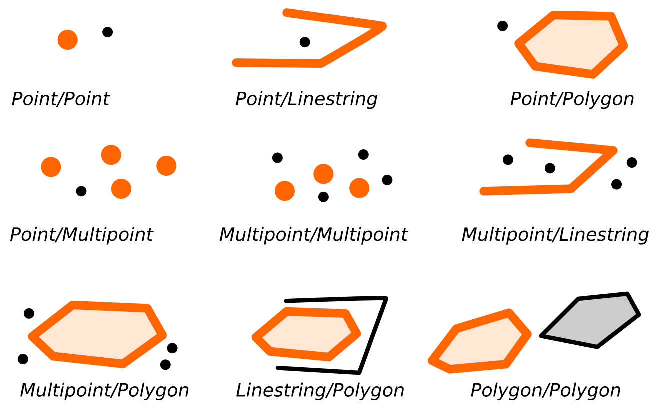
Returning attention to the situation in Figure 5, and comparing with the possibilities of Table 1, it is evident that geometry A intercepts geometry B as well as A crosses B.

Topological operators can also be used with multipart geometries. Some operator possibilities according to geometry types are presented in Table 2. In all cases, a geometry type A, in black color, is related to a second geometry B, in orange color.

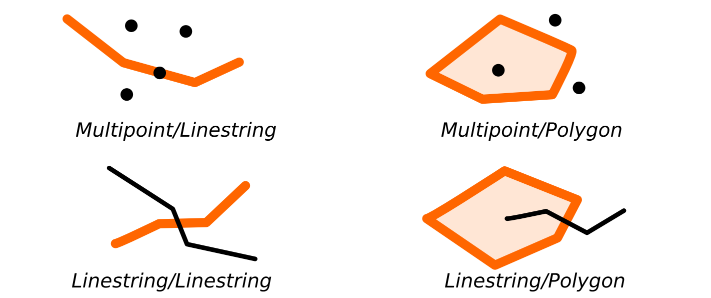
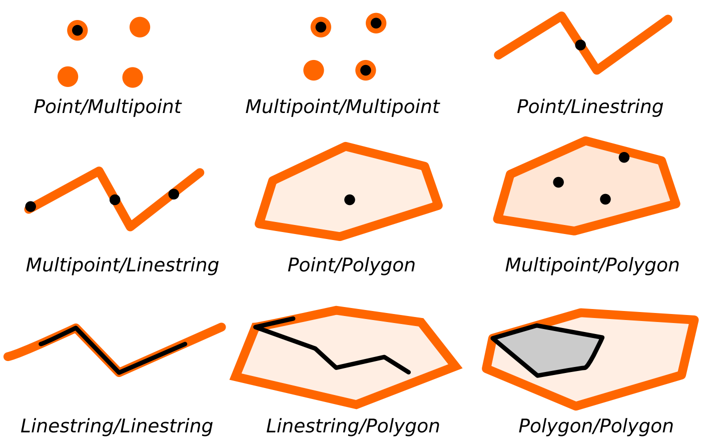
 DOI: http://dx.doi.org/10.14393/rbcv74n3-61508

Table 2 - Examples of topological operators.

To be continued

Topological Operator	Examples
• *Touches*: tests whether geometry A touches a Geometry B, that is, checks whether the geometries have at least one point in common, as long as their interiors do not intersect.	
• *Overlaps:* tests whether a geometry A overlaps another B geometry of the same dimension. Geometries have some, but not all points in common.	
• Contains: tests if a geometry A contains another geometry B, that is, all points of geometry B are points of geometry A and the interiors of the two geometries have at least one point in common. It is the equivalent of "B within A".	
• Equals: tests whether one geometry is equal to another, that is, the interiors intersect and no part of the inside or boundary of one geometry intersects the exterior of the other.	
• Disjoint: tests whether one geometry is disjointed from the other, that is, the two geometries have no point in common.	

Conclusion

Topological Operator	Examples
• Crosses: tests whether a geometry crosses another specific geometry. Geometries have some, but not all interior points in common.	
• Within: tests whether a geometry A is within geometry B. All points of geometry A is a point of geometry B, and the interiors of both have at least one point in common. It's the equivalent of "B contains A."	

Source: Adapted from Esri (2010).

4 TOPOLOGICAL VALIDATION

In the topological validation process, it is common to find invalid geometries or even geometries constructed with inadequate configuration to the cartographic representation of the geographic feature for a given mapping scale. These geometries make it impossible, or misleading, to determine spatial relationships, as well as any other geoprocessing analysis or geometric measurements such as area or perimeter.

For a better understanding of the topological validation process, considering the types of inconsistencies and the verification flow, the following validation levels can be considered in this work:

a) Geometry validation: aims to check geometry validity according to OGC's Simple Feature Specifications (SFS) and other geometry construction rules required by the model.

b) Intraclass (internal) validation: aims to identify topological inconsistencies between features of the same class (IBGE, 2017).

c) Interclass validation (external): aims to identify topological inconsistencies between features of different classes (IBGE, 2017).

Geospatial data must be created (or structured) according to the data model and all inconsistencies in the geometries (either shape or relative position of the vertices) must be checked and repaired in order to adequately represent a real-world feature.

Topological inconsistencies are related to the relative positioning between features within the same class (intraclass) or in different classes (interclass), in order to meet the meaning of the spatial relationship between objects. This verification can be done by creating topological structures, but the most common is the use of analytical methods based on the DE-9IM matrix.

In the case of this work, the data validation process was carried out in non-topological structures, seeking to analyze the topology between features through topological operators, based on the DE-9IM matrix.

4.1 Interclass validation

In interclass validation, topological rules are checked between distinct class features. For the execution of this validation to be effective, it is essential that intraclass and geometries validation are performed preliminary.

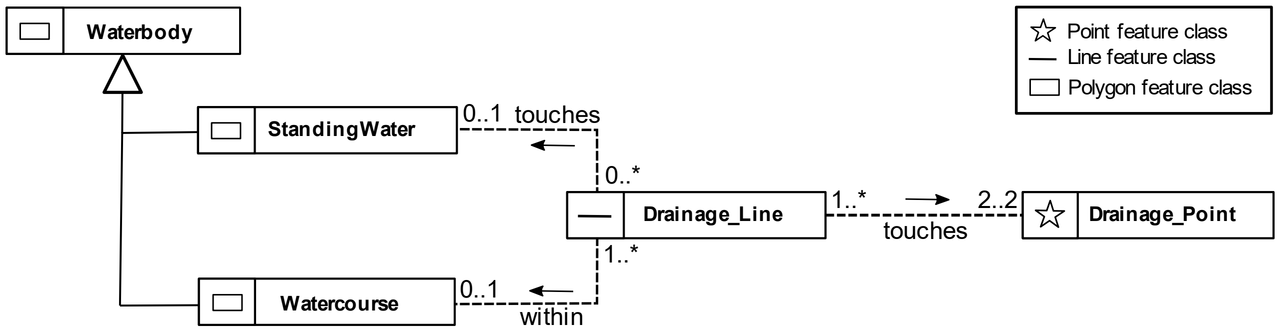
Spatial relationships between features of distinct classes can be indicated in conceptual model diagrams by their conventional name (e.g., contains, touches, etc.) or, in more specific cases, by constraints based directly on the results of the DE-9IM matrix (LIZARDO; DAVIS JR., 2017).

Spatial relationships are characterized by cardinality. Cardinality represents the number of instances of a class that can be associated with instances of another class (BORGES et al., 2005).

The cardinality notation adopted by the OMT-G model is based on the Unified Modeling Language (UML) (DAVIS JR.; LAENDER, 2000), where the minimum and maximum value are indicated, separated by two consecutive points. The asterisk is used when the modeler needs to symbolize the maximum value as "many", for example in: 1..*, which reads: "one to many".

Figure 6 presents a conceptual model diagram in OMT-G for some classes of a watershed. In this diagram it is possible to observe the main spatial relationships between classes and their cardinalities.

Figure 6 - Spatial relationships between classes and their cardinalities.



Elaboration: The authors (2021).

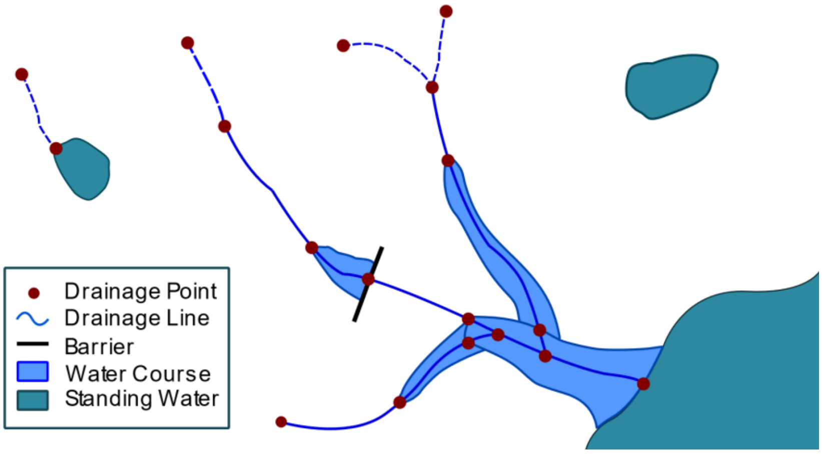
From Figure 6, it can be observed that the “Watercourse” and “StandingWater” feature classes are generalized in “Waterbody”. Table 3 presents the description of the classes presented in Figure 6. Figure 7 illustrates some cases of these classes.

Table 3 - Description of the studied hydrographic feature classes.

Class	***Geometry***	***Description***
Drainage_Point	Point	Connectivity point between drainage lines defined by their start and end points.
Drainage_Line	Line	Linear representation of the main flow of a watercourse.
Waterbody	Polygon	Surface with significant accumulation of water for a given scale. It can be classified into “Standing Water” (lakes, ponds, oceans, etc.) and “Water Course” (rivers, streams, canals, etc.), according to the existence of water flow.

Elaboration: The authors (2021).

Figure 7 - Relationships between hydrographic feature classes.



Elaboration: The authors (2021).

Although the conceptual model makes explicit the main spatial relationships, making them more understandable, other rules may be implicitly defined as part of the meaning of features, and can be deduced from the model (LIZARD; DAVIS JR., 2017).

From the diagram in Figure 6, for example, the following relationship is defined: a drainage line must be within 0 or 1 features of a watercourse. However, the same relationship, in the opposite direction, can be interpreted as: a watercourse must contain 1 or many drainage lines.

Many topological rules can be established to ensure data integrity, and the number of rules may be non-exhaustive (FRANÇA et al., 2020).

Table 4 lists, based on the classes presented in Figure 6, topological rules that should (or should not) occur, considering a certain cardinality.

Table 4 - Examples of rules for interclass validation.

Regra	**Classe 1**	**Deve**	**Topologia**	**Classe 2**	**Cardinalidade**
01	Drainage_Line	yes	*touches*	StandingWater	0..1
02	Drainage_Line	yes	*within*	Drainage_Line	0..1
03	Drainage_Line	yes	*touches*	Drainage_Point	2..2
04	WaterCourse	yes	*contains*	Drainage_Line	1..*
05	Drainage_Point	yes	*touches*	Drainage_Line	1..*
06	Drainage_Point	no	*within*	StandingWater	1..*
07	Drainage_Line	no	*crosses*	StandingWater	1..*
08	StandingWater	no	*intersects*	WaterCourse	1..*

Elaboration: The authors (2021).

It is observed that rules 04 and 05 correspond respectively to rules 02 and 03 in the opposite direction. Rules 06, 07 and 08 are examples of rules to check situations that should not occur, which, although not explicit in the data model, are consensus of understanding the real world.

It is important to note that all topological rules should be included in the conceptual schema, or at least in the data dictionary, in order to optimize the checking criteria and ensuring data integrity. The documentation of these rules in the technical specifications of the dataset, especially when required in contracts, will eliminate any subjectivity in the quality assessment process.

In practical terms, inconsistencies in interclass validation can be identified by SQL queries. Figure 8 shows an example of the PostGIS SQL query, which can be applied to rule 04 in Table 4, to identify polygons that do not contain at least one line.

Figure 8 - SQL query to identify topological inconsistencies between classes.

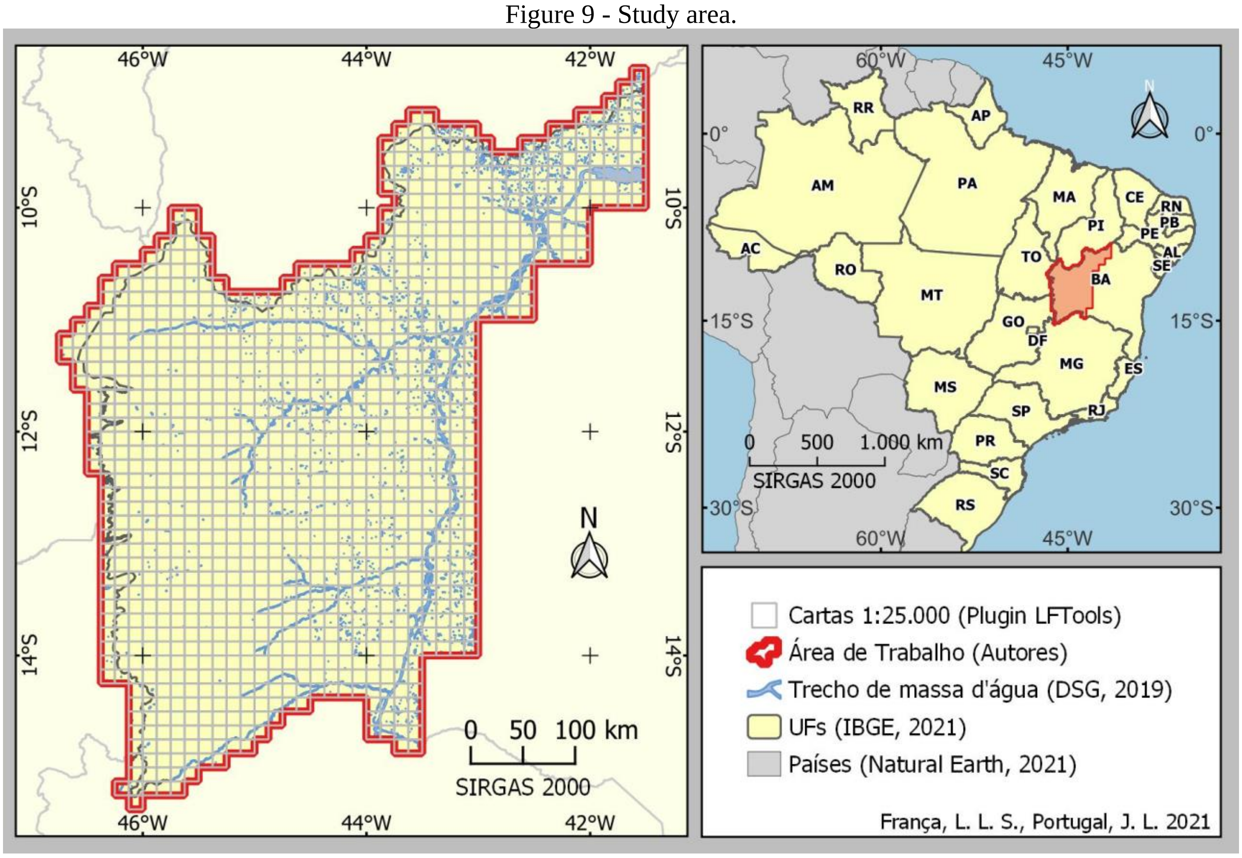
```sql
SELECT
V.id AS id,
V.geom AS geom
--,'hid_trecho_massa_dagua_a deve conter pelo menos 1 feição da classe hid_trecho_drenagem_l' AS erro_msg
FROM(
    SELECT
        C.id AS id,
        C.geom AS geom,
        sum(C.relation::int) AS n_elemts
    FROM(
        SELECT
            T1.id AS id,
            T1.geom AS geom,
            st_contains(T1.geom, T2.geom) AS relation -- relação topológica
        FROM cb.hid_trecho_massa_dagua_a T1, cb.hid_trecho_drenagem_l T2 -- classes
        ) C
    GROUP BY C.id, C.geom
) V
WHERE n_elemts = 0; -- cardinalidade
```

Elaboration: The authors (2021).

5 METHODOLOGY

The study area is located in the west of the Brazilian state of Bahia, whose topographic mapping was executed by the Brazilian Army between 2010 and 2019 in the Mapping Project of the State of Bahia. Figure 9 shows the cartography of the area.

Figure 9 - Study area.



Elaboration: The authors (2021).

This area has a total of 209. 923.5 km² and corresponds to a total of 1,116 topographic charts on the scale of 1:25,000, following the Brazilian systematic mapping (FRANÇA et al., 2017; FRANÇA; FERREIRA DA SILVA, 2018; PASSOS; FRANÇA, 2018; FRANÇA et al., 2019).

The material evaluated consists of the CDGV of the study area, using the modeling of the Technical Specifications for Structuring Vector Geospatial Data (ET-EDGV), version 2.1.3 (CONCAR, 2017).

This data is made available both through the *Banco de Dados Geográficos do Exército* (BDGEx), as well as through the geoportal of the Superintendence of Economic and Social Studies of the state of Bahia (SEI, 2021).

The data were worked using the SGBD PostgreSQL, v. 10.11, with its spatial extension PostGIS, v.2.5.1, and the maps were elaborated with QGIS 3.16 and the evaluations, analysis and reports were made based on SQL queries implemented for this work.

From the CDGV of western Bahia, the following classes of features were selected for evaluation:

a) Drainage Point (*Ponto_Drenagem*): connectivity points between two or more drainage lines;
b) Drainage Line (*Trecho_Drenagem*): corresponds to line-type geometries which represents the water flow, permanent or temporary, contained in a water course;
c) Standing Water *(Massa_Dagua*): body of water represented by polygon, such as ocean, bays, abandoned meanders, lakes, ponds, and dams that do not have water flow; and
d) Water Course (*Trecho_Massa_Dagua*): segments of watercourses represented by polygons, which have water flow.

In the evaluation of topological consistency, rules 01 to 08 were verified, already previously listed in Table 4. The SQL files used in topological validation are available for download through SQL queries (FRANÇA, 2021).

6 RESULTS AND DISCUSSION

The total number of features per evaluated class of the CDGV dataset considering the study area is given by Table 5.

Table 5 - Number of features per evaluated class.

Class	Class name in BD	Number of Features
Drainage Point	*Ponto_Drenagem*	17.051
Drainage Line	*Trecho_Drenagem*	248.809
Standing Water	*Massa_Dagua*	5.735
Water Course	*Trecho_Massa_Dagua*	3.654

Elaboration: The authors (2021).

The amount and percentage of inconsistencies found for the interclass topological rules, defined in Table 4, are given in Table 6.

Table 6 - Total inconsistencies found for interclass rules.

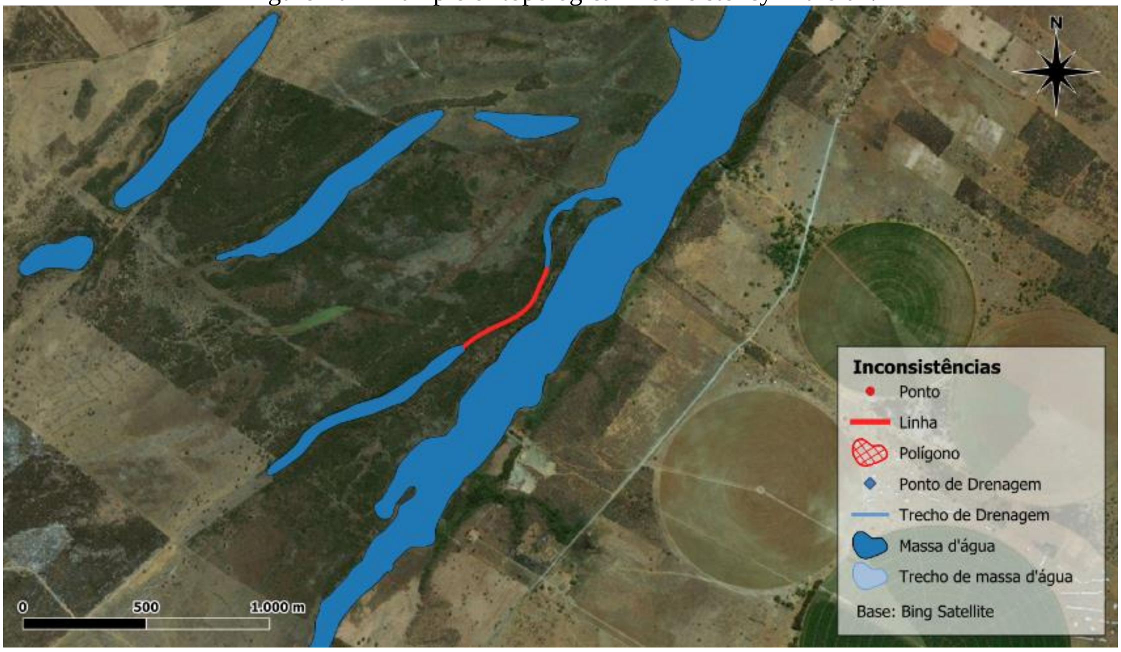
Rule	Description	# Inconsistencies and Percentage
01	*Trecho_Drenagem* you should touch 0 or 1 feature of class *Massa_dagua*.	1 (0,00%)
02	*Trecho_Drenagem* must be within 0 or 1 feature of the class *Trecho_Massa_Dagua.*	0 (0,00%)
03	*Trecho_Drenagem* must touch exactly 2 features of the class *Ponto_Drenagem.*	244.705 (98,35%)
04	*Trecho_Massa_Dagua* must contain at least 1 feature of the class *Trecho_Drenagem.*	649 (17,76%)
05	*Ponto_Drenagem* must touch at least 1 feature of the class *Trecho_Drenagem.*	3.452 (20,25%)
06	*Ponto_Drenagem* must NOT be within more than 0 feature of the class *Massa_Dagua*.	0 (0,00%)
07	*Trecho_Drenagem* must NOT cross more than 0 features of the class *Massa_Dagua*.	18 (0,01%)
08	*Massa_Dagua* must NOT intercept more than 0 feature of the *Trecho_Massa_Dagua*.	5 (0,09%)

Elaboration: The authors (2021).

 DOI: http://dx.doi.org/10.14393/rbcv74n3-61508

For the inconsistency of rule 01, a single occurrence was found (Figure 10), where a drainage line touches two features of the StandingWater class, which would characterize a flow of water between them, in violation of the conceptual model rule.

Figure 10 - Example of topological inconsistency - rule 01.

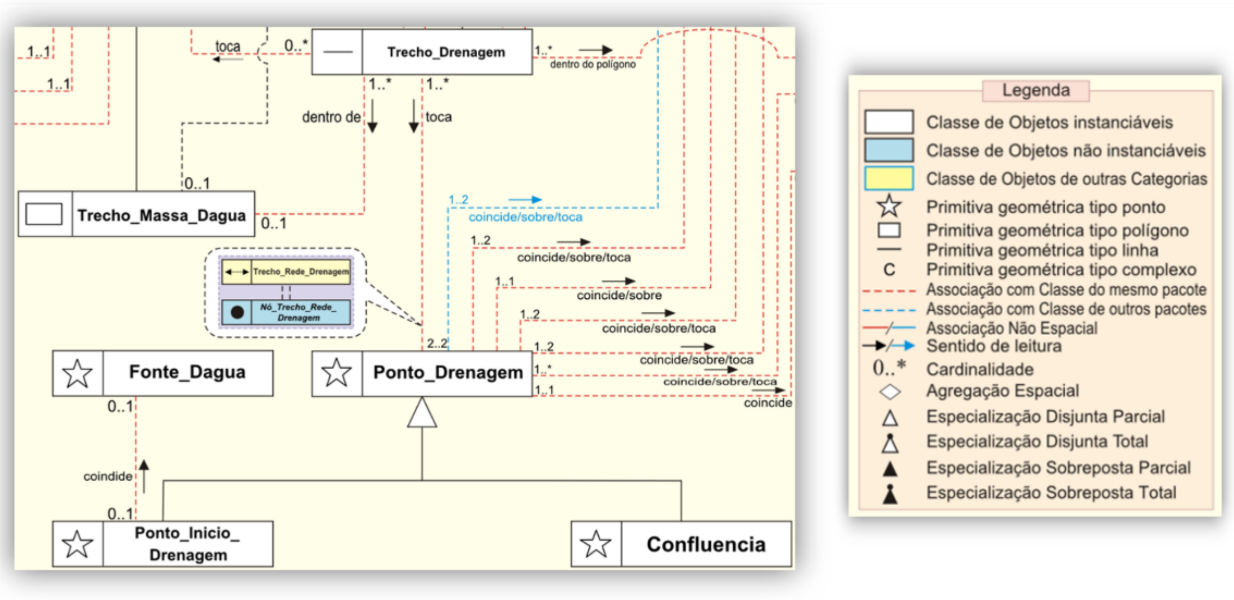


Elaboration: The authors (2021).

In the validation of rule 02, no case of inconsistency was found, that is, there was no case where a line was in two or more features of the Watercourse class.

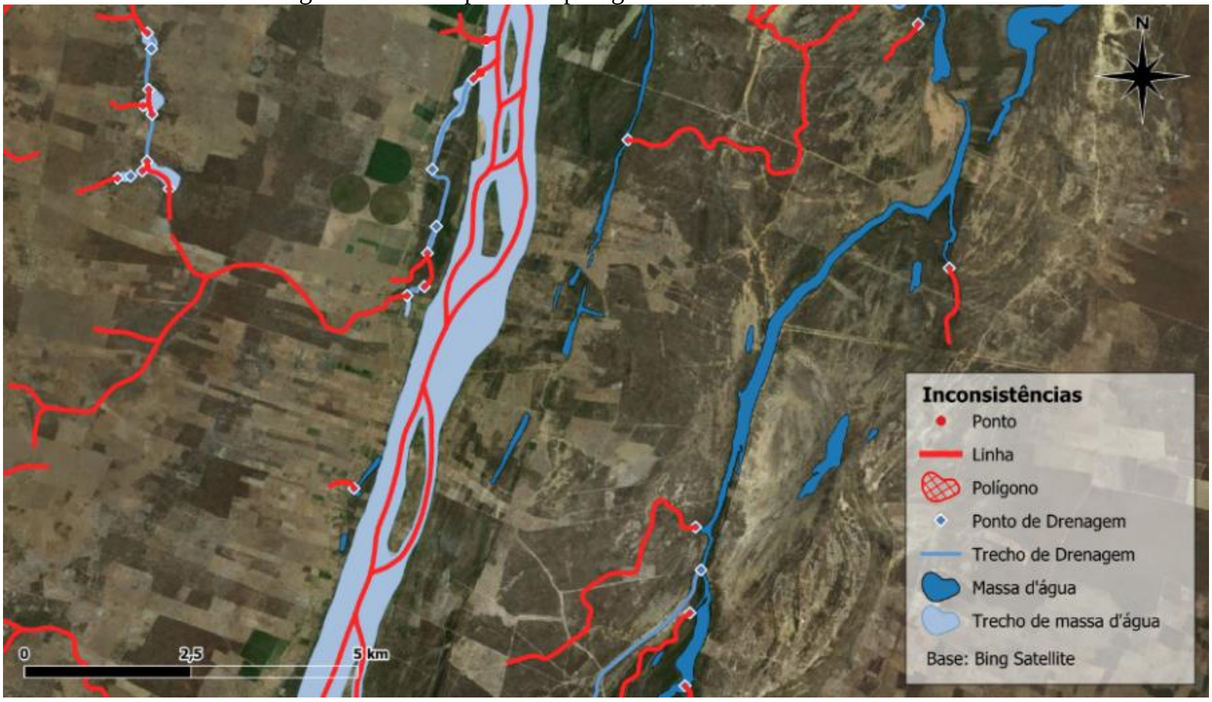
For the inconsistencies of rule 03, it was verified that the points of "confluence" and "drainage start" are not included in the class "drainage point", as provided for in ET-EDGV 2.1.3 (Figure 11). This problem in the database caused more than 98% of drainage lines to be inconsistent in this rule (Figure 12).

Figure 11 - Relationship between *Trecho_Drenagem* and *Ponto_Drenagem*, according to ET-EDGV.



Source: CONCAR (2010).

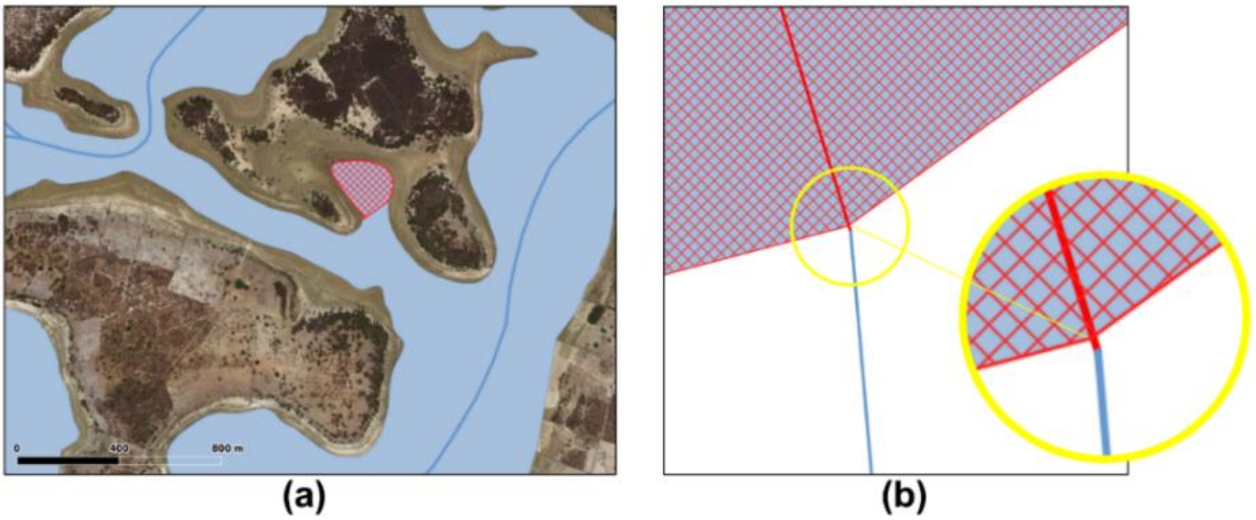
Figure 12 - Examples of topological inconsistencies - rule 03.



Source: The authors (2021).

In the validation of rule 04, there were rare cases of polygons of the class WaterCourse that are not actually crossed by a drainage line, as shown in Figure 13(a). However, the major cause of inconsistencies is related to the lack of precision of the feature coordinates, causing the line geometry to not be completely within the polygon, as shown in Figure 13(b).

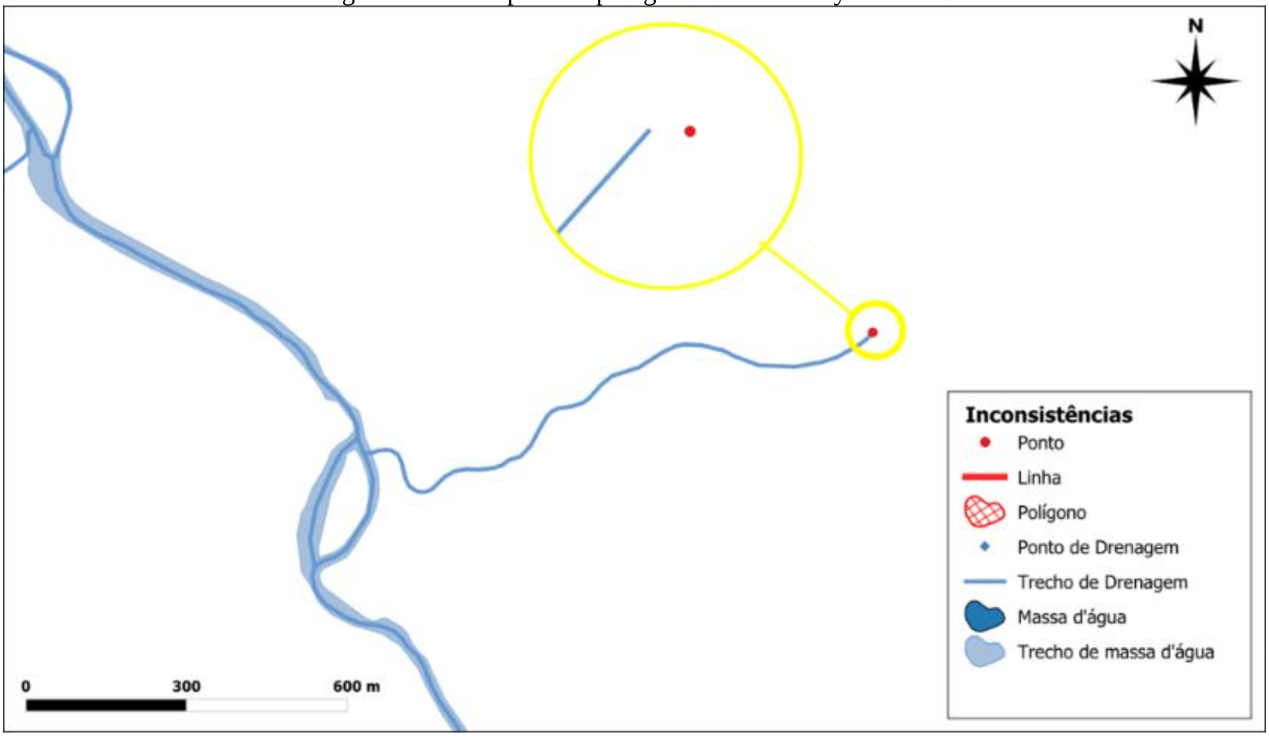
Figure 13 - Examples of topological inconsistencies - rule 04 (a) polygon that contains no line; (b) polygon that does not fully contain the line by coordinate accuracy.



Elaboration: The authors (2021).

In the validation of rule 05, no "isolated" drainage point was identified, that is, without being related to a line of the Drainage Line class. However, the reason for the inconsistencies found by the SQL query is related to the accuracy of the coordinates of some points, as verified in Figure 14.

Figure 14 - Example of topological inconsistency - rule 05.

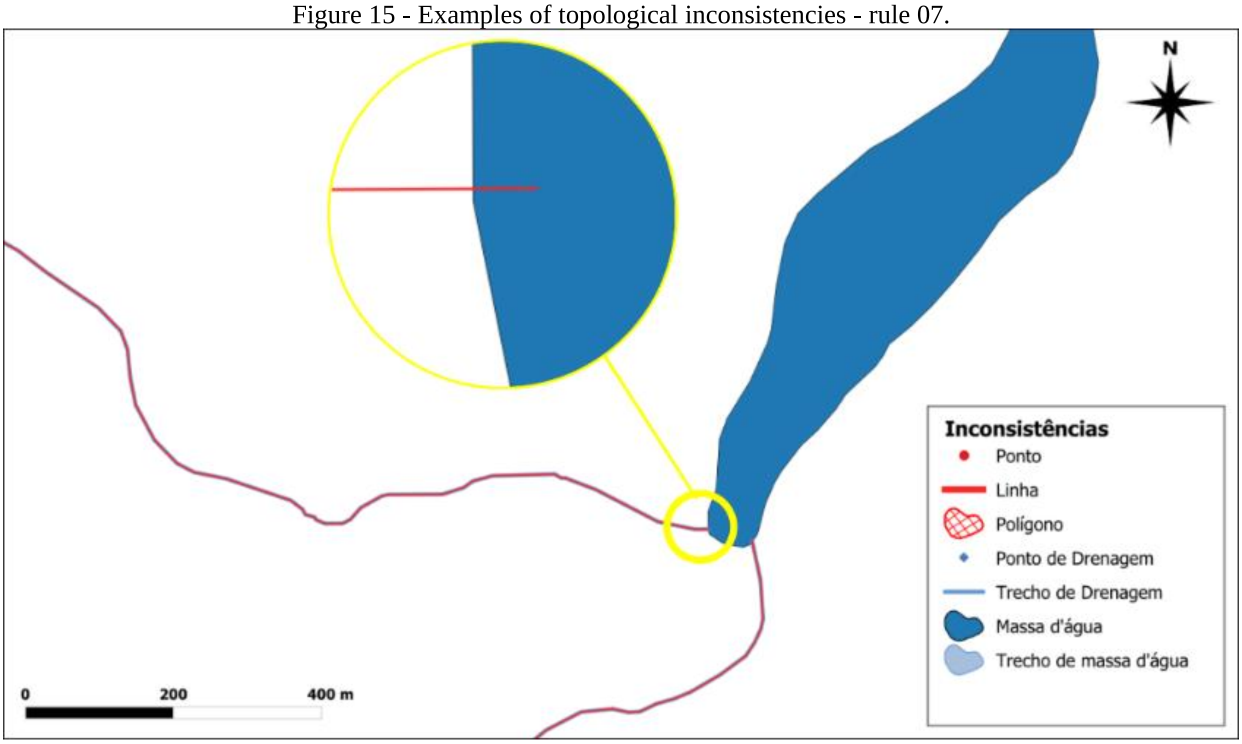


Elaboration: The authors (2021).

For rule 06, where it was checked if any Drainage Point is inside a Standing Water feature, no inconsistency was found.

For rule 07, which seeks to identify drainage lines that cross Standing Water, all 18 inconsistencies found are related to the accuracy of the coordinates, as shown in Figure 15.

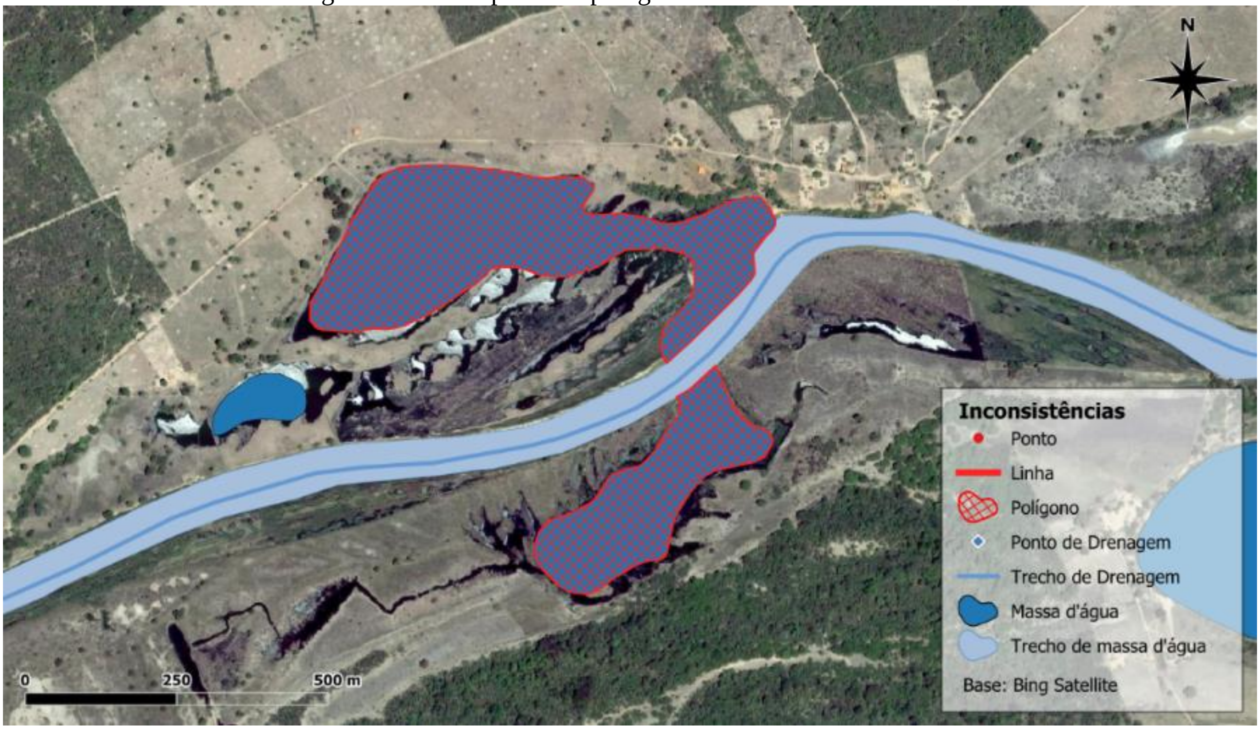
Figure 15 - Examples of topological inconsistencies - rule 07.



Elaboration: The authors (2021).

The last topological interclass rule identified five cases of intersection between polygons of the Standing Water and Water Course classes, as shown in Figure 16.

Figure 16 – Examples of topological inconsistencies - rule 08.



Elaboration: The authors (2021).

7 CONCLUSION

Some producers of geospatial data have great difficulty in ensuring the integrity and reliability of their products due to inadequate verification of topological relationships, causing rework with repeated and inefficient checks.

On the case study, this study performed interclass (external) topological validation in a CDGV previously validated in Gothic, a system used for more than 10 years by Brazilian Geographic Service.

From the results of topological consistency, there was a large amount of inconsistencies related to the accuracy of the coordinates, and others to the implementation of the model in the database, as is the case of the Drainage Points class, in addition to a smaller number of other cases that, in fact, violate the rules of the data model.

The results obtained in this work indicate the need to modernize the procedures adopted in cartographic production, in order to document and implement all possible topological rules of the data model to achieve a complete integrity of the CDGV.

On future research aligned with the theme of this work, it is also suggested the study of techniques that allow the real-time validation of the logical rules listed for the data model at the exact moment of acquisition or vectorization, either in the form of construction of "alert" mechanisms, or in the construction of constraints that will prevent the inconsistent data from being stored in the database.

Acknowledgment

To the professionals of the Brazilian Army’s Geographical Service Bureau who participated in the Mapping Project of the State of Bahia for the construction of the geospatial dataset used in this research.

Authors' Contribution

L.L.S. França contributed to the conceptualization, data curation, methodology, analysis, implementation and writing of the text. J.L. Portugal contributed to the supervision and revision of the text.

Conflicts of Interest

There are no conflicts of interest.

References

ARIZA-LÓPEZ, F. J., RODRÍGUEZ-AVI, J., REINOSO-GORDO, J. F., ARIZA-LÓPEZ, Í. A. Quality Control of "As Built" BIM Datasets Using the ISO 19157 Framework and a Multiple Hypothesis Testing Method Based on Proportions. **ISPRS International Journal of Geo-Information**, v. 8, n. 12, p. 569, 2019. DOI. 10.3390/ijgi8120569.

BAUMGART, Bruce G. A polyhedron representation for computer vision. In: NATIONAL COMPUTER CONFERENCE, **Proceedings of the May 19-22, 1975, national computer conference and exposition**. New York, NY, USA ,1975. p. 589-596.

BORGES, K. A. V.; DAVIS, C. A.; LAENDER, A. H. F. OMT-G: an object-oriented data model for geographic applications. **GeoInformatica**, v. 5, n. 3, p. 221-260, 2001. DOI. 10.1023/A:1011482030093.

BORGES, K. A. V.; DAVIS, C. A.; LAENDER, A. H. F. Integrity constraints in spatial databases. In: **Database integrity: challenges and solutions. IGI Global**, 2002. p. 144-171.

BORGES, K. A. V.; DAVIS JR, C. A.; LAENDER, A. H. F. Modelagem conceitual de dados geográficos. In: CASANOVA, et al. **Banco de Dados Geográfico**: Curitiba: MundoGEO , 2005. p. 83-136.

BORRMANN, A.; RANK, E. Topological analysis of 3D building models using a spatial query language. **Advanced Engineering Informatics**, v. 23, n. 4, p. 370-385, 2009. DOI. 10.1016/j.aei.2009.06.001.

CAO, G. Spatial analysis and modeling: Database Fundaments. 06 aug. 2014, 12 dec. 2014. 99 p. **Notas de Aula**. Department of Geosciences. Texas Tech University. 2014.

CASANOVA, M. A., CÂMARA, G., DAVIS JR, C., VINHAS, L., E QUEIROZ, G. R. **Banco de dados geográficos**. Editora Mundogeo. Curitiba-PR. 2005.

CLEMENTINI, E.; DI FELICE, P. A comparison of methods for representing topological relationships. **Information sciences-applications**, v. 3, n. 3, p. 149-178, 1995. DOI. 10.1016/1069-0115(94)00033-X.

COMISSÃO NACIONAL DE CARTOGRAFIA (CONCAR). **Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV)**. Versão 2.1.3. Brasília, 2010. Disponível em: < https://bdgex.eb.mil.br/portal/media/edgv/ET_EDGV_Vs_2_1_3.pdf>. Acesso em: 07 ago 2022.

DAVIS JR, C. A.; LAENDER, A. H. F. Extensões ao modelo OMT-G para produção de esquemas dinâmicos e de apresentação. In: WORKSHOP BRASILEIRO DE GEOINFORMÁTICA, 2000. **Anais do II Workshop Brasileiro de Geoinformática**, São Paulo. 2000.

EGENHOFER, M. J.; SHARMA, J.; MARK, D. M. A critical comparison of the 4-intersection and 9-intersection models for spatial relations: formal analysis. In: R. MCMASTER AND M. ARMSTRONG (EDS), AUTOCARTO 11**. ASPRS American Society for Photogrammetry**, 1993. p. 1-12.

ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE (ESRI). **ArcGIS Resource Center ArcSDE SDK 10 C and Java API – Predicates**. 2010. Disponível em: <https://help.arcgis.com/en/geodatabase/10.0/sdk/arcsde/concepts/geometry/shapes/spatial_relations/predicates.htm>. Acesso em: 12 dez 2020.

FRANÇA, L. L. S.; DE ALMEIDA, A. D. O.; DA PENHA, A. L. T. Avaliação da qualidade dos modelos digitais de elevação Aster e SRTM para o Estado da Bahia. **Revista Brasileira de Cartografia**, v. 69, n. 9, 2017.

FRANÇA, L. L. S.; SILVA, T. A., A., BORBOREMA C.B.A.B., ALCÂNTARA, L.A. Vetorização de Cobertura Terrestre no QGIS. In: SIMPÓSIO BRASILEIRO DE CIÊNCIAS GEODÉSICAS E TECNOLOGIAS DA GEOINFORMAÇÃO. **Anais do VII SIMGEO**. Recife-PE, 2018. p.393-400.

FRANÇA L. L. S.; FERREIRA DA SILVA, L. F. C. Comparison between the Double Buffer Method and the

Equivalent Rectangle Method for the quantification of discrepancies between linear features. **Boletim de Ciências Geodésicas**, v. 24, n. 3, p. 300-317, 2018. DOI. 10.1590/s1982-21702018000300020.

FRANÇA, L. L. S.; DA PENHA, A. L. T.; CARVALHO, J. A. B. Comparison between absolute and relative positional accuracy assessment-a case study applied to digital elevation models. **Boletim de Ciências Geodésicas**, v. 25, n. 1, 2019. DOI. 10.1590/s1982-21702019000100003.

FRANÇA, L. L. S.; PASSOS, J. B.; PORTUGAL, J. L. Topological validation: a study applied for hydrographic features of a watershed. **Ciências Exatas e da Terra: Aprendizado, Integração e Necessidades do País**. 1ed.: Atena Editora, v., p. 191-207. 2020

FRANÇA, L. L. S. **Consultas SQL para Validação Topológica**. GitHub. 2021. Disponível em: <https://bit.ly/3tgMzBf>. Acesso em: 10 fev 2022.

HERRING, J. OPENGIS® **Implementation Standard for Geographic Information-Simple Feature Access**. Part 2: SQL option. 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Avaliação da Qualidade de Dados Geoespaciais**. Manuais Técnicos em Geociências nº 13. 2017. Disponível em: < https://biblioteca.ibge.gov.br/visualizacao/livros/liv101152.pdf>. Acesso em: 07 ago 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **Geographic Information – Services. Quality management systems-requirements (ISO 19.157: 2013)**. 2013. Disponível em: < https://www.iso.org/standard/32575.html>. Acesso em: 07 ago 2022.

LIZARDO, L. E. O.; DAVIS JR, C. A. A PostGIS extension to support advanced spatial data types and integrity constraints. In: SIGSPATIAL'1. **Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems**. Los Angeles, CA: ACM, 2017. p. 33.

MARAŞ, H. H. et al. Topological error correction of GIS vector data. **International Journal of Physical Sciences**, v. 5, n. 5, p. 476-483, 2010. DOI. 10.5897/IJPS.9000598.

PASSOS, J. B.; FRANÇA, L. L. Processo de reambulação no mapeamento topográfico. **Revista Brasileira de Geomática**, v. 6, n. 2, p. 119-138. 2018. DOI. 10.3895/rbgeo.v6n2.6700.

PASSOS, J. B., CARVALHO, R. B., PENHA, A. D. L. T., E FRANÇA, L. L. S. Estruturação e validação de dados geográficos em ambiente orientado a objeto do Sistema Gothic. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, XVIII, 2017. **Anais do XVIII SBSR**, Santos-SP. 2017.

ROMANHOLI, M. P.; DE QUEIROZ, A. P. Base hidrográfica ottocodificada na escala 1: 25.000: Exemplo da bacia do Córrego Itapiranga (SP). **Caminhos de Geografia**, v. 19, n. 68, p. 46-60, 2018. DOI. 10.14393/RCG196804.

SHIELDS, R. Cultural topology: The seven bridges of Königsburg, 1736. Theory, **Culture & Society**, v. 29, n. 4-5, p. 43-57, 2012. DOI. 10.1177/0263276412451161.

SIEJKA, M.; ŚLUSARSKI, M.; ZYGMUNT, M. Correction of topological errors in geospatial databases. **International Journal of Physical Sciences**, v. 8, n. 12, p. 498-507, 2013. DOI. 10.5897/IJPS2013.3835.

STEMPLIUC, Sergio Murilo. **Modelagem de restrições de integridade espaciais em aplicações de rede através do modelo UML-Geoframe**. Dissertação (Mestrado em Ciência da Computação) – Curso de Pós-Graduação em Ciência da Computação, Universidade Federal de Viçosa, Viçosa. 2008.

STOLZE, K. SQL/MM spatial: The standard to manage spatial data in a relational database system. In: BTW 2003**. Datenbanksysteme für Business, Technologie und Web**, Tagungsband der 10. BTW Konferenz. Gesellschaft für Informatik eV, 2003. 247–264.

STROBL C. Dimensionally Extended Nine-Intersection Model (DE-9IM). In: SHEKHAR S., XIONG H., ZHOU X. **Encyclopedia of GIS**. Springer. 2008. DOI. 10.1007/978-3-319-17885-1_298. 470–476

SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA (SEI). **Geoserviços - Baseados em Vetores - Cartografia de Referência - Cartografia 1:25.000**. 2021. Disponível em: <https://www.sei.ba.gov.br/index.php?option=com_content&view=article&id=3346&Itemid=953>. Acesso em: 10 jan 2021.

WADEMBERE, I.; OGAO, P. Validation of GIS vector data during geo-spatial alignment. **International Journal of Geoinformatics**, v. 10, n. 4, p. 17-25, 2014. DOI. 10.52939/ijg.v10i4.576.

Biography of the first author

Leandro Luiz Silva de França obtained his B.E. degree in Cartographic Engineering from the Military Engineering Institute (IME) with a period studied at Texas Tech University (TTU), USA. He obtained his Master's degree in Geodetic Sciences and Geoinformation Technologies from the Federal University of Pernambuco (UFPE). Python developer of QGIS's tools focused on Cartography, Surveying, PostgreSQL/PostGIS databases, Image Processing, and Quality Control of Geospatial Data.