IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 3, June 2024

# Consistency in Spatial Databases

**Mr. Ajay Bhimrao Thorat**
Student, Department MCA
Late Bhausaheb Hiray S S Trust's Hiray Institute of Computer Application, Mumbai, India
University of Mumbai, Mumbai, India
ajaythorat988@gmail.com

**Abstract***: This paper delves into the consistency challenges associated with integrating multiple sets of spatial data in Geographic Information Systems (GISs). The data sets in question, which provide information about the same geographic features, often come from different sources and times, varying in reliability, accuracy, and scale. This leads to multiple spatial representations for these features. The proposed systematic approach begins with breaking down the consistency issue by identifying various consistency classes that can be checked independently. These classes represent a comprehensive set of properties and relationships necessary to fully identify geographic objects in the data sets. Different levels of consistency—total, partial, and conditional—are then proposed for each consistency class, allowing for the integration of data sets without requiring total consistency in every aspect. The second step involves explicitly representing these consistency classes and levels in the system. As an example, a simple structure storing adjacency relationships is given for representing topological consistency. The paper also suggests explicitly representing the consistent knowledge in the data sets (mainly qualitative) and accounting for the uncertainty or ambiguity inherent in the knowledge.*

**Keywords**: Geographic Information Systems

## I. INTRODUCTION

Integrating data in spatial information systems involves merging diverse types of information from various sources, requiring effective matching of similar entities across data sets and ensuring consistency. Spatial information typically comes in different forms from sources like maps, field surveys, photogrammetry, and remote sensing. These data sets, collected at varying scales or resolutions over different periods, often exhibit differences in reliability and may have missing or undefined details. Incompatibilities between data sets may include discrepancies in dimension, shape, and positional accuracy..

For instance, consider the need to store both a schematic representation of an area and a more accurate representation in a GIS. A schematic map, useful as an interactive tourist guide, might omit many objects and lack positional accuracy. However, it maintains the same relative positions and orientations for the objects it includes. Integrating both representations in the geographic database involves handling multiple spatial representations of the same geographic objects. Different representations might be more efficient for answering specific queries in the GIS. Users should have seamless access to these different data sets, understanding the nature and reliability of the information they use to evaluate the accuracy of any results or analyses.

The term "consistent" is often used in research to describe the initial processing required to reconcile data sets, which might involve rectifying geometric distortions, coordinate registration, or reclassification. Integrating geographic information presents many challenges, including combining vector and raster data and checking topological consistency.

This paper proposes a systematic approach to address the consistency issue in integrating spatial information, comprising the following steps:.

- Analyzing different aspects of equivalence between data sets and identifying a range of spatial equivalence classes to be checked in isolation.
- Studying measures of spatial equivalence applicable to each class and proposing different levels of equivalence—total, partial, conditional, and inconsistent. Data sets can then be ranked according to their


DOI: 10.48175/IJARSCT-18978

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, June 2024

consistency in each class and level. This flexibility allows for the integration of data sets without requiring complete consistency in every aspect.

- Explicitly representing these equivalence classes and levels in the spatial database to facilitate data manipulation and provide users with a clear view of the nature of the data.
- Explicitly representing the set of consistent knowledge and the inherent ambiguity or uncertainty.

As an examplethe representation of the topological consistency class is presented using a simple structure that stores adjacency relationships. This approach aims to provide users with a clear understanding of the data sets handled by the spatial information system. Users can either choose the type of data to manipulate or let the system decide which data best suits their application, always knowing the extent of consistent knowledge available and the nature and measure of any inconsistency.

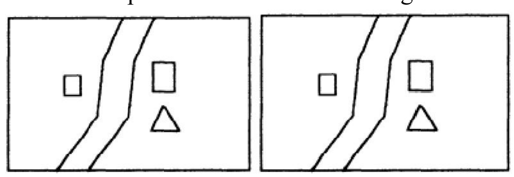
The issues discussed are relevant to GISs and other spatial information systems and applications. This research is part of an ongoing project to develop methods for modeling and manipulating hybrid data sets in a GIS.

The paper is structured as follows: Section 2 identifies different consistency classes. Section 3 proposes different levels of consistency. Section 4 discusses the explicit representation of consistency. Section 5 addresses the representation of uncertainty. Section 6 provides conclusions, and Section 7 lists the references.

## II. ASPECTS OF SPATIAL EQUIVALENCE

In integrating two sets of spatial data which relate to the same area in space, two consecutive steps are needed,

- Object matching: This step identifies corresponding objects in both data sets using spatial equivalence tests. It determines which objects can be considered the same in both sets. For instance, it could involve matching land parcels from an old map with those in a new map or aligning road networks in maps of different scales. Note that these objects might differ in their positional information and geometric structure.



(a) (b)

Figure 1: Existentially non-consistent data sets. Some of the objects in (a) are missing in (b).

Spatial Equivalence representation: his step involves explicitly representing the relationships between the data sets. This allows the system to intelligently manipulate both sets and provide the user with a clear view of the data's nature.

### A. Positional Equivalence

Objects are represented by specific coordinates that describe their spatial extents. Under this reference, two objects from different data sets match only if their coordinates match exactly. Two data sets are considered locationally consistent if any position (x, y, z) corresponds to the same object in both sets.

### B. Object-Based Equivalence Classes

A spatial data set consists of the spatial properties of a set of objects in a defined space. These properties include descriptions of spatial extent, from which the dimension and shape of the object can be derived. An object in the data set can be composite, containing or consisting of other objects. Object-based consistency can be classified using these properties.

#### 1. Object Existence Equivalence

Two data sets are existentially equivalent if all object classes and instances in one data set exist in the other. For example, the two data sets in Fig. 1 are existentially non-equivalent because some objects are missing in one.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, June 2024

**2. Object Dimension Equivalence**

Two data sets are dimensionally equivalent if every object in one set has the same spatial dimension as the corresponding object in the other set. For instance, the two spatial scenes in Fig. 2 are not dimensionally equivalent because objects are represented using different dimensions (e.g., areas by points or lines).

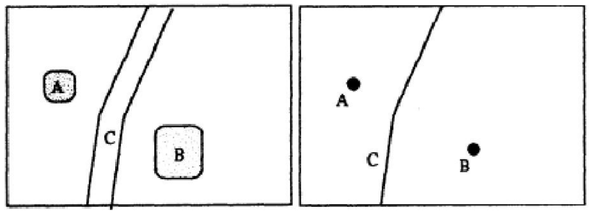



Figure 2: Non-consistent data sets with reference to object dimension.

**3. Object Shape Equivalence**

Shape equivalence can be defined at different levels of accuracy. Strictly, object shapes can be defined using equations of the curve or set of curves defining their boundaries. More flexibly, object shapes can approximate well-known geometric shapes like circles or squares. Two data sets are shape equivalent if each object in one set can be described as shape-equivalent to the corresponding object in the other set. For example, in Fig. 3, object A's shape may be considered equivalent in all three scenes, depending on the acceptable measure of shape distortion in the database.

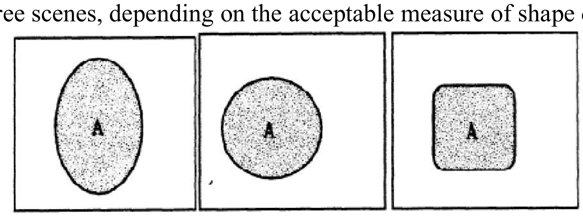



Figure 3: Object A may be considered to have equivalent shape in all three sets according to the allowed shape distortion.

**4. Object Size Equivalence**

Size can be measured by boundary length, area, and volume. Two data sets are size equivalent if every object in one set has a similar size to its corresponding object in the other set.

**5. Spatial Detail Equivalence**

Composite objects may contain several connected or non-connected objects. Two data sets are detail-equivalent if corresponding composite objects in both sets are considered equivalent, as shown in fig.4.

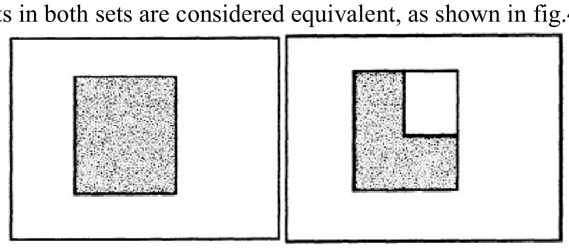


Figure 4: Non-consistent data sets with reference to object details

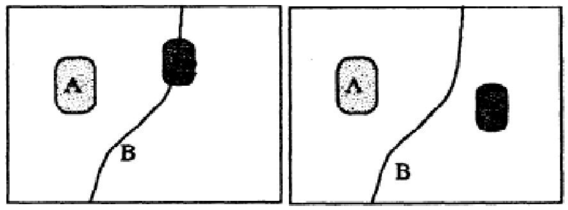
**C. Relation-Based Equivalence Classes**

This type of consistency measures the spatial relationships between objects in the data sets. Three classes of equivalence are classified according to spatial relationships:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, June 2024

**1. Topological Equivalence**

Two data sets are topologically consistent if the topological relationships derived from one set are the same as those derived from the other. For example, the two sets in Fig. 5 are not topologically consistent because object B's relationship with object C differs between the sets

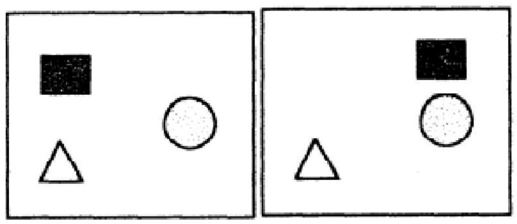



(a) (b)

Figure 5: Topological inconsistency. (a) Object B crosses object C. (b) Object B is disjoint from C.

**2. Direction or Orientation Equivalence**

Two data sets are directionally consistent if the relative directional relationships in one set are the same as in the other. For example, the two sets in Fig. 6 are not directionally consistent.

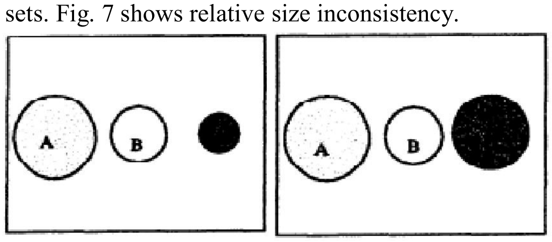


(a) (b)

Directional inconsistency.

**3. Relative Size Equivalence**

Two data sets are size-consistent if the qualitative size relationships (e.g., larger or smaller) are maintained between corresponding objects in the two sets. Fig. 7 shows relative size inconsistency.




(a) (b)

Relative size inconsistency. A > B > C in (a) and A>B

**Interdependency between Equivalence Classes**

Classes of object-based equivalence may exist. The ones listed above are among the most important from a general perspective. These classes are not mutually exclusive. For example, positional consistency implies every other type of consistency (except categorical) and is the strictest measure of spatial equivalence. Shape and size equivalence imply dimension equivalence, and all equivalence classes imply existence equivalence. Shape equivalence may imply spatial detail consistency if the object is composed of non-connected sets. Additionally, a certain degree of inaccuracy is acceptable in measuring some properties, like size and shape, depending on the application's requirements.

## III. DIFFERENT LEVELS OF SPATIAL CONSISTENCY

When integrating two geographical data sets that pertain to the same spatial area, they can exhibit consistency across multiple defined classes. For instance, they may be consistent topologically and dimensionally, or in terms of

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, June 2024

dimension, detail, and category. Some types of consistency inherently assume others. For example, achieving topological equivalence may imply spatial detail consistency.

Up until now, our discussion has focused on a single level of consistency, where all objects in the data sets adhere to the studied consistency class. However, this ideal scenario is not always feasible. Hence, it's crucial to rank the level of consistency across different classes. This ranking provides GIS users with an initial understanding of the nature of the data sets they are working with. Further analysis involves identifying the extent and nature of this consistency within the data stes.

Let's denote S1 and S2 as the sets of knowledge present in two data sets. This knowledge encompasses various types of information derived from each data set, classified according to object-based and relation-based classes. Consider S1i and S2i as subsets of knowledge from S1 and S2, respectively, belonging to a specific class i (such as shape properties, directional or topological relationships)

Four distinct levels of consistency can be identified:

**A. Total Consistency '**

Two data sets S1 and S2 are completely consistent with respect to a certain class i if $S1i \cap S2i = S1i \cup S2i$, meaning S1i equals S2i. Queries to the GIS pertaining to properties of class i would yield identical results from either S1 or S2.

**B. Inconsistency Level**

S1 and S2 are inconsistent regarding a class i if $S1i \cap S2i = \emptyset$, indicating they share no knowledge from that class. Queries to the GIS concerning properties of class i would yield differing results from S1 and $S_2$.

In practice, data sets needing integration typically involve a combination of classes and levels of consistency. For example, two data sets may exhibit partial consistency in shape and dimension but achieve total consistency topologically. They might also be conditionally consistent regarding object detail and partially consistent topologically.

## IV. REPRESENTATION OF DIFFERENT LEVELS OF CONSISTENCY FOR DIFFERENT CLASSES

Determining the class and level of consistency between two data sets involves extracting and comparing the set of properties or relationships for that class. While it's beneficial for users and systems to understand the general class and level of consistency, explicit representation of the consistent set of knowledge becomes necessary in specific application domains.

Upon closer examination, these consistency classes primarily involve qualitative measures (except for location, size, and shape). Therefore, the shared spatial knowledge between data sets can be qualitatively represented. A structural mechanism can be devised to apply to geographic data sets, explicitly representing qualitative properties and relationships and deriving others as needed. This structure operates at the geographic object level rather than the geometric representation level, accommodating multiple spatial representations of the same geographic objects.

Manipulating such a qualitative structure enables spatial reasoning. For example, storing some topological relationships while deriving others using composition tables for similar and mixed types of spatial relations..

Explicitly representing this knowledge facilitates comparison between data sets, seamless manipulation of existing sets, integration of new sets, and consistent updates of existing ones.

## V. CONCLUSION

This paper examines the nature of consistency issues in integrating hybrid data sets, focusing specifically on integrating multiple spatial data sets for the same spatial area.

**Key Points of the Approach:**

- Consistency between data sets is categorized into comparing fundamental object properties and relationships between these objects.
- Nine consistency classes are identified under these categories, each independently checkable.

- Each class can exhibit varying levels of consistency—total, partial, conditional, and inconsistent. Data sets can be ranked accordingly, such as achieving total topological consistency but only partial consistency in object dimensions.
- Explicitly representing different classes and levels of consistency provides a realistic view of database contents.
- The shared set of consistent knowledge within data sets should be explicitly expressed. A proposed qualitative structure manages different types of geographic feature knowledge at the object level, with explicit representation of ambiguity or uncertainty inherent in individual data sets and resulting from their integration.

## REFERENCES

**[1].** *Abdelmoty, A. I., & Jones, C. B. (Year). Towards Maintaining Consistency of Spatial Databases.* Dept. of Computer Studies, University of Glamorgan, Pontypridd, Mid Glamorgan, CF37 lDL, Wales, UK