UNIVERSIDAD DE JAÉN
ESCUELA POLITÉCNICA SUPERIOR
DEPARTAMENTO DE INGENIERÍA CARTOGRÁFICA, GEODÉSICA Y FOTOGRAMETRÍA

TESIS DOCTORAL

AUTOMATIC EVALUATION OF GEOSPATIAL DATA QUALITY USING WEB SERVICES

PRESENTADA POR:
EMERSON MAGNUS DE ARAÚJO XAVIER

DIRIGIDA POR:
FRANCISCO JAVIER ARIZA LÓPEZ
MANUEL ANTONIO UREÑA CÁMARA

JAÉN, JUNIO DE 2017

ACKNOWLEDGEMENTS

We wish to acknowledge the support of the Brazilian Army's Department of Science and Technology, and the support of the Regional Government of Andalucía (Spain) for the Research Group 'Ingeniería Cartográfica' (TEP-164).

ABSTRACT

The geomatics sector is going through a data overload scenario which new geospatial datasets are generated almost daily. However there are few or nothing information about the quality of these datasets, and they should be evaluated aiming to provide users some information about their quality. In this context we propose a solution for the automatic quality evaluation of geospatial datasets using the web services platform. This approach is compound by automatic evaluation procedures for quality control of topological consistency, completeness, and positional accuracy described in the Brazilian quality standard. Some procedures require an external dataset for comparison purposes. Hence we provide a set of synthetic datasets and apply over them an experimental design aiming to select suitable methods to find the correspondences between datasets. The solution has an interoperability tier that links users and automatic procedures using the standardized interface of Web Processing Services (WPS).

Keywords: automation, quality control, matching, WPS

Evaluación automática de la calidad de los datos geoespaciales mediante servicios web

El sector geomático vive un escenario de sobrecarga de datos donde casi todos los días se generan nuevas bases de datos geoespaciales (BDG). Sin embargo, hay poca o ninguna información sobre la calidad de estas BDG. En este contexto proponemos una solución para la evaluación automática de la calidad de los datos geoespaciales utilizando servicios web. Este enfoque está compuesto por procedimientos de evaluación automática para el control de calidad de la consistencia topológica, compleción y exactitud posicional según se especifican en el estándar brasileño. Algunos procedimientos de control requieren datos externos para fines de comparación. Por ello, en este trabajo, proporcionamos un conjunto de datos sintéticos generados según un diseño de experimentos con el objetivo de seleccionar los métodos más adecuados para encontrar correspondencias entre las BDG. La solución desarrollada tiene un capa de interoperabilidad que vincula usuarios y procedimientos automáticos utilizando la interfaz del Web Processing Services (WPS).

Palabras clave: automatización, control de calidad, casado, WPS

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

To date there are lots of geospatial data sources available to generate data almost instantaneously. Imagery from aerial or satellite platforms, and the popularization of Unmanned Aerial Vehicle (UAV), or 'drones', has allowed to generate geospatial datasets in an unmanageable way, what some authors named 'big data' trend (Crampton et al. 2013). 'Terabytes are quite typical today' said Traxler and Hesina (2017). Other important data source is the crowdsourced data, generated by volunteers almost daily (Neis and Zielstra 2014). This overload data scenario brings new challenges for the official spatial data suppliers, or National Mapping and Cadastral Agencies (NMCA). Traditionally, these institutions create and manage authoritative datasets in a standardized way. However, today many data 'producers' represent the same phenomena, geospatial features, following their own rules. This new scenario may lead users questioning the quality of available datasets.

In these cases, few or nothing information about the quality of a spatial dataset is available, so we believe that would be interesting a web service with the capability of assess the quality of a test dataset against a reference dataset. A data quality validation service is an appealing topic in the geospatial research agenda which has been developed in current projects (Kruse 2014).

A recent trend in the geomatics industry is to automatise the most of the productive chain, as we can see in recent projects, e.g. the 'mapping as a service' in Ordnance Survey Ireland (Coumans 2016) and the use of UAV in cadastral mapping (Ramadhani et al. 2016). It is fair to assume that data quality evaluation also experiences this trend.

The state-of-the-art for the automation of quality control for spatial data has shown recent advances. The study of Donaubauer et al. (2008) proposed a web service with the ability to generate quality information of assessed data via web services. The work used well-defined standards when was executed, with the Web Processing Service (WPS) (Schut 2007) interface to process the quality control, and ISO 19115 (ISO 2003b) for the quality report by means of metadata elements. WPS is an open specification from the Open Geospatial Consortium (OGC). Despite the simplicity of the quality procedure, just an overlay of previously tagged data with some quality elements, this study seemed to be the first attempt of an automatic evaluation service in the literature. Other study also indicated that the quality evaluation can be executed through a WPS (Mobasheri 2013). More recently, Meek et al. (2016) presented a solution for quality evaluation of crowdsourced data using service chaining and WPS. This kind of data is also referred as Volunteered Geographic Information (VGI, Goodchild 2007).

Looking for the development projects around the European Spatial Data Infrastructure initiative, INSPIRE, we may identify a semi-automatic evaluator service as an aim of the ESDIN Project (European Spatial Data Infrastructure Network) (Beare et al. 2010). The authors presented the concepts of semi-automatic evaluation services for data quality control. Deserves mention this work pointed out many gains for both users and producers of spatial data while using such web services. The report of Portele (2011), also in the context of the ESDIN Project, introduced quality control test cases using the Radius Studio tool for on-line procedures.

In the Universidad de Jaén emerged a successful research focused on the automation of the positional accuracy evaluation, due to Ruiz-Lendínez (2012). The author proposed a solution for automatic positional accuracy assessment of polygonal features using a matching approach. His thesis presented encouraging results and this is our starting point for the current research.

The free availability of some geospatial information for final users has raised questions about the cost of maintenance for NMCAs (Carpenter and Snell 2013). However, as more data is available for users, more becomes necessary evaluate their quality in order to identify if these data fits the users' requirements. This may be the opportunity for an authoritative data supplier plays the role of data 'validator', providing standardized and useful quality reports about the data users want. Other possibility is the raising of quality certification for geospatial data, as pointed by Ariza-López (2013).

1.1. Research question, hypothesis, and goals

Our main research question arises from the need of an on-line evaluation service: how far can we automate the evaluation of geospatial data quality over a web environment?

The current state-of-the-art of the automation of quality control for spatial data shows some recent research:

- It is possible to generate quality information about a spatial dataset using the WPS interface (Donaubauer et al. 2008), also confirmed by a later work (Mobasheri 2013);

- Studies inside the ESDIN project described semi-automatic data quality evaluation services (Beare et al. 2010);
- Ruiz-Lendínez (2012) proposed and demonstrated a feasible solution for automation of the positional accuracy evaluation using a matching approach;
- Ariza-López (2013) argued that quality of spatial data has received recent and continuous development of international standards, notably the ISO 19157:2013 (ISO 2013);
- Fan et al. (2014) demonstrated that is possible evaluating various quality elements - completeness, positional accuracy, thematic accuracy, and shape accuracy - for building footprints using a test dataset against a reference one, where the first step was the matching between datasets; and
- Brovelli et al. (2017) presented a new procedure to perform comparisons between crowdsourced and authoritative road datasets with a significant degree of automation.

Taking these facts as working assumptions we can formulate our working hypothesis H1: A fully automatic evaluation procedure is possible, without any human intervention, that assesses a test dataset against a reference dataset. We believe a data quality validation service will bring gains for data producers and data consumers. Data producers may benefit themselves by a standardized dataset to evaluate their own contracted products. Data consumers may obtain a quality report for the spatial data using a standard protocol (WPS).

Considering this hypothesis, our main goal is to develop a web service able to evaluate the quality of geospatial datasets using the standardized interface of WPS in a fully automatic way. Starting from this main goal, we can describe our secondary aims:

- A1: Implement a generic WPS server that supports the evaluation service;
- A2: Develop a generic architecture able to accommodate any quality evaluation procedure over a web environment;
- A3: Develop a matching approach able to compare two datasets by combining different techniques found in literature;
- A4: Use a common evaluation framework, if available, or develop a simple quality model as a proof of concept, all based on the new ISO standards; and
- A5: Execute quality evaluation procedures from quality elements: positional accuracy, completeness, and logical consistency.

1.2. Institutional relevance and publications

This study was developed in the research group GIIC (Grupo de Investigación en Ingeniería Cartográfica) at the Universidad de Jaén, Spain. This research group (TEP-164) has produced relevant studies in the geospatial data quality area, among them we can cite: Ariza López and Atkinson Gordo (2008), Ariza-López et al. (2011), Ariza-López and Mozas-Calvache (2012), Ariza-López and Rodríguez-Avi (2014), Ruiz-Lendínez et al. (2016) and Gil de la Vega et al. (2016).

The research project was supported by the Brazilian Army's Department of Science and Technology (DCT), which sponsored this project on behalf of the Geographic Service (DSG). DSG leads the geospatial information in the Brazilian Army. According to Brazilian law (Brasil 1967) DSG is responsible to generate and maintain the technical standards for the national land mapping.

This research project generated the following publications at the present date:

- Banco de dados geográficos do Exército Brasileiro: arquitetura e resultados. In V Jornadas Ibéricas de Infra-estruturas de Dados Espaciais, Lisbon, Portugal (Xavier et al. 2014);
- Proposal of a web service for positional quality control of spatial data sets. In International Workshop on Spatial Data and Map Quality, Valletta, Malta (Ariza-López et al. 2015);
- Aplicações, tendências e desafios em infraestruturas de dados espaciais. In Bahia Análise & Dados, 25:4 (Xavier et al. 2015a);
- Web service for positional quality assessment: the WPS tier. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W5, La Grande Motte, France (Xavier et al. 2015b);
- WPS for positional quality control applying the method proposed in UNE 148002. In VI Jornadas Ibéricas de Infraestructuras de Datos Espaciales, Sevilla, Spain (Xavier et al. 2015c);
- A survey of measures and methods for matching geospatial vector datasets. In ACM Computing Surveys, 49:2 (Xavier et al. 2016a); and
- Using third party data to update a reference dataset in a quality evaluation service. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B4, Prague, Czech Republic (Xavier et al. 2016b).

We have submitted other two studies to consideration:

- Métodos de evaluación de la calidad posicional en Hispanoamérica: análisis de la situación (Ariza-López et al. 2017); and
- MatchingLand, geospatial data testbed for the assessment of matching methods (Xavier et al. 2017).

1.3. Structure of the document

This thesis is composed of six chapters. The remainder of this document is structured as follows: In Chapter 2, we present the methodological background and related studies that serve as basis for this research. In Chapter 3 we develop our proposal for a framework for automatic data quality evaluation of geospatial data. This framework is composed by the architecture of a solution towards quality assessment through web services. Chapter 4 brings the experiments executed in order to validate our approach. The essays are designed to assess the proposed framework using both real and synthetic data. Conclusions are drawn in Chapter 5, as well as the suggestions for future work. Finally, Chapter 6 introduces an extended abstract that provides an overview of this study in the Spanish language.

CHAPTER 2
RELATED WORK

This chapter presents the methodological background and related studies that serve as basis for this research. It is divided into five sections.

The main goal of this study is to develop a quality control service for geospatial data, so the first section presents the concepts related to the geospatial data quality evaluation. Some quality assessment procedures require an external dataset as a reference for comparison purposes. The way to find the correspondences between this reference and a test dataset is called matching. As we pointed in a previous work (Xavier et al. 2016a), the matching methods are closer related to the similarity measures. Taking into account these relationships, the Section 2.2 presents the similarity measures and the Section 2.3 describes the matching methods.

It is possible to note that there are a plethora of techniques facing the geospatial matching issue. Xavier et al. (2016a) indicated that the methods can use only one measure, or multiple measures, which can be combined in many different ways, like a normalized score, a weighted combination, the probabilistic theory, optimization processes, the belief theory, genetic algorithms, or regression model. In order to find a suitable matching solution to our quality control service we used the concepts of experimental design, which is briefly explained at Section 2.4.

Finally, Section 2.5 illustrates the concepts of geospatial web services, the technological platform on which we implement the quality control procedures.

Sections 2.2 and 2.3 are extended versions of a review published in Xavier et al. (2016a). In this same line, Section 2.5 is an extended version of a study about web processing services previously published in Xavier et al. (2015b).

2.1. Quality evaluation of geospatial data

The geospatial data quality evaluation has experienced a highlighted attention in the last years. In 2013, the International Organization for Standardization (ISO) got its data quality standards – ISO 19113, ISO 19114, and ISO 19138 (ISO 2002, 2003a, 2006) – which had been used for ten years, and replaced them for a single new standard ISO 19157 (ISO 2013). In 2014, the American Society for Photogrammetry and Remote Sensing (ASPRS) modernized its positional standard from ASPRS 1990 to ASPRS 2014 (ASPRS 2015). This later version was readly adopted by some American government institutions, like the Army Corps of Engineers (USACE 2015), and the Federal Emergency Management Agency (FEMA 2016). In 2015, the committee 148 of the Spanish Association for Standardisation and Certification (AENOR) released a brand new positional control standard named UNE 148002 (AENOR 2015).

Following this context of changing, in 2016 the Brazilian Army published the first version of the Brazilian standard for geospatial data quality, named CQDG (DCT 2016a). Taking into account that this standard provides quality evaluation procedures for all geospatial data products in Brazil, this standard plays the role of quality model in this research project (aim A4).

In the following subsections we present the ISO's model to describe the quality of geospatial data since it is the basis of the Brazilian standard (Section 2.1.1). Section 2.1.2 briefly describes the Brazilian standard, what is adopted here as the quality model. The last section (Section 2.1.3) introduces some studies facing the automation of geospatial data quality assessment.

2.1.1. Describing the quality of geospatial data

The ISO 19157 is an International Standard that establishes the principles to describe the quality of geospatial data by means of: (1) defining components (which includes data quality elements); (2) specifying a structure for data quality measures; (3) describing general procedures; and (4) establishing principles for reporting data quality (ISO 2013).

Figure 2.1 illustrates an overview of the components of data quality according to the ISO's standard. In this model, the quality of geospatial data is described using a *data quality unity*, which is the combination of a *scope* and its associated *elements*. Scope identifies which part of data (e.g. whole dataset, spatial/temporal extent, data layers etc.) will be assessed in each data quality element. The scope can also be related to a dataset series. In the CQDG, the more common scopes are the whole dataset and a subtype by kind of geometry (e.g. all points, all lines).

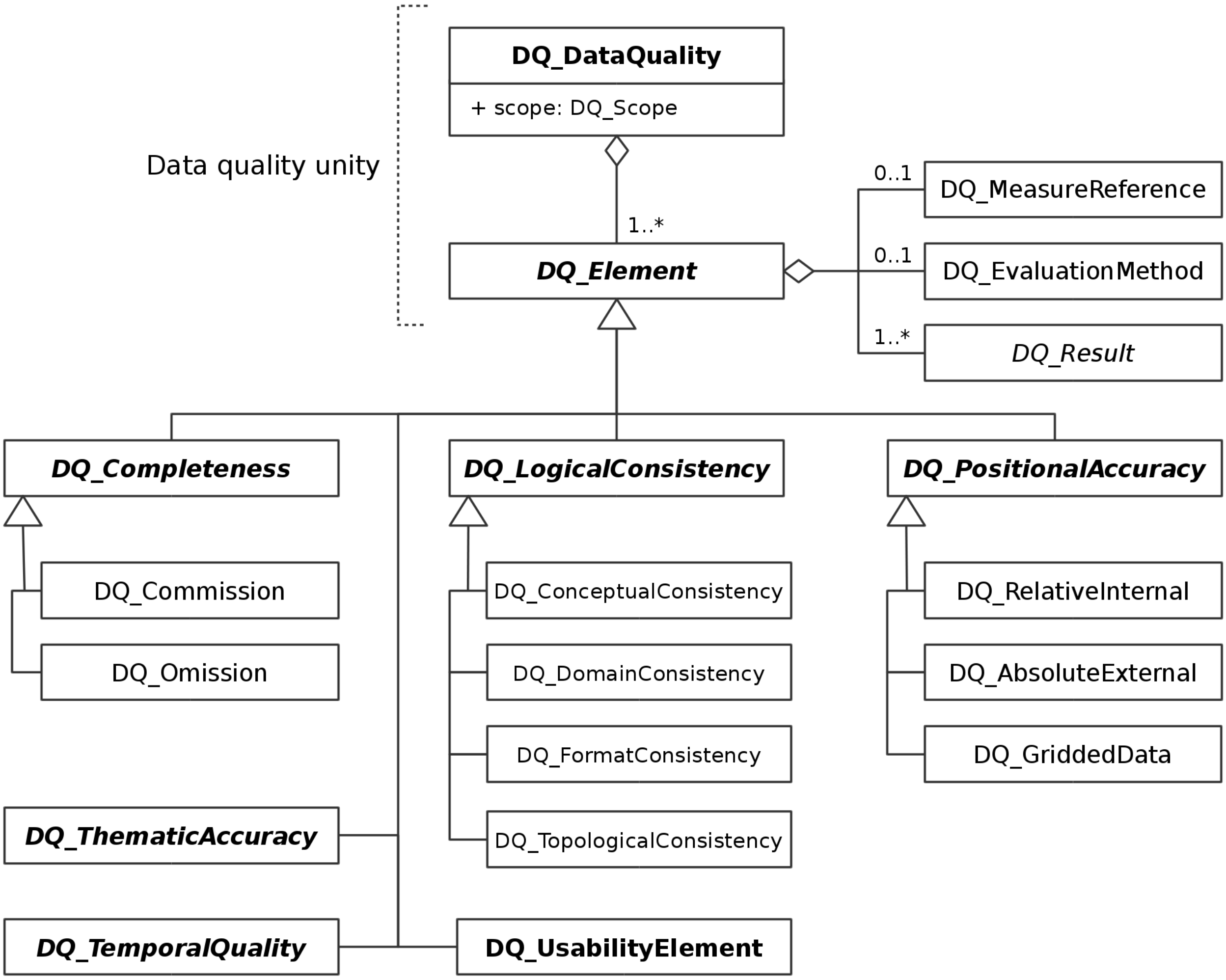


Figure 2.1. Overview of data quality components. Adapted from ISO (2013).

A data quality element describes a specific aspect of the assessed scope using a *measure*, an *evaluation method*, and always present a *result*. Data quality measure is a standardized way to generate comparable results in quality assessment (ISO 2013). An evaluation method describes the process to obtain a data quality result for a specific data quality unity using some measure. Evaluation methods can be classified into three types:

- Direct internal: evaluates the data quality based on inspection of items in a dataset;
- Direct external: the same as direct internal, but comparing the items in the test dataset against items in a reference dataset;
- Indirect: evaluates the data quality based on outer knowledge, which can be subjective.

Data quality elements can be grouped into six categories (see Figure 2.1): completeness, logical consistency, positional accuracy, thematic accuracy, temporal quality, and usability element (ISO 2013).

Completeness refers to the presence or absence of features in a dataset (ISO 2013). This category has two quality elements: *commission* and *omission*. Commission refers to the presence of features that should not be in the dataset. Omission refers to the occurrence of missing features in the assessed dataset.

Logical consistency is a category of quality elements that assesses the degree of fidelity of a test dataset in relation to a set of logical rules (ISO 2013). This category group four quality elements: conceptual consistency, domain consistency, format consistency, and topological consistency. The first two elements refer to the degree of adherence to the conceptual schema and its value domains. Format consistency refers to the obedience to data storage rules. Topological consistency refers to the correctness of topological relations in the test dataset.

Positional accuracy refers to the integrity of data position in a spatial reference system (ISO 2013). Ariza-López and Rodríguez-Avi (2014) pointed that the positional accuracy is an issue of renewed interest owing to the raising of spatial data infrastructures and their interoperability needs. This category gather three quality elements: (1) absolute or external accuracy – compare the coordinates with an external reference (which supposes to being true); (2) relative or internal accuracy – compare the coordinates inside the dataset (only internal comparisons); and (3) gridded data position accuracy – refers to grid data position relative to an exterior reference.

Thematic accuracy refers to the closeness of tested quantitative attributes, the correctness of non-quantitative attributes, and the classification of geospatial objects (ISO 2013). Temporal quality deals with the accuracy, consistency, or validity of temporal attributes and temporal relationships. Lastly, the usability element is based on user requirements that cannot be described using the other five categories (ISO 2013). Thematic accuracy, temporal quality, and usability element will not be used in this study.

The ISO 19157:2013 standard introduces the metaquality for reporting the 'quality of quality'. Metaquality describes the quality of the performed data quality results according to some characteristics (ISO 2013). Inside the data quality components, the metaquality works as a quality element. There are few references in the literature regarding the evaluation of metaquality of geospatial data. For instance, the Spanish standard UNE 148002 (AENOR 2016) deals with metaquality for positional assessment.

The ISO technical committee 211 (TC 211) had been working in data quality standard from more than fifteen years. Since its first quality standards, like ISO 19113 (ISO 2002), the TC 211 continues raising new standards, like ISO/TS 19157-2:2016 (ISO 2016). These standards serves as basis for some national mapping standards, for instance the Brazilian standard that is presented in the next section.

2.1.2. Brazilian standard (CQDG)

The international standard ISO 19157 (ISO 2013) indicates that this standard can be used for defining data quality conformance levels in data product specifications based on the other standard ISO 19131 (ISO 2007). The Brazilian Army published a national standard for geospatial data product specifications based on the ISO 19131:2007, named PCDG (*Produtos de Conjuntos de Dados Geoespaciais* – geospatial data products) (DCT 2016b). In this context arises the Brazilian standard for geospatial data quality, named CQDG (*Controle de Qualidade de Dados Geoespaciais* – quality control of geospatial data) based on ISO 19157:2013 (DCT 2016a). The main aim of CQDG is providing a standardized way to assess the quality of geospatial data products created for the national mapping system, and to inform this quality results.

CQDG standard has five secondary aims: (1) establishes a common vocabulary relative to geospatial data quality; (2) defines a set of measures for quality assessment; (3) describes quality evaluation methods (which includes sampling schemes); (4) defines how reporting the quality results; and (5) establishes the conformance levels for the data products defined in the PCDG standard (DCT 2016b). The standard is divided into three main parts: measures, quality evaluation, and quality report.

The standard defines measures and evaluation methods in five categories from ISO 19157: completeness, logical consistency, positional accuracy, temporal quality, and thematic accuracy. CQDG does not deal with usability element nor metaquality. Figure 2.2 illustrates an example of how a quality measure is presented in CQDG standard using the structure defined in ISO 19157:2013.

In the quality evaluation chapter, CQDG sets up conformance levels for data products and defines the sampling scheme. There are conformance levels and their corresponding quality evaluation procedures for the following kinds of data products: vector datasets (small and large scales); topographic sheets (small and large scales); ortoimagery mapping sheets (small and large scales); digital elevation models; and orthoimagery. In this study we focus only on vector datasets. CQDG adopts a mixed sampling scheme, which the sampling sizes are determined in function of spatial tesselation built over the assessed product, and the sampling procedures for inspection by attributes defined in ISO 2859-1 (ISO 1999), for lot-by-lot inspection, and ISO 2859-2 (ISO 1985), for isolated lot inspection (DCT 2016a). This mixed sampling scheme is explained in details in Section 3.4.

The last part of CQDG standard is dedicated to define how to report the results of quality assessment using standard metadata in XML (eXtensible Markup Language), or by means of a standalone quality report. It is important to note that XML metadata follows the rules of former standard ISO 19115 (ISO 2003b). This occurs because the Brazilian metadata profile, name Perfil MGB (CONCAR 2008), was based on ISO 19115:2003 in the version when the CQDG was launched. Also, the updated ISO standard for encoding data quality reports in XML was published later: ISO/TS 19157-2 (ISO 2016).

Tabela 5 - Medida porcentagem de itens ausentes.

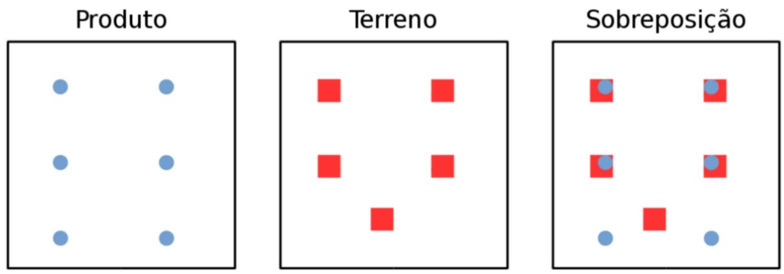
Linha	Componente	Descrição
1	Nome	Porcentagem de itens ausentes
2	Nome alternativo	-
3	Elemento de qualidade	Omissão
4	Medida básica	Taxa de erro
5	Definição	Proporção de objetos ausentes no conjunto avaliado em relação à quantidade de objetos que deveriam estar presentes
6	Descrição	Para uma certa amostra de teste, conta-se a quantidade de objetos presentes no universo de discurso que não possuem correspondente no conjunto avaliado. O resultado da medida é a proporção entre os objetos ausente e a quantidade total de objetos no universo de discurso expressa como porcentagem ([0, 100%]).
7	Parâmetro	-
8	Tipo de valor	Real (porcentagem)
9	Estrutura do valor	-
10	Referência da medida	-
11	Exemplo	Considerando a situação da figura:  Percebe-se que há um objeto na referência (terreno) que não encontra correspondência no produto. Logo a medida retorna: $Medida=\frac{itens\,em\,falta}{tamanho\,da\,referência}=\frac{1}{5}=20\%$
12	Identificador	CQDG:103

Figure 2.2. Example of measure definition in CQDG standard (DCT 2016a).

2.1.3. Automatic evaluation of data quality

The automatic evaluation of data quality has been investigated in diverse research fronts, as we can see in a study about the quality of article revisions in all English Wikipedia (Javanmardi and Lopes 2010), or in the automatic evaluation of syntactic and semantic quality of thesauri (Lacasta et al. 2016). Regarding GISciences, most of approaches for automatic evaluation of data quality are related to matching or conflation studies (Ruiz-Lendínez 2012).

Among these studies we can cite the papers of Mascret et al. (2006), Koukoletsos et al. (2012), and Fan et al. (2014), these last two focused on VGI quality. Mascret et al. (2006) developed new measures adapted to sinuous coastlines in order to automatically assess if a satellital platform would be suitable for coastline mapping. Koukoletsos et al. (2012) developed a new feature matching method to align VGI and authoritative data in order to assess the completeness of the first. Fan et al. (2014) presented an approach to automatically evaluate the quality of 2D building data available in OpenStreetMap (OSM). The authors assessed diverse quality elements of building footprint data, as completeness, semantic accuracy, positional accuracy, and shape accuracy.

In our Research Group emerged a successful research focused on the automation of the positional accuracy evaluation, due to Ruiz-Lendínez et al. (2013). The authors proposed a solution for automatic positional accuracy assessment of polygonal features using a matching approach. The proposed methodology was able to increase significantly the number of features used in the quality evaluation procedure.

In a more recent paper, Brovelli et al. (2017) presented that the research community developed some automated and semi-automated methods to evaluate the quality of VGI data against authoritative datasets, both governmental or commercial datasets. The authors proposed a new method with a significant degree of automation that permits compare OSM road datasets with other road network datasets.

The literature indicates that all automatic evaluation methods have a common feature: they are not error-proof, mainly for those that compare two datasets. Yet, it is possible to recognize full automatic quality procedures in those that are intrinsically direct internal, like some procedures for the quality element logical consistency (e.g. Mobasheri 2013, Tagg 2015).

2.2. Similarity measures

The matching between geospatial data plays an important role in the quality control service developed in this study. These matching methods requires similarity measures in order to evaluate whether two geospatial datasets are similar or not. Similarity measures define some objective measurements for eliminating, or at least mitigating, the uncertainty inherent to this process. Measures applied to the matching process have received different classifications throughout the years. Samal et al. (2004) classify them into context dependent and context independent measures. Frontiera et al. (2008) proposed that spatial similarity is a function of metric, topological and directional functions. Based on the complementary classifications of Tong et al. (2009) (geometric, spatial relationship, attribute) and Zhang et al. (2012) (geometric, semantic, contextual), we organize the matching measures according to the nature of the measured quantity: geometry, topology, attributes, context, and semantics.

Geometric measures refer to the location of features, as well as area overlap, geometric properties (size, area, etc.), and shape (e.g. elongation). Topologic measures assess the relationships between two near features. Attribute measures evaluate the non-geometric properties of a spatial object (e.g. name). Context measures investigate the geographic context of a feature relative to its neighbourhood. Finally, semantic measures refer to concepts, frequently associated to an ontology. Figure 2.3 illustrates the proposed taxonomy for similarity measures.

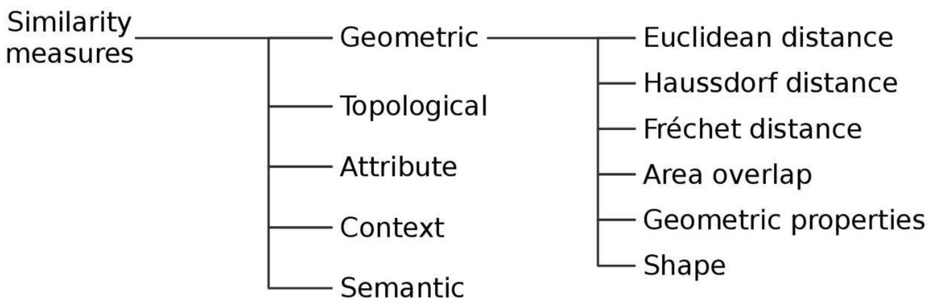


Figure 2.3. The taxonomy of similarity measures (Xavier et al. 2016a).

In this study we use only geometric and context measures. So, in the following sections we present the related work on similarity measures of these two classes.

2.2.1. Geometric measures

The most common types of measure used in geospatial object matching are the geometric measures, due to the spatial nature inherent to this kind of data. Geometric measures refer to the location and shape of a feature, as well as the associated geometric properties (e.g. length) and area overlapping. Some geometric measures take into account the absolute position of the feature, like the distance methods (Euclidean, Hausdorff, Fréchet) and the area overlapping methods. Geometric properties are measures that quantify some peculiar geometric characteristics, like the length of a polyline and the area of a polygon. Shape measures assess the geometric form of objects, notably for areal features. The following subsections present geometric measures beginning with distances, then area overlap, geometric properties, and shape.

2.2.1.1. Euclidean distance

As Deng et al. (2007) point out, the most commonly used distance in GIScience is defined by Euclidean geometry and Cartesian coordinates. Euclidean distance is a widely-used approach to matching point data in the related literature, as can be seen in Yuan and Tao (1999), Beeri et al. (2004), Mustière and Devogele (2008), and McKenzie et al. (2014). However, this measure is limited to calculating unambiguously only the distance from point features. If another kind of geometry is used, including a point set, it is necessary use other concepts, like minimum or maximum distance or centroid distance, as shown by Deng et al. (2007).

2.2.1.2. Hausdorff distance

According to Rucklidge (1996), given two point sets A and B, the Hausdorff distance between A and B is defined by the equation:

$$d_H(A,B) = max\{d_h(A,B), d_h(B,A)\} \quad (2.1)$$

where $d_h(A,B) = \max_{a \in A}\left\{\min_{b \in B}\|a-b\|\right\}$; max{●} represents the maximum value; min{●} represents the minimum value; and ||●|| some underlying norm defined on the plane. The most usual cases adopt the L_2 (Euclidean) norm (Rucklidge 1996, Jones et al. 1999, Li and Goodchild 2010). The undirected Hausdorff distance, $d_H(A, B)$, is the maximum of $d_h(A, B)$ and $d_h(B, A)$, while $d_h(A, B)$ is called the directed Hausdorff distance (Rucklidge 1996). Other authors termed $d_h(A, B)$ and $d_h(B, A)$ forward and backward Hausdorff distances of A to B, respectively (Deng et al. 2007). The resulting values of these two distance functions are often not equal.

The Hausdorff distance (HD) can be used to measure the distance not only between point sets, but also between all geometric primitives and their compositions (multi-geometries). Figure 2.4 illustrates the versatility of this measure in dealing with different representations. HD is a popular measure in GIS used for line-line matching (Yuan and Tao 1999, Chen and Walter 2010) and polygon-polygon matching (Jones et al. 1999, Göesseln and Sester 2004, Huh et al. 2011).

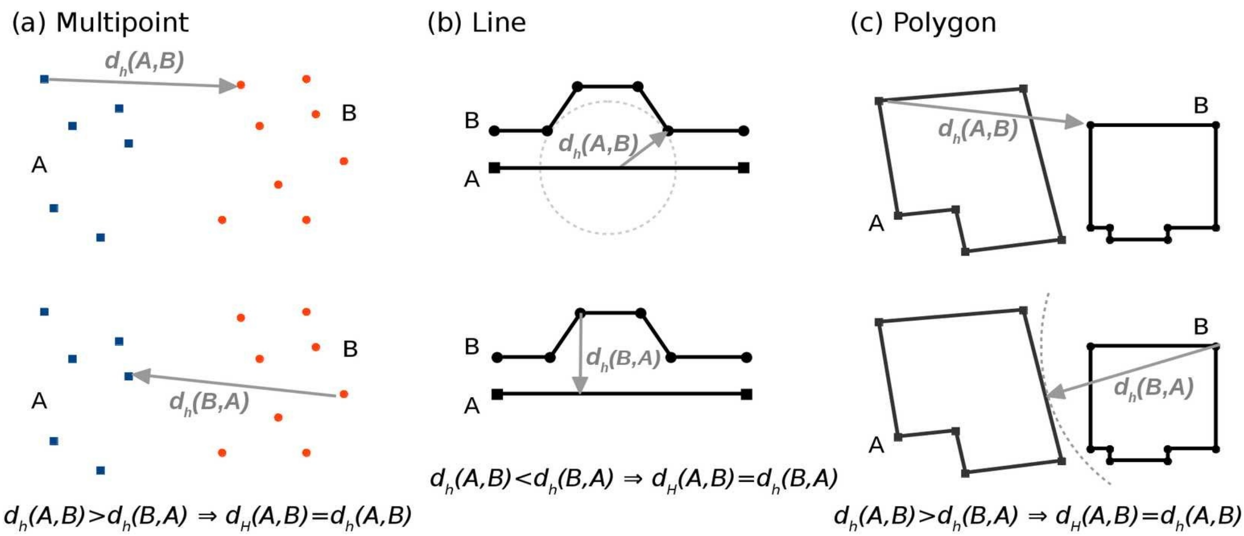


Figure 2.4. Hausdorff distance between spatial objects with different representations.

The research community has proposed some extensions to the original HD. Inside the Java Conflation Suite project, Davis (2003) proposed the Vertex HD as a simplification of the HD that considers only vertices to calculate the distance. Aiming to compare networks with different scales, Mustière and Devogele (2008) indicated the Hausdorff semi-distance, another simplification of HD but now considering only the directed HD from the object in the most detailed network to the other ones.

Another interesting extension to HD is due to Deng et al. (2007). The authors proposed the median HD to measure the central tendency of the distance distribution aimed at minimizing the influence of outlying portions of spatial objects. The median HD (Md_{dH}) is defined in a similar way to HD, but for each directed component the Md_{dH} selects the median value instead of the maximum value, as in HD. Based on Deng et al. (2007), Tong et al. (2014) developed the short-line median Hausdorff distance (SMHD), in order to handle length anomalies in road network data. The proposed SMHD is defined for line objects and takes into account the length of objects. The authors point out that SMHD is more suitable to measuring the distance between line objects, but it has not been tested on other kinds of data (areal features, point sets, etc.).

Despite the popularity of HD in GIS applications, some authors argue that this measure is not suitable for real-world natural entities, like curves representing rivers or borderlines (Alt and Godau 1995). According to Alt and Godau (1992), the reason for this problem is that HD only considers sets of points in both evaluated curves, but does not reflect the course of the curves. The authors believe HD is an appropriate measure for many applications, however they argue that some curves may have a small HD, but nonetheless greatly differ one from another.

2.2.1.3. Fréchet distance

According to Devogele (2002), the Fréchet distance measures the greater distance between two oriented lines. Each oriented line can be mapped to a continuous function f and g, so the Fréchet distance (dF) between these functions is defined as

$$d_F(f,g)=\inf_{\alpha,\beta}\left\{\max_{t\in[0,1]}\|f(\alpha(t))-g(\beta(t))\|\right\} \qquad (2.2)$$

where α and β are arbitrary continuous monotonic (non-decreasing) reparametrization functions α, β [0, 1] → [0, 1] with α, β(0) = 0 and α, β(1) = 1; and ||•|| is the Euclidean norm. These reparametrization functions (α and β) lead the distance calculus to occur in the homologous points (not necessarily vertices) of the assessed curves.

The Fréchet distance is sometimes known in folklore as the 'dog-man' distance (Efrat et al. 2007). This measure has been used in GIScience for merging lines (Devogele 2002), coastline matching (Mascret et al. 2006), matching road data with GPS traces (Brakatsoulas et al. 2005, Chen et al. 2011), and analysis of moving objects (Buchin et al. 2010, Ranacher and Tzavella 2014).

Some authors see the Fréchet distance as a better option for measuring sinuous lines (Alt and Godau 1995), and that it gives a more natural distance between curves (Driemel et al. 2010). However, this measure is more difficult to compute when compared to the Hausdorff distance (Alt et al. 2004). In fact, the exact Fréchet distance calculus proposed by Alt and Godau (1992) has a complexity of $O(mn \log^2 mn)$, where m and n are the number of vertices in compared lines. As an example, Wenk et al. (2006) reported that the calculus of the Fréchet distance can take several hours in matching one trajectory in a large dataset.

In order to overcome this performance issue, some researchers have worked on other solutions. An important algorithm for approximately computing the Fréchet distance is due to Eiter and Mannila (1994). The authors developed the discrete Fréchet distance (d_{dF}) as an approximation that computes in time $O(mn)$ by considering only the vertices into the calculus. Based on this work, Devogele (2002) proposed the partial discrete Fréchet distance (d_{pdF}) that is able to match one line to a part of another line. Another approach derived from Eiter and Mannila (1994) is the average Fréchet distance (d_{aF}) proposed by Mascret et al. (2006). This method uses the average Euclidean distance between vertices pairs and has been proposed as a complementary measure to the d_{dF}. Figure 2.5 illustrates the differences among the Fréchet distance (d_F), its discrete version (d_{dF}), and the Hausdorff distance (d_H).

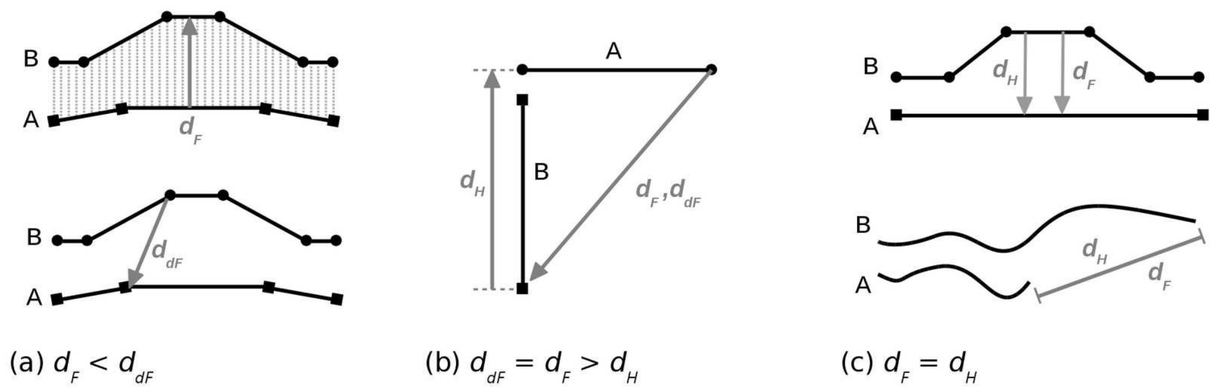


Figure 2.5. Fréchet distance in different situations. (a) Fréchet and discrete Fréchet. (b) Fréchet, discrete Fréchet and Hausdorff. (c) Same value for Fréchet and Hausdorff.

A more recent algorithm for calculating the nearly exact Fréchet distance was proposed by Driemel et al. (2010). The authors associated the approximate Fréchet distance to efficient computability of the distance for polygonal curves. Using a simplification of this work, Chen et al. (2011) presented a new measure for map matching that runs in near linear time for matching GPS routes and network data. The search for a near linear time method of calculating the Fréchet distance continues to produce new studies (e.g. Agarwal et al. 2014).

Despite the fact that the Fréchet distance has been used to measure the distance between two oriented lines, Devogele (2002) also propose its use to assess polygons, a point of view also shared by Buchin et al. (2006). In the case of areal shapes the easier to compute HD may be used in some cases, since Alt et al. (2004) have proved that the Hausdorff and Fréchet distances have the same values for convex closed curves.

2.2.1.4. Area overlap

Overlapping area is a technique used since early matching studies to find corresponding areal objects (Van Wijngaarden et al. 1997). The ratio of overlapping area between two polygon features can be calculated using the value of the intersection area over the average value of their individual areas (Hill 1990). Sometimes this ratio is calculated from one direction, using the value of one area instead of the average area in the denominator part (Fu and Wu 2008). This asymmetric approach can be called directional ratio of overlapping area. Another related method can consider only the minimum bounding box (MBB) of features in order to calculate the ratio of overlapping area (Ruiz-Lendínez et al. 2013), and in this case it is possible extend its use to line elements.

2.2.1.5. Geometric properties

Other geometric properties of spatial features have been used to assess similarities between linear and areal features. In linear objects we can cite length and orientation. The orientation is the angle in tangent of a linear feature, calculated from the starting point and ending point of a line (Olteanu-Raimond and Mustière 2008). Orientation is sometimes referred to as direction, and can be calculated for each segment (Yang et al. 2013). For polygonal objects, area, perimeter, and inertial axis are commonly used (Ruiz-Lendínez et al. 2013). Orientation can also be used in areal features, by using angles of the diagonal lines of the MBB (Tong et al. 2009); using the wall statistical weighting measure from Duchêne et al. (2003); or its adapted version of Zhang et al. (2012). Despite these geometric properties having been used in matching methods, Kieler (2007) and Ruiz-Lendínez (2012) demonstrated that these descriptors have little importance in the matching procedure of areal features.

2.2.1.6. Shape

Shape measures describe and quantify the geometric form of objects, notably for areal features. These measures have been investigated for many years in the research fields of pattern recognition (Loncaric 1998), image retrieval (Smeulders et al. 2000), product design (Cardone et al. 2003), and computational geometry (Van Kaick et al. 2011). In GISciences, shape measures are often used with other measures in order to support the matching process (Samal et al. 2004, Tang et al. 2008).

A common way to measure a shape of a spatial object is by using the compactness indicator, which describe the form based on how far it deviates from a given norm (e.g., circle, square) (Wentz 1997). MacEachren (1985) compared eleven compactness measures applied in a geographic context and concluded that the Relative Distance Variance (Bachi 1973) is the more suitable approach. The author concludes that indices based on perimeter-area measurements and on individual parameters of circles normally present insufficient accuracy. So the search for an efficient measure of compactness continues (Li et al. 2013).

The spider function is another shape measure presented by Rosen and Saalfeld (1985), and later used by Saalfeld (1988), to assess nodes by measuring the directions of the arcs emanating from a node. This function can divide the angle area in 2^n cases (e.g., 8, 16, or 32 sectors), and codify these sectors using some pattern, for example using an eight bits code. A similar approach is the azimuth code method proposed by Fu and Wu (2008). The authors used 24 angle areas and considered a code for each segment greater than 1/10 of the polyline's length, forming a string to be compared to other ones. In a more recent study Dongcai (2013) introduced an optimal spider coding method by combining 16 sectors and angles using a 32 bits code.

An important shape measure that has influenced later work is the distance between turning functions (θ) proposed by Arkin et al. (1991) for polygonal shapes. The turning function $\theta_A(s)$ is a signature of the shape and represents the angle of the counter-clockwise tangent as a function of the arc length s measured from some reference point in the boundary of polygon A. The turning function has been used in GIS studies to measure the shape similarity of on-demand maps (Frank and Ester 2006); data matching at feature level and also at vertices level (Ruiz-Lendínez 2012); and to evaluate the shape accuracy of building data (Fan et al. 2014).

There are other simple approaches to assessing shape similarity of spatial data. In order to evaluate the similarity of areal features, Samal et al. (2004) used the simple buffer method (SBM) proposed by Goodchild and Hunter (1997) to describe the positional accuracy of a linear feature. Fan et al. (2014) proposed the use of rectangularity, which is the ratio of a polygon area over its MBB area, as a shape descriptor. Zhang et al. (2014) defined elongation as the ratio of the width over the length of a rotated MBB for quantifying object shapes. Another option is to use a closed set of patterns for shape classification by comparing line patterns against this set (Chen and Walter 2009, Yang et al. 2014b). Lastly, a more sophisticated shape similarity measure based on the Fourier transform was proposed by Ai et al. (2013) for building data.

Shape measures are useful as a complementary technique, used to improve the precision of the whole matching method. Despite this low accuracy, compactness methods based on perimeter-area are easy to compute, so they are still used as a shape measure (e.g. Zhang et al. 2012). The spider functions are easy to work, although they may not be effective when angles are near but in different sectors, which may lead to a mismatch. The use of turning functions has presented interesting results. On the other hand, the pattern classifications for linear data seem to us to be a weak approach when compared with the widely used topological measures.

2.2.2. Context measures

Context measures allow us to quantify the geographic context of features in order to assess their similarity. According to Samal et al. (2004), 'geographic context refers to the spatial relationships between objects in an area', notably the relationships between an object and a limited set of landmarks. Tversky (1977) argues that the perceived similarity of two objects is decreased by the presence of a third instance. This observation leads us to consider that geographic context has the potential to play an important role in the matching scenario. Zhang et al. (2014) affirm that in ambiguous cases, the similarity of geographic features depends on the context. Figure 2.6 illustrates this situation well. In the first scenario (Figure 2.6(a)), it is fair matching a1 to none, and a2 plus a3 to b1. Yet if we consider a wider scenario like Figure 2.6(b), the uncertainty increases and maybe it would be better to match a1 to b1, and a2 plus a3 to b2. But in a still broader scenario (Figure 2.6(c)), regarding another type of feature, the uncertainty decreases and matching pairs proposed in the previous scenario becomes more clear.

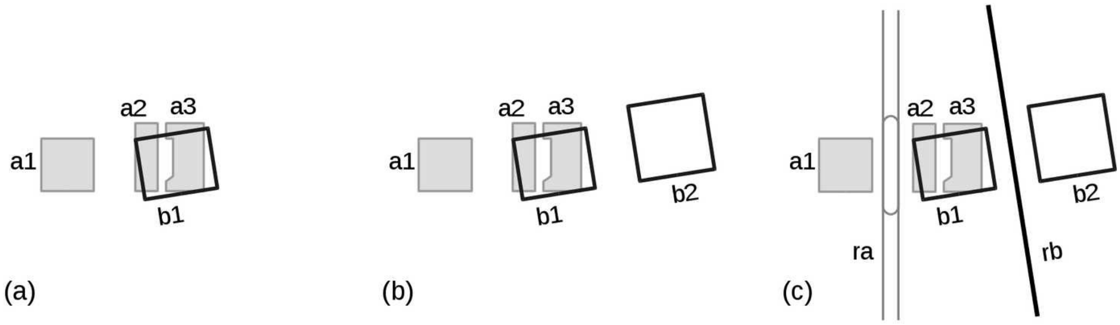


Figure 2.6. Context influences similarity decisions (Xavier et al. 2016a).

One of the first proposals was the compatibility coefficient of pairs introduced by Rosenfeld and Kak (1982). This coefficient is defined as a function to quantify the relative difference between one pair (A_i, B_j) against another pair (A_h, B_k). Any kind of measure can be employed to determine the joint compatibility, so Song et al. (2011) used the Euclidean distance in order to match point pairs. Inspired by this latter study, Yang et al. (2013) calculated the compatibility coefficient between linear data pairs combining distance and direction.

Considering the context as a broader concept beyond matching pairs, Samal et al. (2004) applied landmarks to similarity assessment. By combining many metrics, such as positional and attribute, the authors also proposed the use of landmarks to build a proximity graph in order to compare similarity between features. The proximity graph is a weighted directed graph defined to assess geographic context similarity based on 'proximity' relative to some pre-selected landmarks. The similarity is measured using the total vector offset of the corresponding objects in both datasets (see Figure 2.7(c)).

Kim et al. (2010) extended the previous context approach by using Voronoi diagrams and triangulation geometry. Figure 2.7 presents the similarities and differences between these two landmark-based methods. The aim is to compare a test data (B) against a reference data (A), then five landmarks are selected (L1-L5 and L1'-L5'). Next, the landmarks are connected to assessed features a ∈ A and b ∈ B in reference and test data (Figure 2.7(b-b')). So in the final analysis the two approaches differ. While Samal et al. (2004) use the total offset vector to quantify the context similarity (Figure 2.7(c)), Kim et al. (2010) use the area/perimeter ratio in the assembled triangles (Figure 2.7(d)).

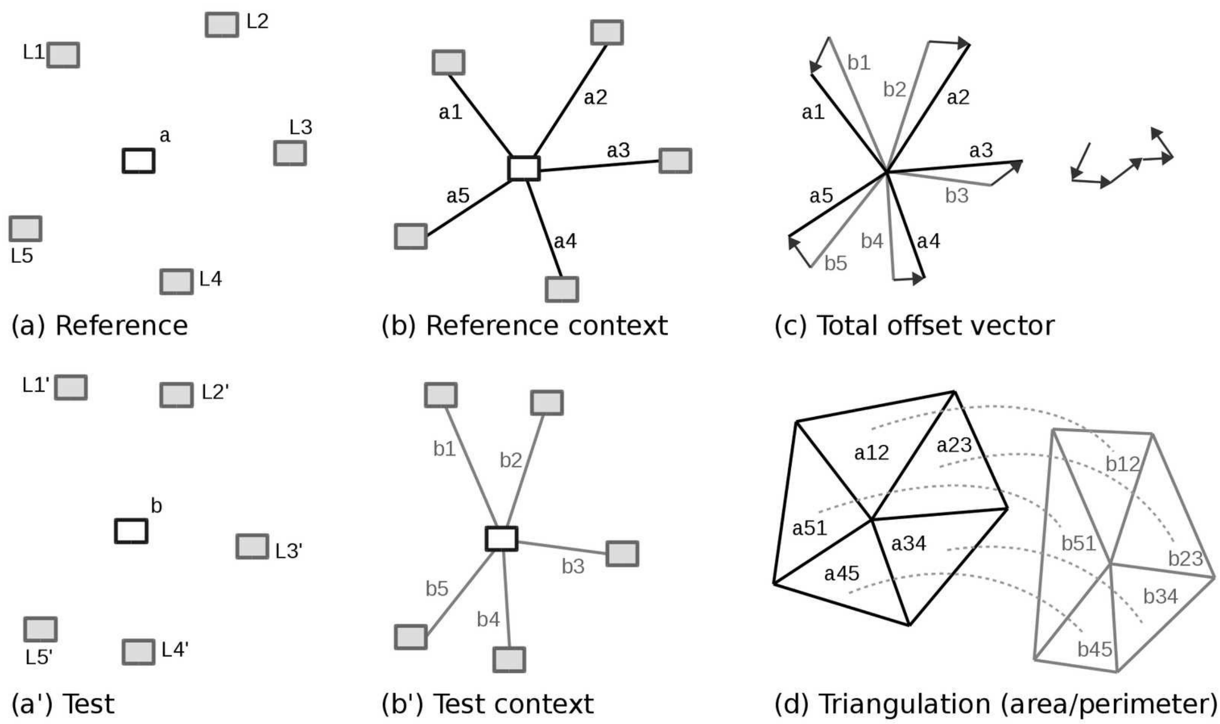


Figure 2.7. Different measures of context. Based on offset vector (c) (Samal et al. 2004) and triangulation (d) (Kim et al. 2010). Source: Xavier et al. (2016a).

Zhang et al. (2014), inspired by the k-nearest approach of Zheng and Doermann (2006), proposed the use of Delaunay triangulation to define the neighbourhood of objects, considering a continuous influence from the closest objects.

Context measures seem to be a promising solution to employ in geospatial data matching procedures. This kind of technique is particularly interesting when there is little information about the assessed datasets, or when there exists a large displacement between them. As shown in Figure 2.6(c), the context may help to reduce uncertainty when searching for corresponding features.

2.3. Matching of geospatial data

The quality control service developed in this study requires finding the correspondences between two datasets in order to execute direct external evaluation methods. This process is called geospatial data matching (Xavier et al. 2016a). Various classifications of matching methods have been proposed in the last 30 years. Rosen and Saalfeld (1985) classified matching criteria in some taxonomies: as discrete or continuous, geometrical or topological, local or semi-local, independent or dependent. While working with multi-scale databases, Devogele et al. (1996) proposed three kinds of data matching: semantic, topologic, and geometric. Similar classification was adopted by Yuan and Tao (1999), but using attribute instead of semantic matching, although with the same meaning. Volz (2006) proposed a classification according the geometric primitive: point-based, line-based, area-based, and mixed approaches. Similar classification was proposed by Dongcai (2013): dot-, line-, and plane-entity matching. From another point of view, Tong et al. (2009) classified matching methods as statistical and non-statistical.

We believe these classifications are not adequate to describe matching methods because most matching methods combine two or more kinds of perspectives in order to achieve better results. Reviewed studies indicate that these approaches differ fundamentally in the levels where they can actuate, and the corresponding cases supported by them. Based on this assumption we propose classifying matching methods according to: level of actuation and case of correspondence (Figure 2.8).

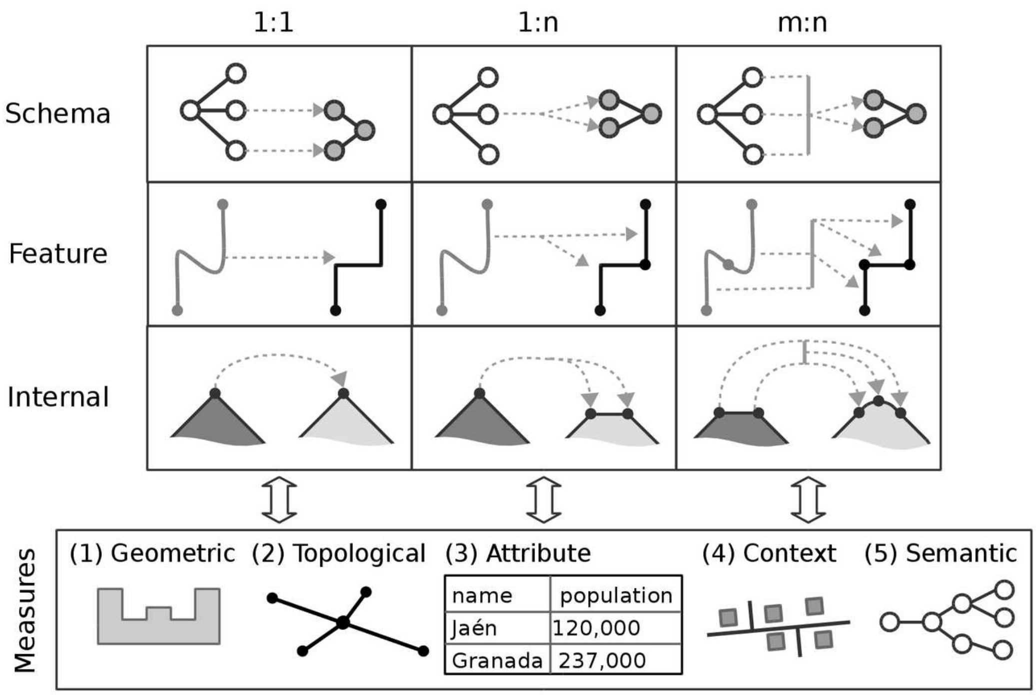


Figure 2.8. The taxonomy for matching methods taking advantage of all kinds of similarity measures (Xavier et al. 2016a).

Classification according to level refers to the 'height' of the level where the matching occurs. In our classification we propose the use of three values to classify matching approaches regarding the level of actuation: schema, feature, and internal. This classification is not exclusionary, i.e., a given method can actuate in both one level and another.

Classification according to the case of correspondence refers to the degree of correspondence supported by the method. Following the possible correspondence cases, from null matching (zero-to-one, one-to-zero, 0:1, 1:0) to multiple matching (many-to-many, m:n) (Van Wijngaarden et al. 1997), we propose three values with respect to the supported case of correspondence: one-to-one, one-to-many, and many-to-many. In our proposal this classification is accumulative, such that if an approach supports a many-to-many case, it also permits one-to-many and so on.

The next two subsections present related studies using the proposed classification. The last subsection presents how to assess the performance of matching methods.

2.3.1. Matching techniques according to the level of actuation

Matching techniques can be classified according to the level where the correspondence occurs: schema, feature, and internal. Schema matching occurs at the highest level: the modelling. Feature matching occurs at the object level, using instances. Finally, internal matching occurs inside a feature by using its interior components, like vertices in an area boundary. Figure 2.9 illustrates the proposed assortment.

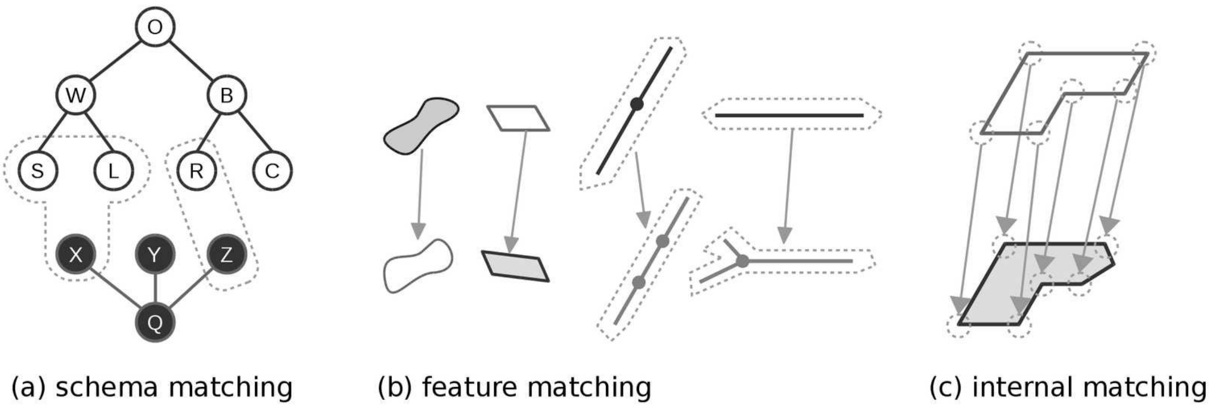


Figure 2.9. Classification according to level (Xavier et al. 2016a).

However, a more complete approach can actuate on all three levels described. It is possible that a schema matching approach uses feature-based methods, if this is a bottom-up approach (from features to classes, e.g. Volz (2005)). Along the same lines, internal methods are often combined with some feature matching approach, in order to determine corresponding features first, before beginning to analyse them internally (e.g. Huh et al. 2011).

In this study we deal with data matching at feature level and at internal level. So, in the following sections we present the related work on matching methods that actuate at these levels.

2.3.1.1. Feature matching

Feature matching methods begin with the assumption that the schema matching is already solved. Olteanu et al. (2006) call this kind of approach isolated matching. These techniques usually compare distances (geometric and attribute) between evaluated objects in order to find the correspondences. Feature-based approaches were present in early conflation studies; Saalfeld (1988) states that the measurement of similarities between features is the key component of any conflation algorithm. However, the most of these studies are focuses on linear data, e.g. road data.

Walter (1997) presented the buffer growing method, a feature-based method for network data that was used in various later studies (Walter and Fritsch 1999, Mantel and Lipeck 2004, Volz 2006, Zhang and Meng 2007, Chen and Walter 2009, 2010, Ying et al. 2011). Walter and Fritsch (1999) proposed a relational matching approach based on statistical investigations of the datasets. These investigations included geometric measures, buffer growing, and connections between features. The matching problem was mapped to an information communication problem solved using an A* algorithm. Despite the fact that the statistical method does not require weight factors, thresholds, or any kind of parameter, this technique demands training of datasets (manually matched).

Following the buffer growing principle, Mantel and Lipeck (2004) describe a three-step approach that encompasses division into classes, symmetrical buffer growing, and selection. With the aim of promoting street data integration in an SDI environment, Volz (2006) developed an iterative node and edge matching technique using buffer growing to handle one-to-many cases. Zhang and Meng (2007) proposed an automatic matching approach for road networks by extending the buffer growing in an unsymmetrical way. Chen and Walter (2009) extended the buffer growing method to allow matching not only between lines, but also between lines and points.

Jones et al. (1999) developed a feature matching approach based on Bayesian maximum likelihood methods (Johnson and Wichern 1998). This technique uses 'evidences' to condition probabilistic functions, including Hausdorff distance and line length. The drawback of the method is that it requires a large training site.

Other important study in feature matching is that of Beeri et al. (2004). The paper presents three location-based algorithms for finding homologous objects: the mutually-nearest method, the probabilistic method, and the normalized-weights method. Their work has influenced various later studies (Beeri et al. 2005, Safra et al. 2006, 2010, 2013, Tong et al. 2009, Ying et al. 2011). Initially these methods were developed to determine one-to-one correspondences in two datasets using one distance metric for points.

Beeri et al. (2005) adapted their original work to three or more sources, using two combining approaches: sequential and holistic. The road network matching approach of Safra et al. (2006, 2013) was able to match polylines by considering only endpoints as isolated points and applying the mutually-nearest method. In Tong et al. (2009), multiple measures are integrated in an extended version of the probabilistic method, and one-to-many cases are also enabled. Ying et al. (2011) used the buffer growing principle

(Walter 1997) in order to handle 1:n and m:n cases. This study applied multiple measures and multiple representations (point, line, area) using the probabilistic method.

Approaches concerned to linear data are common in feature matching. Pendyala (2002) proposed an algorithm that actuates on node, segment, and edge matching, by combining both top-down and bottom-up computations. Xiong and Sperling (2004) introduced a semi-automated network matching method that starts with a seed node screening solved by a cluster-based algorithm. Using the measure for cluster changing presented in a previous work (Luan et al. 2011), Luan (2012) developed a structure-based approach that transforms the road network matching problem into the Maximum Common Subgraph problem. Koukoletsos et al. (2012) described another feature-based automatic method of matching linear data, but their objective was to assess the completeness of VGI data. This multi-stage approach splits the datasets into smaller areas and found corresponding road entities using geometric (e.g. distance between segments) and attribute (e.g. road name) constraints.

In the research for matching geospatial data, the methods that actuate in the feature level are the most common. Among these techniques, most consider the features individually, without taking into account the relatedness among them. We believe that considering an isolated feature in a dataset could lead to a mismatch. One example comes from robot navigation research, where Neira and Tardós (2001) concluded that the matching pairs shall be reconsidered using a joint compatibility test in order to reduce the amount of erroneous pairs.

2.3.1.2. Internal matching

Internal matching techniques refer to methods focused on matching basic component parts of a geometry, like identify vertices in a polygon boundary or in a line string. This kind of approach can be found in studies focused on contour matching (Huh et al. 2011) or quality assessment (Fan et al. 2014, Ruiz-Lendínez et al. 2015). These studies are described here despite having a part of their methods focused on matching at the feature level because their internal procedures differentiate them from other feature correspondence studies.

Huh et al. (2011) present a method for detecting homologous point pairs in two polygon datasets using a modified version of the Vertices-Attributed-String-Matching (VASM) algorithm from Kaygin and Bulut (2002). This original algorithm matches the contours of compared polygons using their vertices as primitives in the calculus of the minimum-cost edit sequence required to convert a test vertex string into the reference vertex string. The Huh’s approach begins with a polygon matching using area overlapping, then the Hausdorff distance to filter (rejection criterion) some pairs. The modified VASM requires a training site in order to calculate appropriate thresholds for matching the conjugate-point pairs.

Another interesting approach that focuses on the internal parts of an areal feature is the study of Fan et al. (2014). The authors presented an approach to evaluating the quality of 2D building data available in OpenStreetMap for the city of Munich (Germany). Their approach to matching areal building features is a two-way area overlap, a technique able to identify many-to-many cases. When a one-to-one relation is identified they apply the procedure in order to find internal corresponding points, based on the assumption that there is not much divergency in shapes between VGI and reference data. This procedure first uses a Douglas-Peucker algorithm (Douglas and Peucker 1973) to find key-points of the building footprint. Then an MBB (not north-oriented) is calculated and aligned with the edges of the simplified footprint. So the test MBB is shifted to the reference MBB and the MBB edges are compared. Using this procedure, the authors have been able to investigate positional accuracy by computing the average distance between the homologous points.

Following the same line of quality assessment, Ruiz-Lendínez et al. (2015) presented an approach towards to increase expressively the number of features used in the evaluation of positional accuracy. The authors developed a technique based on genetic algorithms (Herrera et al. 1998), for feature matching, and boundary descriptors (Arkin et al. 1991) for internal matching. The homologous points in corresponding features are found using a modified version of Arkin's method that uses external angles and a perimeter rescaled to one. Thus it is possible to match homologous vertices by examining at the turning functions.

An overview of previous work indicates that internal matching methods are fundamental for quality assessment tasks of areal features. Finding internal vertices-pairs is the first step in a point-based positional quality analysis, like the National Standard Spatial Data Accuracy (NSSDA) (FGDC 1998). However, if the positional quality is evaluated through line-based methods, like the simple buffer method (Goodchild and Hunter 1997), the matching at feature level is sufficient, as demonstrated by Ruiz-Lendínez et al. (2013).

2.3.2. Matching techniques according to the case of correspondence

The supported case of correspondence for each method guides this classification into one-to-one (1:1), one-to-many (1:n), and many-to-many (m:n). One-to-one matching is the simplest approach present from early studies of conflation (Lynch and Saalfeld 1985) to more recent research (Safra et al. 2013). This conventional set also comprises the null matching cases, although some approaches do not support it. At the next level, one-to-many (and also many-to-one) matching methods are able to deal with aggregate objects. This case is particularly useful in a multi-scale environment whose data has different LoD. In the most general case, many-to-many matching is able to manipulate all possible cases. Figure 2.10 illustrates the proposed assortment.

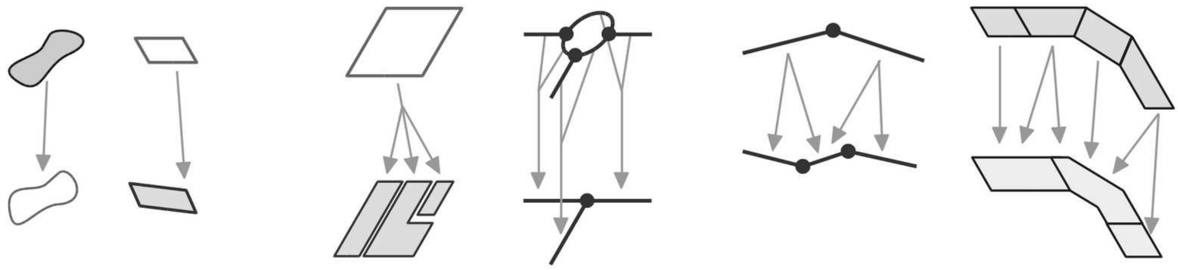


Figure 2.10. Classification according to case (Xavier et al. 2016a).

Below we present and review some related papers using the proposed classification.

2.3.2.1. One-to-one matching

The one-to-one matching case (1:1) is the simplest approach to matching geospatial data. Beginning with the work of Saalfeld (1988), this approach was followed by later studies (Beeri et al. 2004, Safra et al. 2010, Song et al. 2011). According to Li and Goodchild (2010), this strategy is unable to handle certain real application cases, such as when one object has several parts in another dataset (1:n), or various objects corresponding to a different number of objects (m:n). Notwithstanding this limitation, the 1:1 case is the prevailing situation in near-scale matching experiments (Huh et al. 2011, Fan et al. 2014), but some authors argue that these ideal cases are rare (Song et al. 2006).

A new approach to dealing with one-to-one matching cases, called the Opt method, was presented by Li and Goodchild (2010). The authors proposed a matching algorithm developed by converting the matching problem into an assignment problem (in operations research) formulated as an objective function that can be solved by an optimization model. At the core of this objective function lies the constraint that permits only 1:1 cases. Despite this limitation, the approach was later extended to accommodate 1:n cases (Li and Goodchild 2011) and even m:n cases (Tong et al. 2014). The main gain of the Opt method is that it focuses on the search for corresponding objects rather than choosing some measures and using a greedy strategy.

Based on the iterative relaxation labelling algorithm initially proposed for point matching by Ranade and Rosenfeld (1980), Song et al. (2011) integrated a distance metric with the compatibility coefficient, which designates the correlation of one candidate matching pair to another. This point-based approach matches road intersections but it has limited scope since it ignores multiple cases.

Another interesting one-to-one matching approach was proposed by Zhang et al. (2012), aiming at the maintenance of a multiple representation database. The authors transferred the data matching problem into a pattern classification problem, which they felt should be able to solve the problem of normalization and weighting multiple measures. They tested the approach with four supervised classifiers: C4.5 algorithms (Quinlan 1993), the classification and regression tree (Breiman et al. 1984), the Naive Bayes classifier (Rish 2001) and Support Vector Machines (Hsu and Lin 2002). The probabilistic Naive Bayes model was also used in a later study (Zhang et al. 2014). Although the four classifiers outperformed the weighted average of applied measures in terms of precision, these classifiers require large training samples, of almost the same size as the test data.

Although some authors feel that the one-to-one matching case is insufficient to handle datasets which differ in topology (Walter and Fritsch 1999, Zhang 2009), there are studies on matching road networks limited to 1:1 cases (Diez et al. 2008). This limitation may decrease the recall rates (see Van Rijsbergen 1979) of a method (e.g. Olteanu 2007), since multiple cases are treated as null matching, or even mismatching. In truth, the research indicates that there are authentic one-to-many and many-to-many cases using different geometric primitives or distinguished classes, like buildings and roads (Fu and Wu 2008).

2.3.2.2. One-to-many matching

One-to-many (and also many-to-one) matching methods (1:n, n:1) can manipulate composite objects, a common case while using a data in different LoDs. One-to-many cases are also common while manipulating networking data, like roads or rivers. Some studies using linear data sometimes split polylines and create virtual vertices (Volz 2006, Stankute and Asche 2009, Yang et al. 2013) in order to handle the multiplicity of corresponding objects. Sometimes existing one-to-many enabled approaches are derivatives from previous studies limited to one-to-one cases.

In her first study using the theory of evidence, Olteanu (2007) was only able to handle 1:1 matching cases. The support for multiple (1:n) cases was added in the later version (Olteanu-Raimond and Mustière 2008). In this study the authors developed a networking matching approach method on the Belief Theory (Dempster 1967, Shafer 1976). Using this technique Olteanu-Raimond and Mustière (2008) combined various matching criteria including positional, geometric, semantic, attribute, and topological (neighbourhood) criteria, in order to match French road networks at different scales. However, this approach requires a detailed analysis of data and expert knowledge in order to model the masses of belief (weights among criteria).

Another two papers describe one-to-many approaches using a greedy method: Tang et al. (2008) and Kieler et al. (2009). In the first study the authors proposed a method for areal feature matching that selects various characteristics of areal features: position (Euclidean distance between MBBs), shape and size. The final similarity is calculated by a weighted mean that can be obtained by training or by human expertise. The study by Kieler et al. (2009) aimed to match river datasets at different scales. The authors developed a three-step procedure to find corresponding river features, including using different representation (areas and lines). Under this method, the areal features are collapsed to center-lines using a skeletonization algorithm that preserves the topology. Then the subsequent matching steps occur in the line network, by using distances between nodes and angles between arcs.

In other one-to-many method Li and Goodchild (2011) play on the asymmetry of directed Hausdorff distance to settle m:1 and 1:m correspondences extending their Opt method (Li and Goodchild 2010). In this later paper (Li and Goodchild 2011), the authors focus more on measure than method, also using other measures like the Hamming distance to feature names and angles between edges.

Matching methods that support one-to-many cases can attain results impossible for the one-to-one approach. Although this class of techniques has increased the matching capabilities, they are not adequate to handle all possible cases in real world applications. Only the most generic approaches (m:n) can reach some cases (see Figure 2.11).

2.3.2.3. Many-to-many matching

Many-to-many matching (m:n) is the broadest approach that encompasses all possible corresponding cases. The support for m:n cases has been seen as fundamental to reaching authentic complex cases present from early work (Brown et al. 1995) until more recent research (Fan et al. 2014). Figure 2.11 illustrates how the same phenomena in real world ((a) source imagery) can be acquired in different manners from two distinct specifications ((b) and (c)), and in diverse representations ((1)-(3)). And many times these discrepancies lead to m:n matching cases.

One of the first studies to enable the management of many-to-many matching cases was Van Wijngaarden et al. (1997). Their goal was to propagate building updates between topographic maps at two different scales. The authors used an area overlapping percentage criterion and a geometry-enabled database management system to find corresponding buildings.

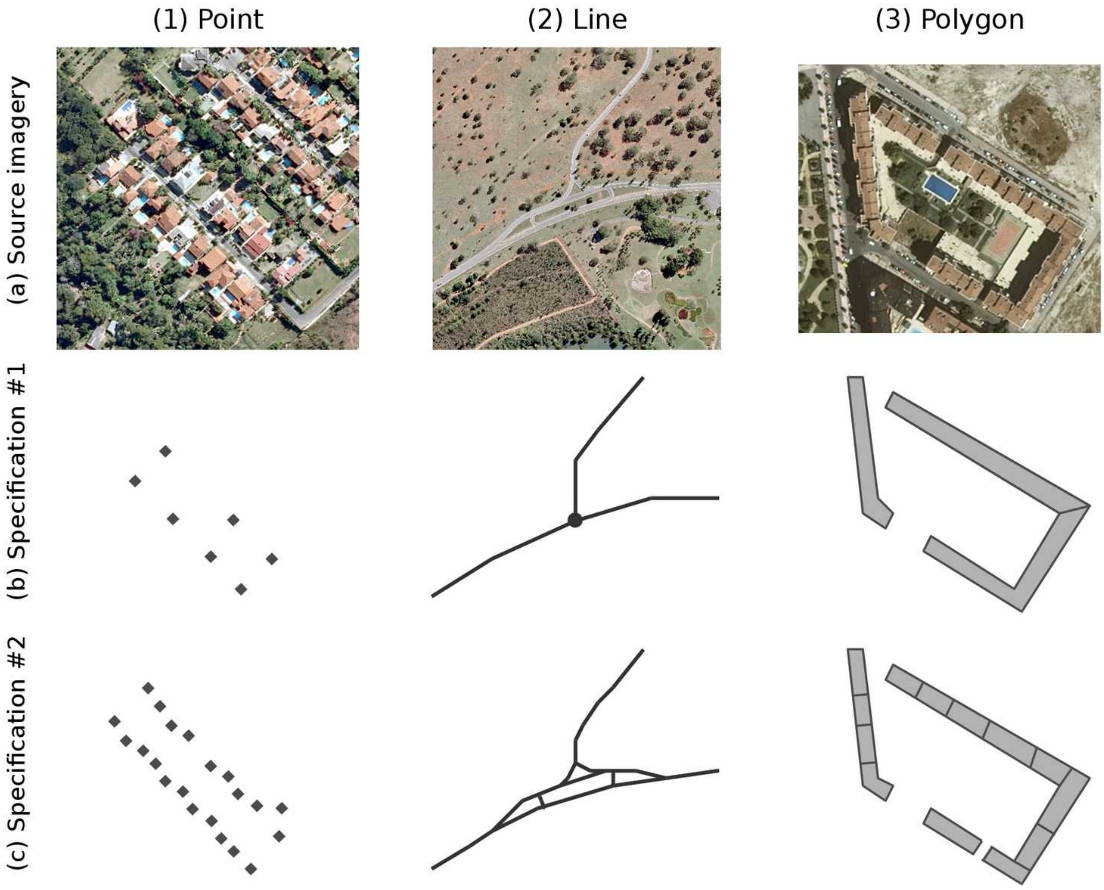


Figure 2.11. Same real world phenomena acquired using different specifications (Xavier et al. 2016a).

Another many-to-many approach for areal features was proposed by Butenuth et al. (2007). The authors expanded the work of Göesseln and Sester (2004) who aimed to develop a framework for the integration of geo-scientific datasets in a federated database environment. Their matching strategy first creates potential matching pairs using intersections. Then these candidates are assessed using symmetric difference (the union of both geometries minus their intersection) and azimuth histograms. In the case of many-to-many relations being identified these features are grouped together into a 'relation set' that can be processed as a simple one-to-one case.

A new matching approach that can handle many-to-many cases, termed the Delimited-Stroke-Oriented (DSO) algorithm, was developed by Zhang (2009). DSO utilizes three assisting methodologies: matching guided by *structure* (geometric and topologic properties), by *semantics* (attributes), and by *spatial index* (to increase matching speed). In order to deal with more generic cases, Zhang (2009) also proposes the concepts of partial correspondence (incomplete matching) and equivalent correspondence (topological inconsistency).

Based on Christmas et al. (1995), Yang et al. (2013) developed a probabilistic relaxation matching approach. The authors establish an initial probabilistic matrix using distance, length and direction. The compatibility coefficient is used during the relaxation process to iteratively recalculate the initial matrix until it reaches convergence values. In order to handle many-to-many cases, Yang et al. (2013) propose an additional five-step procedure based on structural similarity – the sum of matching probabilities of one candidate road matching pair and its neighbouring pairs. This method was also used to integrate VGI points of interest and street linear data (Yang et al. 2014). The framework described in the later work first executes a linear clustering algorithm based on DBSCAN (Ester et al. 1996) to determine the line patterns in VGI POIs, then it constructs a complete graph structure which represents these virtual lines. Finally, the matching problem between the 'lines' and a set of professional road data is solved by the probabilistic relaxation method.

Another two recent approaches to dealing with m:n cases are those developed by Zhang et al. (2014) and Tong et al. (2014). Zhang et al. (2014) extended the relaxation labelling techniques of Parent and Zucker (1989) which aimed to integrate areal building features using a context measure. Tong et al. (2014) developed a new technique by extending the Optimization method (Li and Goodchild 2010) with logistic regression models. The logistic regression model enables both one-to-many and many-to-many matching relationship cases, in this case proving useful to integrate linear road data.

Although m:n cases sometimes indicates an inconsistent matching (Van Wijngaarden et al. 1997), it is possible to find genuine cases of many-to-many matches (Huh et al. 2011, Ying et al. 2011). Some authors have argued that it is necessary to go beyond m:n cases, using the concepts of partial correspondence and equivalent correspondence (Zhang 2009). However, a deeper analysis indicates that these are particular to cases of many-to-many matching, so they can be framed within this classification.

2.3.3. Describing the performance of methods

Common tools for assessing the results of matching studies are the precision and recall measures from the information retrieval (IR) field. Van Rijsbergen (1979) presents precision and recall using the IR concepts of *retrieved* and *relevant*, for the results obtained and the rule of thumb, respectively. In a geospatial matching scenario we can call the test data 'retrieved', and the reference data 'relevant'. So a matching method tries to increasing the number of correct test data against reference data and avoids wrong matches (false positives) and non-matches (false negatives). Thus, precision and recall can be defined in terms of the equations:

$$Precision = \frac{correct}{correct + wrong} \qquad (2.3)$$

$$Recall = \frac{correct}{correct + unmatched} \qquad (2.4)$$

Some authors also consider ambiguous cases in the denominator of precision (Yang et al. 2013). Huh et al. (2011) argues that the F-measure is also an option, because this method is popular in the ontology matching field (see Euzenat et al. 2009). The F-measure is a parametrized combination of precision and recall, where an α parameter ([0, 1]) allow relative importance to be assigned to precision or recall (Do and Rahm 2002). In the cases where precision and recall are considered equally important we have α=0.5, and the expression becomes the harmonic mean between these values.

$$F\text{-}measure(\alpha)=\frac{Precision\cdot Recall}{(1-\alpha)\cdot Precision+\alpha\cdot Recall} \qquad (2.5)$$

Not all matching studies indicate the precision and recall values found in their experiments. Some studies report an *average precision*, not using the IR concepts, but do not detail whether errors are mismatches or non-matches (e.g. Walter and Fritsch 1999, Volz 2006, Li and Goodchild 2011). Other papers describe performance using precision and recall relative to the length of assessed features (e.g. Kieler et al. 2009, Koukoletsos et al. 2012). Most of studies that uses precision and recall is in the feature level. There is no study of this metric at the schema level, and just that of Huh et al. (2011) at the internal level.

2.4. Design of experiments

In the previous section (Section 2.3) we noted that the geospatial data matching is an active topic in GISciences, that has been investigated by many researchers. This interest had generated many approaches to deal with this issue. This diversity of solutions raises a question: which one would be suitable to our quality control service? We believe that the experimental design is the appropriate tool to find this answer.

A designed experiment is that in which is possible to make intentional changes in some controlled factors of a system or process, notice the resulting variations in the observed variables, and then analyse the influence of those factor in these variables (Montgomery and Runger 2003). For Park (2007) the Design of Experiments (DOE) has been used for executing effective experimentation and analysis of their results. Some authors prefer the term experimental design (Seltman 2015).

The main concepts around experimental design are: (1) controlled variable, (2) factor, (3) level, and (4) treatment. The *controlled variables*, sometimes characteristic functions, are the objective of the experiment in which we want to maximize or minimize (Park 2007). Alba (2013) argues that *factor* is the independent variable in which we want to determine its effect over the controlled variable, and the different values of this factor are named *levels*. The author presents *treatment* as the combination of different levels of considered factor in an essay, or in case of a single-factor experiment, the treatment is the level itself.

The key-tool for DOE is the analysis of variance (or ANOVA, for short) which permits to obtain the optimal results (Rodriguez-Avi 2011). ANOVA allows to identify if there is influence of an assessed factor over the controlled variables (Alba 2013). The null hypothesis (H_0) of ANOVA is the equality of effects in the means of a controlled variable ($\tau_i = \mu - \mu_i$). So the test involves verify the alternative hypothesis (H_1), where at least one treatment has effect over the variable (Montgomery and Runger 2003). The ANOVA assessment can be done by analysing the corresponding F test for ANOVA, or using the P-value that represents the smallest level of significance to reject the null hypothesis (in this case, the equality of means).

$$\begin{aligned} &H_0: \tau_1 = \tau_2 = \ldots = \tau_t = 0, \ \ t \text{ treatments} \\ &H_1: \tau_i \neq 0, \ \ \text{for at least one } i \end{aligned} \quad (2.6)$$

If after running ANOVA over a set of observations we accept the null hypothesis, this means that there is no effect of the considered treatments over the controlled variable. On the other hand, rejecting the null hypothesis means that at least one treatment has affected the controlled variable (Alba 2013). When this later case occurs (reject H_0) we can apply analysis over data in order to find which treatments are equivalent. These techniques are named *multiple comparison methods*, and are presented in the next section (Section 2.4.1).

As Montgomery and Runger (2003) pointed out, statistically designed experiments bring gains to the experimental process, particularly the objectivity when drawing out conclusions. In fact, these tools have been applied in research areas as diverse as water quality (Vega et al. 1998), service quality (Caruana 2000), magnetic resonance imaging (Zou 2004), and materials science (Ozcelik and Erzurumlu 2006).

Concerning GISciences, ANOVA has also been used in some research projects. The early study of Anselin and Getis (1992) indicated that ANOVA is an appliance that could be included in an analytical module of a GIS. Since then, ANOVA tools were successfully applied in different applications inside GISciences, for instance: geospatial sampling (Davidson and Csillag 2003), spatial interpolation (Luo et al. 2008), land-use change (Dendoncker et al. 2008), climate change (Persson et al. 2012), or to test a geographic regression model (Harris et al. 2013). More recently, ANOVA was used in a study towards the visualization of positional uncertainty (McKenzie et al. 2016), and it was used to evaluate the difficult in generating crowdsourced data (Salk et al. 2017).

2.4.1. Multiple comparison methods

Alba (2013) presents three multiple comparison methods: Tukey's HSD (Honestly Significant Difference) test, Fisher's LSD (Least Significant Difference), and Bonferroni's method. There is no consensus about which one should be used, however Dunn (1961) pointed out that Tukey HSD might be preferable when the number of means tend to be small.

Following Alba (2013), the Tukey HSD is preferred by researches because it maintains the same level of significance for all comparisons.

The pairwise comparison method of Tukey is a conservative approach (Abdi and Williams 2010) that has been used in studies as diverse as ecology (Chase and Knight 2003) or language (VanPatten 1990).

2.5. Geospatial web services

In computer science, a service is a self-describing, open component that supports rapid, low-cost composition of distributed applications (Papazoglou and Georgakopoulos 2003). Service provider is the organization that implements, and provides both technical and business support, and supply descriptions of a service. These descriptions may include a set of information about a service, commonly interface, location, behaviour, and even quality.

Web services provide a systematic and extensible way for application-to-application interaction, built on top of existing Web protocols and based on open standards (Curbera et al. 2002), most of them based on XML. The Web services framework is divided into three areas: communication protocols, service description and service discovery. Simple Object Access Protocol (SOAP) is the communication protocol; Web Services Description Language (WSDL) is used to service description; and Universal Description, Discovery, and Integration (UDDI) plays the role of a service directory (Newcomer 2002).

One important organization in the development of web services is the Open Geospatial Consortium (OGC). This consortium is formed by over 520 companies, universities and government agencies. Its objective is to promote the development of technologies that enable interoperability among systems that use spatial information (OGC 2017). This consortium publishes some specifications known as geospatial web services (GWS): Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS), Catalog Services – Web (CSW), and Web Processing Service (WPS) (Beaujardiere 2006, Vretanos 2010, Whiteside and Evans 2008, Nebert et al. 2007, Mueller and Pross 2015). OGC standards also encompasses file formats, like Geography Markup Language (GML) and Keyhole Markup Language (KML) (Portele 2007, Burggraf 2015).

The proposed quality control service is implemented over a WPS interface, so in the following section we review the WPS specification and related studies.

2.5.1. Web processing services

Web Processing Service (WPS) is an OGC standard that defines an interface aiming to publish and to use geospatial processes, as well as the discovery of and the binding to those processes by clients (Schut 2007). In the early March 2015, OGC consortium released the version 2.0 of this specification (Mueller and Pross 2015). The WPS interface is defined by means of six operations: three mandatory and included in WPS

1.0 – GetCapabilties, DescribeProcess and Execute; two optional operations to handle with asynchronous processes – GetStatus and GetResult; and the Dismiss operation. This last operation is defined in the Dismiss extension, and it can be used in both situations: synchronous and asynchronous jobs. These operations are presented in the sequence diagram of Figure 2.12 using the Unified Modelling Language (UML) notation.

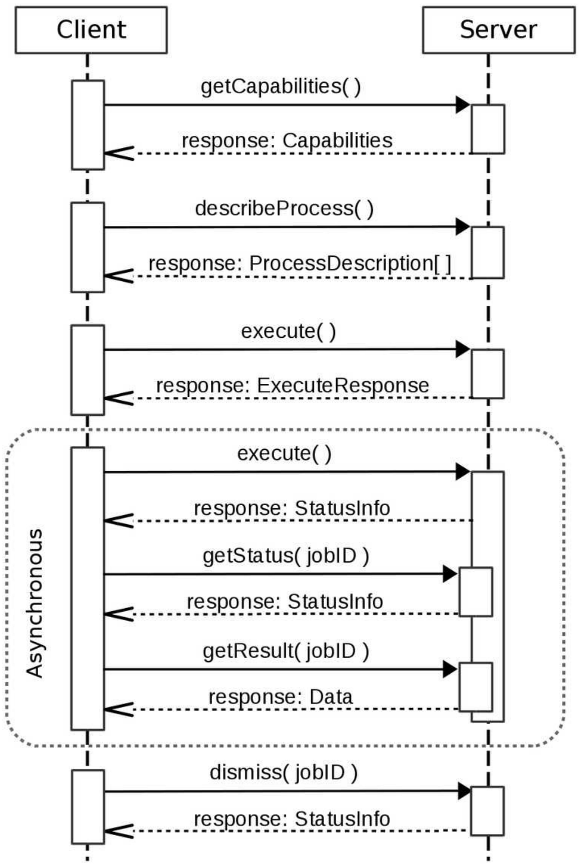


Figure 2.12. Sequence diagram for WPS operations (Xavier et al. 2015b).

According to Mueller and Pross (2015), GetCapabilities operation returns the service metadata and a list of process available at the server. DescribeProcess provides detailed information about a list of selected processes. Execute is the key-operation of a WPS – it permits a client to execute some process given a list of parameters. GetStatus is an operation that allows a client to query status information of some asynchronous processing job. GetResult operation allows a client to recover the final result of an asynchronous job. Lastly, Dismiss operation permits a client to cancel an asynchronous job, or for finished jobs this operation will release all allocated resources, like temporary files or result files.

The research community around the Geographic Information Systems (GIS) environment has investigated some aspects of the on-line processing of geospatial data for a variety of purposes.

Kiehle et al. (2007) developed an open-source WPS server, the degree WPS, in order to analyse the applicability of this specification. The authors concluded that including complex processing tasks, like a model for global climate change, can be encapsulated inside a WPS. Brauner et al. (2009) proposed three main topics for a research agenda of geoprocessing services: (1) service orchestration strategies, (2) semantic description of processes, and (3) improve the performance of these services. These authors also argued the WPS interface provides an efficient communication mechanism using its asynchronous messages capabilities.

Friis-Christensen et al. (2009) introduced the term Distributed Geographic Information Processing (DGIP) while developing an on-line application for forest fire risk analysis. Their architecture was based on OGC specifications, among them the WPS, and service orchestration provided by the Web Services Business Process Execution Language (BPEL) (OASIS 2007). Some authors have argued the BPEL has becoming the *de-facto* standard for service chaining (Akram et al. 2006). Biodiversity applications also have demanded on-line geoprocessing tools. In this sense Fook et al. (2009) developed a conceptual framework that enables the collaboration in biodiversity by allowing sharing species distribution modelling experiments. Granell et al. (2010) took advantage of the standard WPS interface to develop an open architecture that allows the calibration and the running of hydrological models.

Other interesting study is due to Zhao et al. (2012). The authors introduced the term geoprocessing web as a broader concept that covers all aspects toward distributed and collaborative geoprocessing over the web. Interoperability is one of the characteristics of the geoprocessing web, and the WPS specification plays this role. Hofer (2013) evaluated the commonalities and differences between geoprocessing web (Zhao et al., 2012) and geospatial cyberinfrasrtuctures (Yang et al., 2010). The author concluded that both concepts have the function of data analysis and knowledge generation, and also encompass the resource of distributed geoprocessing and web services.

The automation of quality control for spatial data has also shown recent works using the WPS interface. The study of Donaubauer et al. (2008) proposed a web service with the ability to generate quality information of assessed data via web services. The work used WPS to process the quality control, and ISO 19115 (ISO 2003b) for the quality report by means of metadata elements. Despite the simplicity of the quality procedure, just an overlay of previously tagged data with some quality elements, this study seemed to be the first attempt of an automatic evaluation service in the literature. Other study also indicated that the quality evaluation can be executed through a WPS (Mobasheri 2013). In a more recent paper, Meek et al. (2016) presented a new approach to WPS orchestration and chaining towards the quality control of crowdsourced geospatial data.

CHAPTER 3
FRAMEWORK FOR GEOSPATIAL DATA QUALITY EVALUATION

In this chapter is presented the framework for geospatial data quality evaluation developed in this research project. This framework is composed by the architecture of a solution towards quality assessment through web services.

We propose a three-tier architecture for a web services platform focused on quality control of geospatial data (see Figure 3.1), which we are calling the *quality control service*. From a bottom-up point of view, the first tier, *Data Access*, is used by external evaluation methods to manage reference data: retrieve and matching. The second tier, named *Evaluation*, implements the different quality evaluation procedures available at the service. The last tier, named *WPS*, handle client requests using the standardized interface of OGC WPS. This architecture was first discussed at Ariza et al. (2015).

Data Access tier manages the relation between test and reference data. Since direct external evaluation procedures depend on reference data for comparison, this tier provides the correspondences between datasets in order to permit compare them. There are two ways to facilitate reference data: (1) remote: who calls the service provides the reference data; or (2) local: the service itself has its own reference dataset. The Data Access tier manages the access to local reference data, and also provides a *Matching* module that provides data matching between assessed and reference data (local or remote). According to what is been requested by the external method, this matching can be in the feature level, or in the internal level, i.e., by considering vertices of a geometry. Feature matching is detailed in Section 3.1, while internal matching is detailed in Section 3.2. The information about which reference dataset was used in an evaluation shall be included in the corresponding quality report.

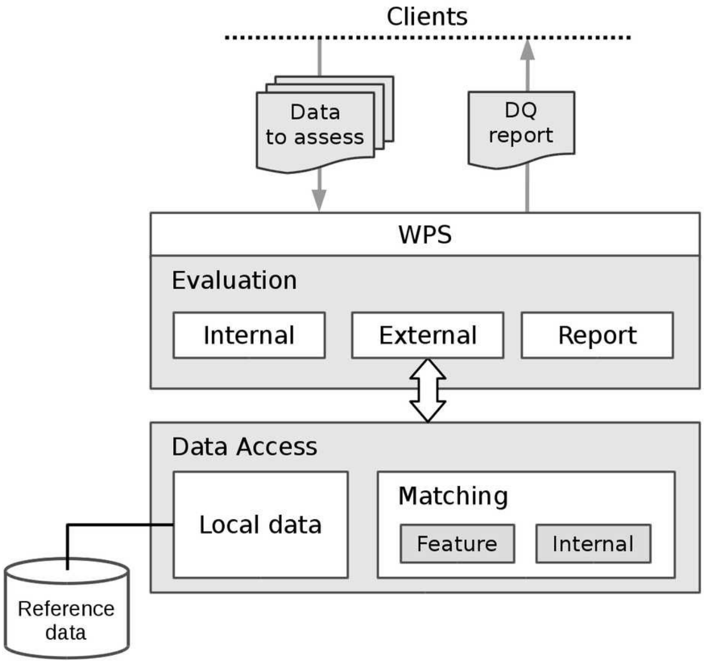


Figure 3.1. Proposed architecture for a quality control service.

Evaluation tier contains the implementations of evaluation methods, notably direct internal and direct external, since internal methods are not recommended in the ISO 19157 (ISO 2013). Direct external methods requires an external reference that is managed in the Data Access tier. Evaluation tier also contains the Report module that is responsible for generate the quality report in different ways: a human-readable report, or an XML report in ISO format, current (ISO 2016) or legacy (ISO 2007). Each implementation of an evaluation method shall be able to provide the metaquality report relative to its own assessment. Evaluation tier is detailed in the Section 3.3.

In the proposed architecture, the WPS tier is the point of contact with the clients. This tier handles requests and responses using the WPS interface. This implementation should support asynchronous requests by using the 'status' parameter of WPS specification in order to permit possible long-term processing jobs. This tier is detailed in the Section 3.4.

The proposed architecture is intended to be general for automatic quality assessment, and should be applied independently of datasets or software platform.

3.1. Feature matching module

In geospatial data quality assessment, the direct external evaluation methods require a reference dataset in order to compare the test dataset. In the proposed architecture of a quality control service the Matching module plays the role of finding the correspondences between these two datasets (reference and test). These correspondences can be at the schema level (among concepts), the feature level (among objects), or the internal level (among parts of objects, e.g. vertices). In this approach we assume that the schema matching is already solved, so we focus on feature matching (this section) and internal matching (Section 3.2).

There are a plethora of approaches facing matching at feature level, as discussed in Xavier et al. (2016a). So we decided to investigate which ones would be adequate to our service. In order to achieve this goal we opened three working fronts: (1) development of similarity measures; (2) preparing a matching testbed; and (3) over this testbed we applied some matching methods under the control of design of experiments.

Similarity measures are discussed in Section 2.2 and the new proposed measures are presented in Section 3.1.1. Matching testbed is a set of synthetic geospatial datasets and their correspondences (matching pairs) prepared with the aim to be a benchmark data when comparing matching methods. This testbed provides a rule of thumb for matching methods and is described in Section 3.1.2. Hence we apply the similarity measures using matching methods over the testbed under the control of a designed set of experiments (details in Section 3.1.3).

3.1.1. Similarity measures

This section presents the similarity measures that were developed in this study for the matching of geospatial data. The first measure uses the concepts of shape context developed by Belongie et al. (2002) to quantify the geographic context. The second measure quantify the line orientation using closer parts (set of segments). The next subsection details these measures.

3.1.1.1. Geographic context measure

There are few context approaches available in the literature in order to match point features, like the studies of Samal et al. (2004), Kim et al. (2010), and Zhang et al. (2014). However, the two first approaches are based on the selection of landmarks, what can be difficult to determine in an automatic system. The Zhang's approach is based on the Delaunay triangulation to define the neighbourhood, what can be limited for many closing objects.

We developed a context distance based on the shape context descriptor developed by Belongie et al. (2002). Despite of the original study has used the shape context for finding correspondences between shapes, we adapted those concepts in order to quantify the geographic context of point features by using its relative position against the closer objects.

The context distance is computed as follows. For each point in a dataset we compute a common histogram with the polar coordinates of all other close points relative to the assessed point. Figure 3.2(a) illustrates how the histogram bins are positioned in the space considering the assessed point in the center. It is possible to note that not all points are used, just those inside a neighbourhood limit, or r_{max}. So, for each bin, we count the number of point features and fill the respective value. After counting all neighbours inside the limit, we have the shape context for the assessed point, which is the context 'signature' for that feature.

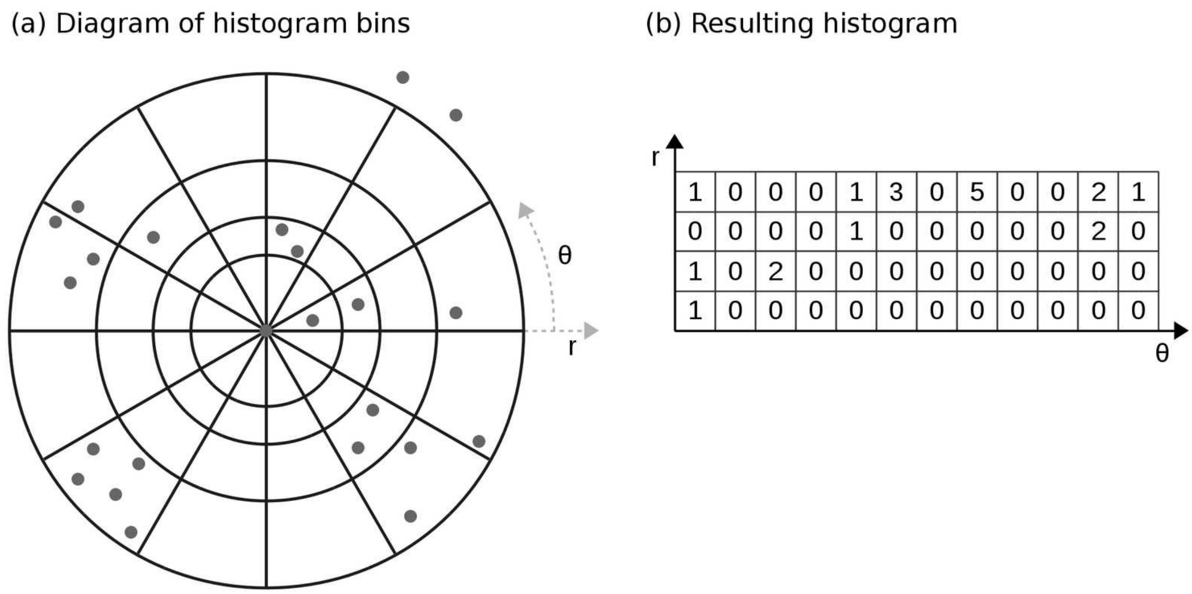


Figure 3.2. Geographic context measure. (a) Diagram of histogram bins. Note that not all points are considered. (b) Resulting histogram representing the number of points in each bin.

The position of the histogram bins differ from the original Belongie's work, where the authors considered a log-polar coordinate system (Belongie and Malik 2000, Belongie et al. 2002). In our study we propose that the increment of radius r should occur by the length of the previous arc. So it works as a geometric progression that begins with an r_0 and has a common ratio equals to the angular step in radians.

We adopted the cost function proposed by Belongie et al. (2002) as the context distance. This distance measures the similarity between histograms, i.e., between shape contexts, using a χ^2 test statistic for normalized histograms. This distance assume values from zero (completely similar) to one (dissimilar).

$$C(p_i, q_j) = \frac{1}{2} \sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \qquad (3.1)$$

Where C represents the cost function that measures the context similarity between points p_i and q_j, and $h_i(k)$ and $h_j(k)$ represent the K-th normalized histogram bin for points p_i and q_j, respectively.

3.1.1.2. Partial orientation

Orientation is a geometric property that has been used in matching methods for linear data (Yang et al. 2013, Kang et al. 2015). For Olteanu-Raimond and Mustière (2008) the orientation is calculated in a linear feature from the starting point to the ending point.

In this study we propose a slight difference in this measure. We are using the orientation between a short line against the corresponding closer part in a greater line. Figure 3.3 illustrates an example. It is possible to note that the orientation of line a1 (o(a1)) is very different of the orientations of lines b1 and b2 (o(b1) and o(b2)). However, when we consider the orientation of line a1 in relation with line b1 (o(a1, b1)), it is possible to note that it is more similar to o(b1) than considering o(a1) singly.

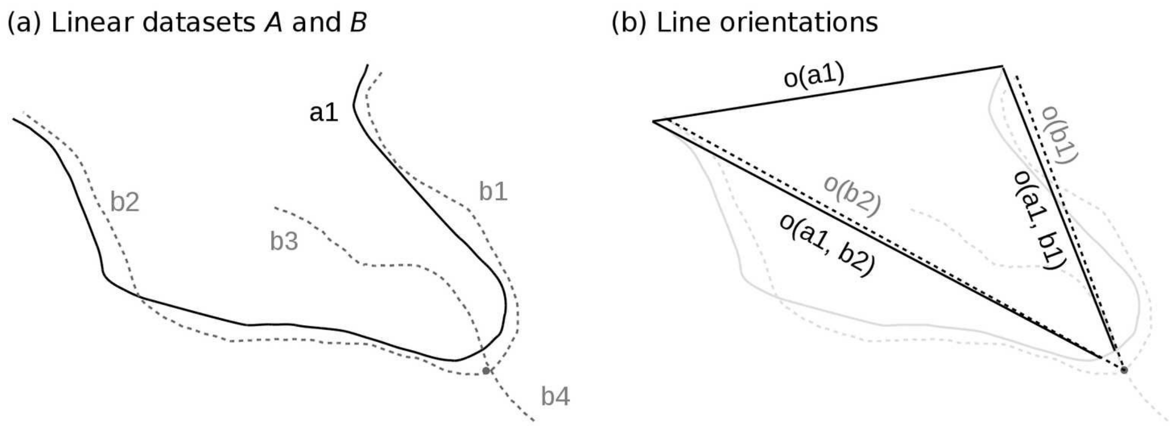


Figure 3.3. Partial orientation method. (a) Linear datasets. (b) Line orientations. o(a1) means the orientation angle of line a1. o(a1, b2) means the partial orientation of a1 in relation to the short line b2.

The partial orientation measure can be used to increase the precision of a feature matching method by permitting assess the relative orientation between candidate line pairs.

3.1.2. Preparing a matching testbed

The quality control service is designed to be general, i.e., it should work with any kind of vector geometries, in any data conditions: presence or absence of features, different morphologies of data, in presence of systematic or random errors, etc. So we propose this testbed in order to simulate the possible problems that a matching method might encounter when actuating in a quality control service. This section is part of a paper that was submitted for consideration at a journal (Xavier et al. 2017). The datasets created in this testbed are used in the experimental design for feature matching, described in Section 3.1.3.

The geospatial data matching testbed is composed of four groups of datasets: (1) initial, (2) morphology modified, (3) systematic disturbance, and (4) random disturbance. Initial datasets are originated from mapping provided by official mapping agencies of Spain at scales 1:25,000 and 1:10,000. The other three groups are derivative of the initial datasets at scale 1:25,000. Morphology modified datasets are composed of synthetic objects in some morphology class for linear and areal features. Systematic disturbance are composed of datasets that were generated from affine transformations over initial data. The last group of datasets (random disturbance) is formed by data influenced by displacement vector fields applied over the initial datasets. Each group of datasets is compounded by the datasets for the three geometric primitives: point, line, and area, except for the morphology modified group which does not have point data. Each test dataset (scale 1:25,000) can be divided into nine regions. Each region can be identified by the first number of the object identifier (OID), e.g., 1023 is in the first region, and 9128 is in the ninth region. Figure 3.4 illustrates the dataset groups in the matching testbed.

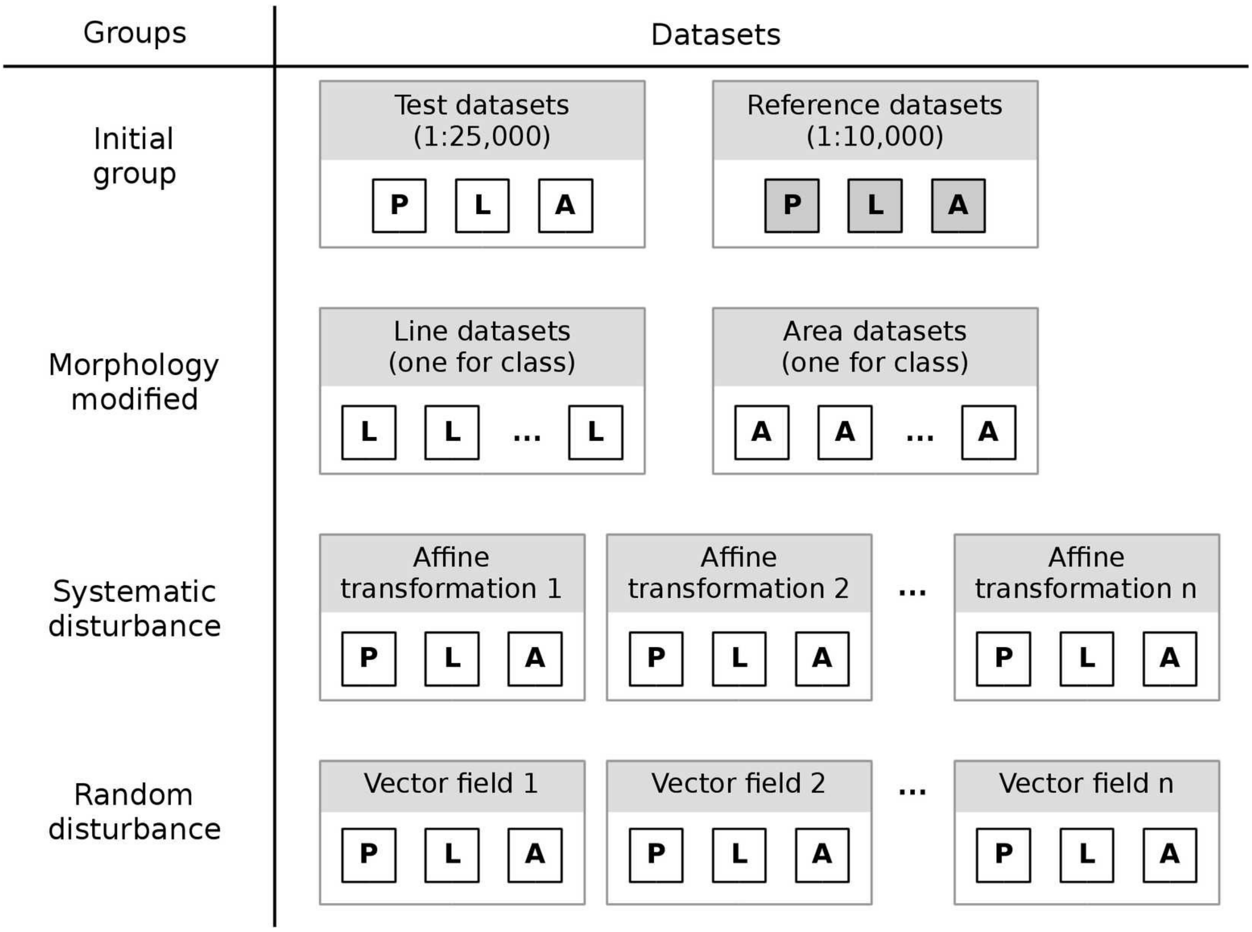


Figure 3.4. Dataset groups available in the matching testbed. Each dataset is compound by one kind of geometry: point (P), line (L) or area (A). Source: Xavier et al. (2017).

The following subsections detail the initial datasets and the methods used to prepare each group of datasets.

3.1.2.1. Initial datasets

The initial datasets in this testbed are formed by test datasets at scale 1:25,000 and reference datasets at scale 1:10,000. Test data is originated from the Base Topográfica Nacional 1:25,000 (BTN25) of national mapping provided by the Instituto Geográfico Nacional of Spain (IGN 2015). Reference data is originated from the Base Cartográfica de Andalucía 1:10,000 (BCA10) of regional mapping provided by the Instituto de Estadística y Cartografia de Andalucía (ICEA 2015). Test datasets were divided into nine regions (S1-S9) with their corresponding regions in reference datasets (B1-B9).

The test data was selected from BTN25 as follows: Point data were created from the *Building* class, so the areal features were converted to points using their centroid. First we selected buildings with an area less than 1000 m² and compactness index relative to a square (MacEachren 1985) less than 1.2. Then for the regions S1-S6 we randomly selected more than 100 objects, since the performance of matching methods is measured in percentage (Section 2.3.3). For the regions S7-S9 we selected objects from some agglomerated areas in order to be close to an urban environment (more than 100 objects in each area).

Area data were also obtained from the Building class, but excluding those objects selected as point data. After excluding point features, we randomly selected more than 100 objects for the regions S1-S6. For the regions S7-S9 we selected some near objects in order to represent an urban environment. In these cases, more than 250 objects were picked in each dataset. For the linear data the first step was to homogenize the road data for the initial datasets (BTN25 and BCA10) using the same topological rules. These procedures avoided 'broken' lines and 'long' lines. So line data were selected from BTN25 in this order: (1) motorways, (2) roads, (3) links, and (4) tracks. In order to reach at least 125 objects in each region, some tracks were selected randomly.

After these selection procedures over the test data we performed a manual matching between the selected objects from the BTN25 test data against the BCA10 reference data. We used regional orthoimagery to help us in this task in order to dismiss any doubts.

The last step was translating the areas to a generic place of the world, since they no longer represent any reality. So we also mirrored, rotated or translated the data in order to decharacterize the original site.

3.1.2.2. Morphology modified

The morphology of linear and areal objects is a factor that may affect the performance of geospatial matching procedures. The quality control service may evaluate the quality of geospatial data with distinct levels of detail, so it is possible that the same feature might be acquired with more details in one dataset than in another. In order to deal with this factor we adopted a roughness classification for lines and developed a complexity classification for areas. Based on these morphology classifications we developed a method for generating synthetic data from some source data for a specific morphology class.

The line roughness classification is based on the road-line classification developed by Ariza-López and García-Balboa (2008) where the authors used a back-propagation artificial neural network (BANN) over a moving window. Since we use road data in our experiment, this method seems to be adequate for our purposes. The BANN method defined five roughness classes for road data: (1) very smooth, (2) smooth, (3) sinuous, with stable directionality, (4) sinuous, with variable directionality, and (5) very sinuous. Figure 3.5 presents examples of lines classified according to this method.

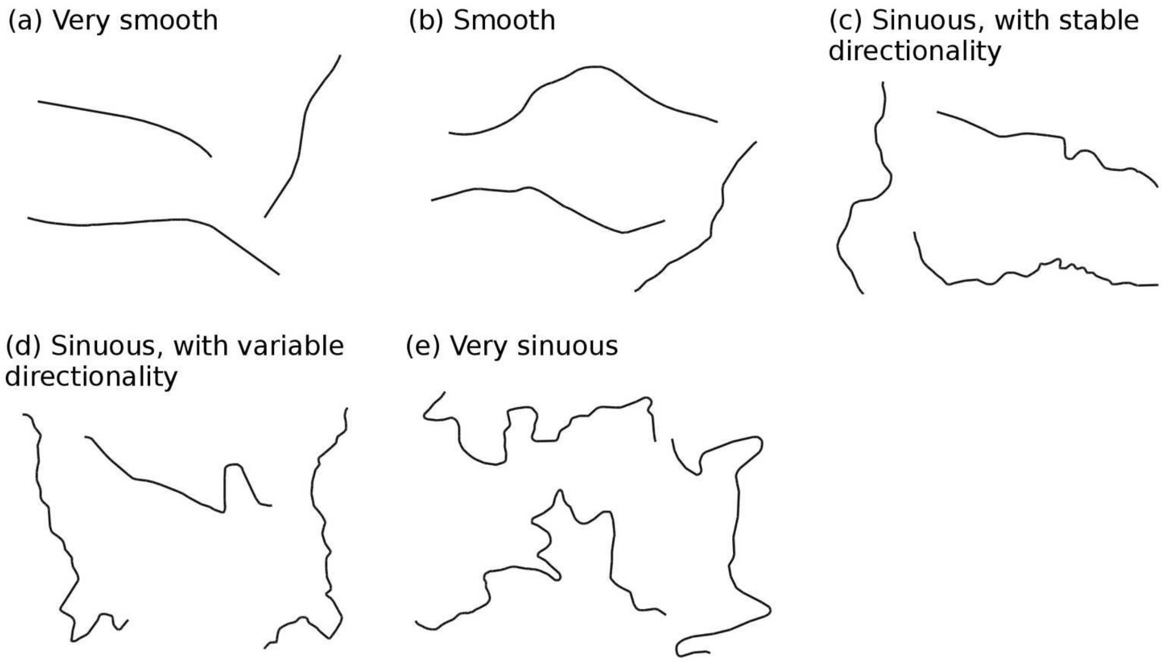


Figure 3.5. Samples of the morphology classification applied to lines (Xavier et al. 2017).

The area complexity classification developed in this study is drawn for building data at small scale. This method is based on two concepts: convexedness and Arkin's turning function (Arkin et al. 1991). We propose a complexity classification for area building data defined in four classes: (1) very simple, (2) simple, (3) complex, and (4) very complex. Class 1 areas group the simpler objects that are the convex ones without holes. Class 2 areas are those convex ones with holes, and also those objects that are similar to some standard, like 'L' or 'U' objects. More complex objects (class 3 and beyond) are determined according to their number of 'turns', i.e., the number of times that the external ring changes its current turning (left or right). Class 3 are those objects with less than or equal to 10 turning changes, while class 4 (very complex) buildings are those that exceed this limit. Figure 3.6 shows some areas classified according to this method.

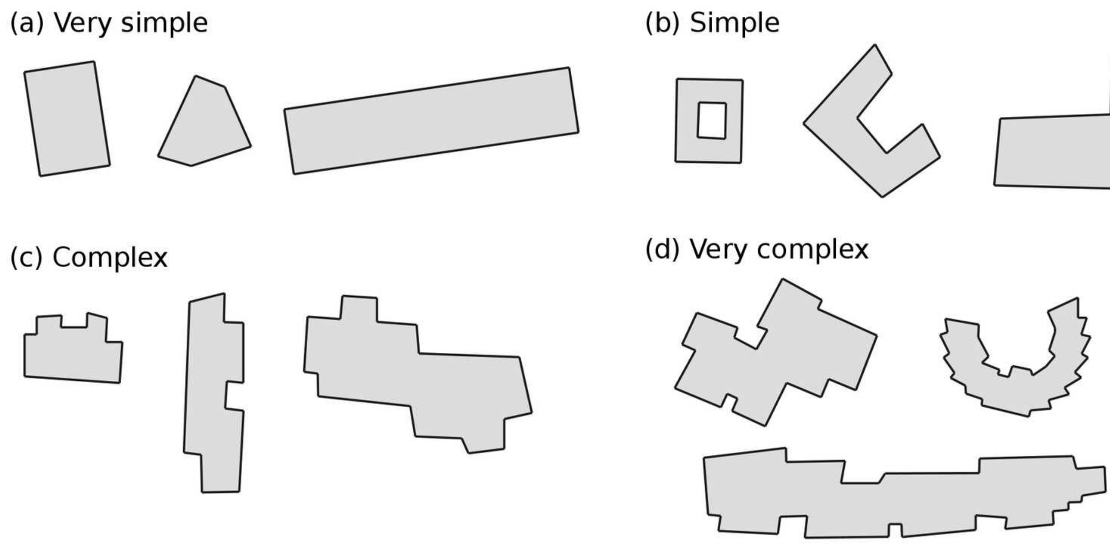


Figure 3.6. Samples of the morphology classification applied to areas (Xavier et al. 2017).

With the aim of incrementing the population of lines and areas in each morphology class we propose two methods for generating synthetic data, modifying original sources according to the desired morphology class.

The method for lines works as follows: For each line in the original data, we compare its morphology class with the desired morphology class. If the difference is greater than two, or the object already has the desired class, the object remains unaltered. Otherwise, a procedure for smoothing or roughing the line should be applied in order to reach the desired morphology class. The smoothing procedure is a combination of Douglas-Peucker simplification (Douglas and Peucker 1973) with Gaussian filtering (Babaud et al. 1986, Plazanet 1997) of sigma 4 (García-Balboa 2006). The roughing procedure applies random displacements along internal curves of the line (clockwise or counter-clockwise). These procedures do not affect the first or last points of manipulated lines. As these procedures do not take into account the neighbouring objects, some lines required manual edition in order to maintain the topology. Figure 3.7(a) shows an example of how a line, originally classified in class 3, can be flattened to class 1 or can be roughened to class 5.

Similar to the line method, the area method also generated synthetic data. The procedure does not affect class 4 areas (very complex). For each area in the original data, we compare its morphology class with the desired morphology class. If there is no difference, the object remains the same. Otherwise, we apply a procedure that randomly disturbs or simplifies the object's geometry in order to achieve the desired morphology class. The disturb procedure raises the complexity of polygons by means of perforating one without holes or creating a 'corner' at a random vertex. The simplify procedure acts over non-convex polygons by removing the corners that least influence the area size. Figure 3.7(b) presents an example of an area that in the source data was classified as simple (class 2 – 'L' shape) that was simplified to class 1 and was disturbed to class 3.

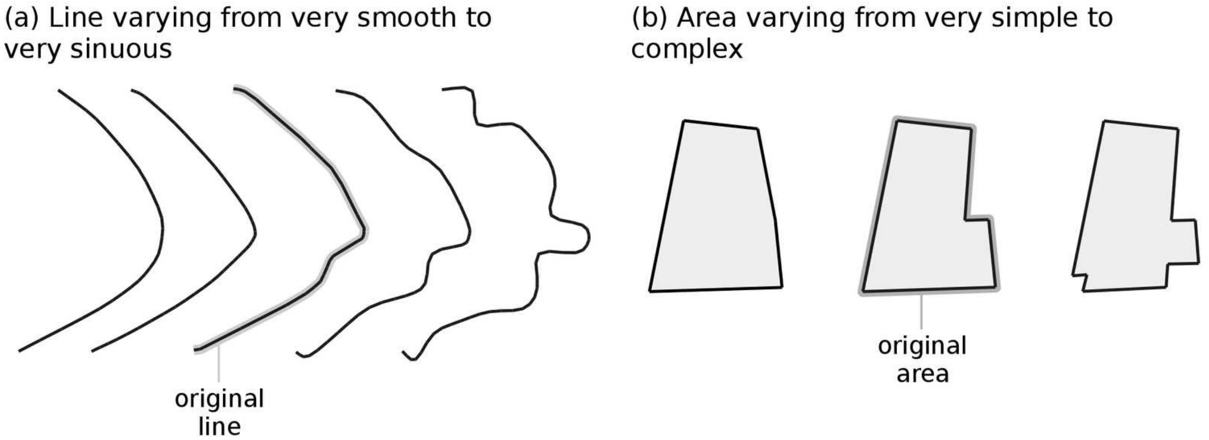


Figure 3.7. Examples of lines (a) and areas (b) created according to the morphology modified procedure (Xavier et al. 2017).

3.1.2.3. Systematic disturbance

The presence of positional systematic disturbance is a factor that potentially affects the performance of geospatial data matching procedures. The quality control service may assess datasets with some kind of systematic disturbance, with distinct amplitudes, so it is important to identify the influence of intentional systematic perturbations in position over matching procedures. Our methodology is similar to the study of Mozas-Calvache and Ariza-López (2014), where the authors simulated several displacements over original data, such as translations, rotations, and scaling.

We propose generating synthetic data from the original data by applying a set of systematic disturbances represented by means of an affine transformation. This transformation is a composition of translations, rotations, and scaling (Weisstein 2016). Hence our systematic disturbance method is designed to reflect these three kinds of transformations. The approach requires a set of *standard displacements* that define the entire process and it also requires a minimum bounding rectangle (MBR). For each displacement we generate a set of systematic disturbances, that are: (1) translations in eight directions, (2) counter-clockwise and clockwise rotations over three different pivots, and (3) two scaling factors (dilation and shear).

The translations are determined by the standard displacement applied in eight directions, beginning at 0 with increments of 45° (Figure 3.8(a)). The rotation angle is calculated for each dataset, taking into account the standard displacement and the half of MBR's diagonal (Figure 3.8(b)). Using this angle we have six possible rotations: two directions (counter-clockwise and clockwise) by three rotation pivots in relation to the MBR (lower-left, centre, and upper-right) (Figure 3.8(c)). Finally, the scaling factors for dilation and shear are calculated using the relation between half of the MBR's diagonal and the standard displacement, as we can see in Figure 3.8(d).

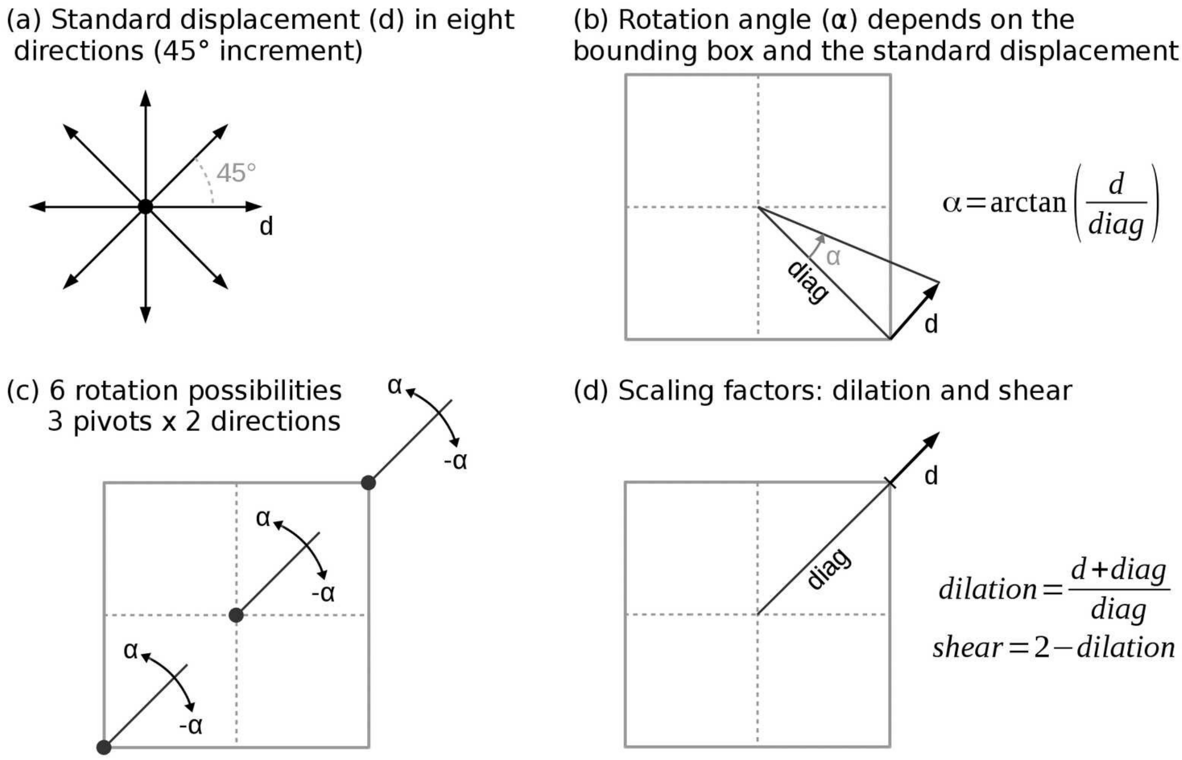


Figure 3.8. Systematic disturbances in function of a standard displacement. (a) Translations. (b) Calculating the rotation angle. (c) Rotations. (d) Scalings (dilation and shear). Source: Xavier et al. (2017).

After determining the translations, rotations, and scalings for each standard displacement, these perturbations must be combined to create a set of affine transformations that will be used to generate the synthetic perturbed data. The no-disturb configuration (no translation, no rotation, no scaling) is added prior to creating the affine transformations. Then all possible combinations are generated. For instance, if we have only two distinct standard displacements, the number of combinations is calculated as follows: combinations = translations × rotations × scalings = (1 + 8 × 2) × (1 + 6 × 2) × (1 + 2 × 2) = 17 × 13 × 5 = 1105. So we have more than 1000 different transformations using only two standard displacements. For each affine transformation, a new synthetic dataset is created.

3.1.2.4. Random disturbance

This last group of datasets acts similarly to the systematic disturbance, but this approach uses random perturbations over original data in order to simulate this random behaviour. The proposed quality control service may assess datasets with some kind of random variation, since the uncertainty, which includes the positional uncertainty, is natural within geospatial data (Fisher et al. 2006). Therefore we intend to check the influence of intentional random perturbations in position over matching procedures. Other studies (Mozas-Calvache and Ariza-López 2014) adopted random errors in each vertex, including correlated displacement by lines. In this study we propose a new methodology to disturb geospatial data using vector fields created for a given *standard displacement*, a concept shared with the systematic disturbance (Section 3.1.2.3).

The key-concept of our methodology for random disturbs is the displacement vector field (Nagel and Enkelmann 1986). This vector field works as a 'force' field that modifies the geospatial features by acting over their coordinates by means of random displacement.

There are three parameters in this approach: standard displacement, field resolution, and sigma. The standard displacement works as in the systematic disturbance (previous subsection), i.e., it defines the amplitude of perturbation. The vector field in this method is created according to a regular tesselation of the source data MBR, so we need a field resolution in order to define the cells. Finally, the sigma value represents the internal variation of the standard displacement. For instance, if we use a sigma of 10%, the random displacements will vary ±10% in amplitude in relation to the standard displacement.

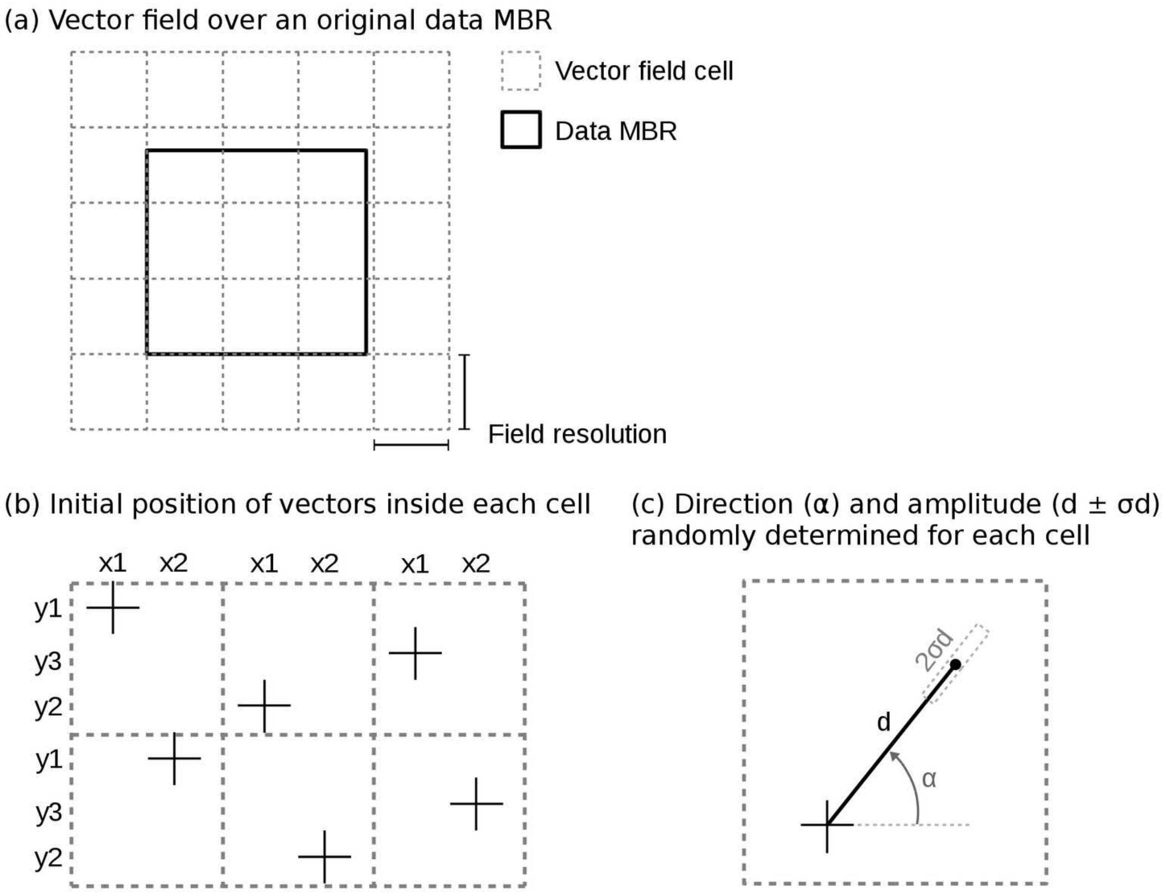


Figure 3.9. Vector field method. (a) Tesselation over original data. (b) Unaligned position of generators. (c) Direction and amplitude are randomly determined. Source: Xavier et al. (2017).

The displacement vector field method works as follows: After defining the parameters, it creates a regular tesselation using the field resolution over the dataset's MBR. This tesselation has at least two additional columns and rows with the aim being to guarantee that the data border will fit inside the vector field (Figure 3.9(a) illustrates an example). Then, using an unaligned systematic pattern (Morrison 1970) it randomly creates a set of x values (one for each row) and a set of y values (one for each column). These values define the coordinates for each generator of our vector field, one per cell (Figure 3.9(b)). The next step is to define the direction and amplitude for each field generator. The direction is randomly determined while the amplitude is calculated in function of the given standard displacement plus a random variation limited by the sigma (σ) parameter (Figure 3.9(c)). In the end we have a vector field with a displacement vector for each cell in the tesselation.

The displacement vector field is composed of a set of displacement vectors which quantify the disturbance to be applied in a dataset. We propose using this vector field as a geometric transformation over the original dataset. The influence of each vector in a coordinate of perturbed data should be determined in function of an interpolation function. In this approach we adopt the inverse distance weighting (IDW) interpolation with pow 2 and search radius of 2.5 times the field resolution, as indicated by Gumiaux et al. (2003). The synthetic perturbed data is generated for each random vector field created from the given parameters (standard displacement, field resolution, and sigma). Figure 3.10 shows an example of how works the geometric transformation created from a displacement vector field.

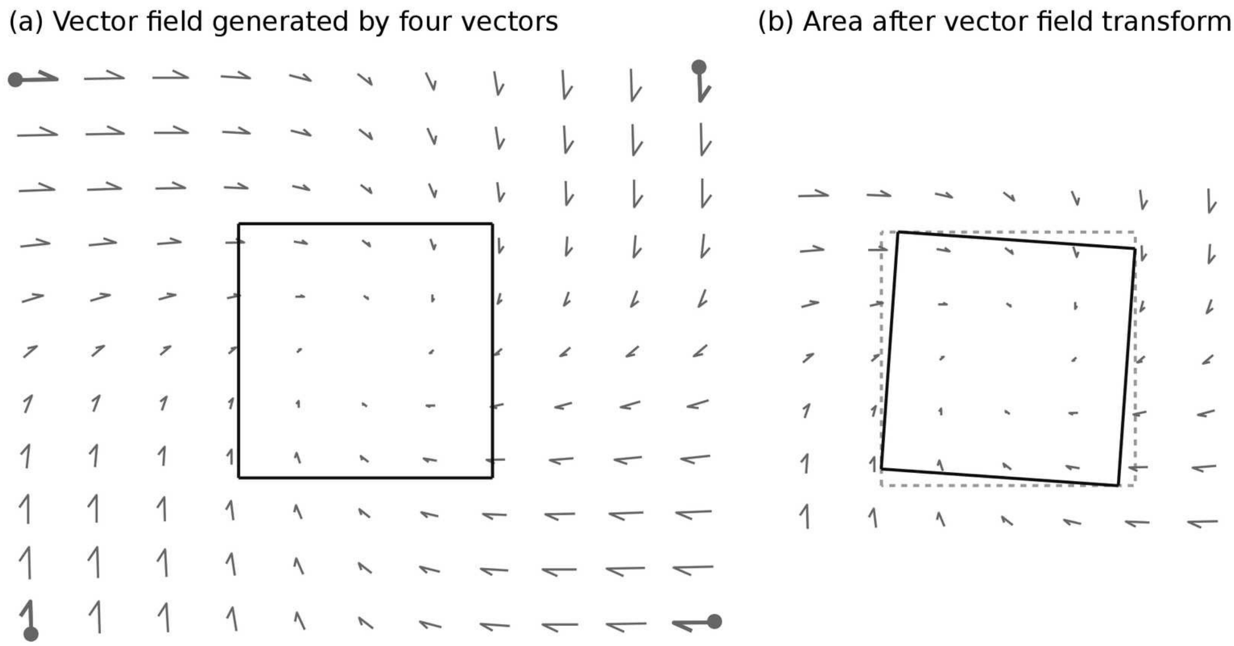


Figure 3.10. Vector field geometric transformation. (a) Vector field generated from four displacement vectors on the edges (gray dots). (b) Area after suffers the vector field transformation.

3.1.3. Experimental design for feature matching

The quality control service needs a matching method with high precision and recall in order to provide a trustworthy evaluation of assessed datasets. Since the service should run over a web environment, the computational cost in time is other relevant issue. A poor matching method certainly would influence the data quality results. Looking for the related literature hitherto, it is possible to identify many approaches addressing this issue, many of them with good results in their test sites. We believe that a designed experiment, as embracing as possible, over the matching testbed (Section 3.1.2) has the potential to indicate the more suitable approach(es) for the proposed quality control service.

Design of experiment techniques were developed to assess how some variables are influenced by a list of factors. In this study we chose four variables: precision, recall, F-measure, and time. We selected six factors that might influence these variables: similarity measures, matching methods, morphology of features, geographic context, systematic disturbance, and random disturbance.

Regarding our four variables, precision and recall are concepts that come from the information retrieval field (Van Rijsbergen 1979). Precision evaluates the presence of wrong matches (false positives) against the real matches (true positives); while recall evaluates the presence of non-matches (false negatives) against those real matches. The F-measure represents the harmonic mean between precision and recall. These three variables are detailed in Section 2.3.3. The last variable considered is the time, i.e., the time consumed to run some matching procedure. While the three first variable control the matching quality, the time is related to system performance, which might be a decisive aspect for some applications (e.g., on-line systems).

The study of Xavier et al. (2016) identified that the geospatial data matching problem can be organized in two key aspects: similarity measures and matching methods. So these are the two first factors to be investigate: measures and methods. The third controlled factor is the morphology of objects, i.e., how the roughness of lines or areas influences the variables. Other factor is the geographic context of features, which refers to the spatial relationships between objects in a neighbourhood (Samal et al. 2004). The last two factors refers to some disturbances applied to source data: systematic and random perturbations. In theses cases is possible to assess the robustness of the investigated matching techniques.

The next subsections detail each factor.

3.1.3.1. Factor: similarity measure

Tested measures are divided according to the geometric primitive: point, line, and area, despite of some commonalities.

For point features we adopted two measures: Euclidean distance and context distance. The ubiquitous Euclidean distance is used in many studies in order to match points, for instance Samal et al. (2004), Li and Goodchild (2010). The geographic context distance proposed in this study (Section 3.1.1.1) is used to assess the neighbourhood of point features.

We have adopted three classes of measures for line features: geometric properties, distances and buffer-based measures. The geometric property used here is the partial orientation described in Section 3.1.1.2. The distances are presented in Section 2.2.1:

- Hausdorff distance (Rucklidge 1996);
- Short-line median Hausdorff distance (Tong et al. 2014);
- Discrete Fréchet distance (Eiter and Mannila 1994); and
- Partial discrete Fréchet distance (Devogele 2002).

In this study we consider two buffer-based measures: the single buffer method (SBM) and the double buffer method (DBM). SBM was proposed by Goodchild and Hunter

(1997) for assess the positional accuracy of lines. This measure returns the proportion of a tested line that lies within a buffer generated from a reference line. DBM was developed by Tveite and Langaas (1999) as a part of a methodology related to the positional accuracy of linear objects. In this method two polygonal areas are generated as buffers in the compared lines. The measure represents the relation between the intersection area over the union area of these two buffers. Geometric properties and distance measures were selected because they had been used in other matching studies (Xavier et al. 2016a). Buffer-based measures were selected because they had been used in quality evaluation studies (e.g., Mozas and Ariza 2011, Ruiz-Lendínez et al. 2016), so the aim is assess its feasibility in matching methods.

In order to evaluate the similarity of areal features we have selected four sets of measures: overlap measures, combination of area descriptors by means of a genetic algorithm, context measure, and line distances.

The overlap-based measures are usually applied to match areal objects, as we can see in Section 2.2.1.4. In this work we adopt two variations: overlap area, that measures the amount of intersected area; and overlap rate, that measures the relation between intersected area and the smaller area (Fan et al. 2014).

Based on the study of Ruiz-Lendínez et al. (2013), we have combined a set of descriptors against polygon-pairs using weights determined by means of a genetic algorithm (GA). Ruiz-Lendínez et al. (2013) identified five relevant descriptors to matching areal objects: minimum bounding rectangle (MBR), which we use the relation between intersection area and the smaller MBR; geometric properties: perimeter and area; moment of inertia considering only the exterior ring; and the Arkin Graph Area, the region below the turning function defined by Arkin et al. (1991). These descriptors are combined in a single measure in order to describe the similarity between area objects varying from 0 (dissimilar) to 1 (similar).

The context measure used in areas is the same used for point matching (Section 3.1.1.1). Each area object, in both comparing datasets, is converted to its centroid, and these sets of points area compared using the geographic context measure. This kind of measure was used in other studies facing matching of area objects (Samal et al. 2004, Zhang et al. 2014).

At last, we selected two line distances to assess the areas: short-line median Hausdorff (Tong et al. 2014) and partial discrete Fréchet (Devogele 2002). In these cases, we use the exterior ring of each polygon in order to calculate the linear distances. Polygon holes are ignored. These measures were selected in order to assess its feasibility in matching of area objects.

3.1.3.2. Factor: matching method

There are manifold methods to deal with the geospatial matching problem (Xavier et al. 2016a). In this study we define a matching method using three elements: a set of measures, their associated thresholds, and the selection criteria.

There are many ways to use a similarity measure in a matching method. It can be used alone, combined with other measures in a single value, or used as an exclusion rule. The first case is the simpler and can be found in many studies (e.g., Beeri et al. 2004, Tong et al. 2014). The combination of several measures in a single similarity value is other popular strategy, as can be seen in the studies of Walter and Fritsch (1999) and Zhang and Meng (2008). The measures can also be used as an exclusion rule, i.e., if the measure does no reach a determined threshold, the other similarity measures are not considered (e.g. Kang et al. 2015).

After the measure (or the set of measures), other factor that affects the performance of a matching method is choosing the thresholds for these measures. According to Xavier et al. (2016a) the threshold selection is almost ubiquitous in geospatial matching solutions, except for those techniques based on optimization (e.g., Li and Goodchild 2010). Hence thresholds are part of a matching method, and in this study we consider the thresholds associated to each measure (or to a combined measure).

The last element that compounds the matching method is the selection criteria. This element defines if two assessed features should be marked as a 'match' or not. In this study we consider that a selection criteria is applied only inside the given thresholds, and we adopt two criteria: both nearest, and closer.

In the both nearest criterion two features in different datasets (e.g. datasets A and B) are marked as a match when the maximum similarity (or minimum distance) occurs in the forward matching (from dataset A to dataset B), and in the backward matching (from dataset B to dataset A). I.e., we only mark a matching pair a:b, $a \in A$, and $b \in B$, if a points to b and b points to a. Figure 3.11(a) illustrates an example with the point datasets A and B, the Euclidean distance as measure, and the similarity threshold t. Using the both nearest criterion we can mark the matching pairs a1:b1 and a3:b2 because they are the corresponding closer elements in both directions (from A to B, and from B to A). The object a2 has no match because its closer object is b1, but it is not a mutual relation since the closer object from b1 is a1. This selection criteria has the inherent drawback to permit only 1:1 matches.

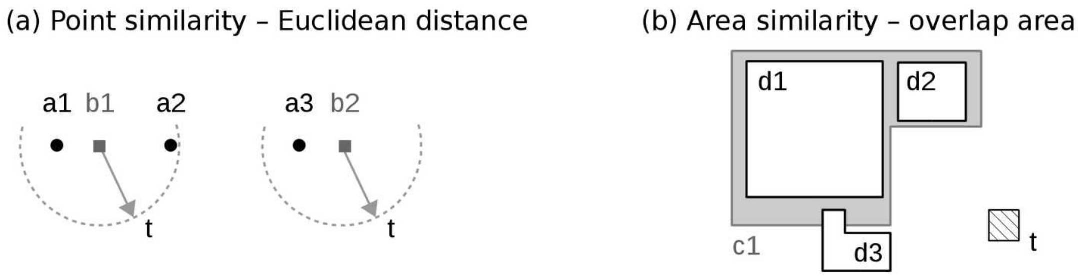


Figure 3.11. Selection criteria example.

The selection criteria 'closer' marks two features as a match when the maximum similarity (or minimum distance) occurs in the forward matching or in the backward matching. Yet considering the example in the Figure 3.11(a), if we use the closer criterion the selected matching pairs are a1,a2:b1 and a3:b2. Considering the area datasets C and D in the Figure 3.11(b), if we use the overlap area measure with the threshold t, the closer criterion will select the matching pairs c1:d2,d3. In this case, the object d3 has no match because the intersection area between it and c2 is below the threshold. The closer criterion permits us handle m:n corresponding cases.

3.1.3.3. Factor: morphology

In this study we consider the morphology of linear and areal objects as the third factor that may affect the performance of geospatial matching procedures. In order to assess this factor we adopted a roughness classification for lines and we have developed a roughness classification for areas. There are five classes for line roughness: (1) very smooth, (2) smooth, (3) sinuous, with stable directionality, (4) sinuous, with variable directionality, and (5) very sinuous. There are four classes for area roughness: (1) very simple, (2) simple, (3) complex, and (4) very complex. These morphology classifications are detailed in the Section 3.1.2.2.

3.1.3.4. Factor: geographic context

The fourth factor considered in this experiment is the geographic context. Samal et al. (2004) defined geographic context as a concept beyond the topology of objects, which takes into account the spatial relationships between features. Hence we investigate the influence of different neighbourhoods over geospatial data matching techniques for point and area data.

In this study we defined three classes for the geographic context: (1) uncertain, (2) intermediary, and (3) distinct. A class 1 context (uncertain) means that there are other objects in the considered dataset with similar contexts. A distinct context (class 3) if for those objects whose context is almost unique, i.e., there is no way to mistake its context 'signature' with others. The intermediary context (class 2) is for those objects that do not fall into previous classes (uncertain or distinct), so there is some ambiguity in their context. Figure 3.12 illustrates this classification.

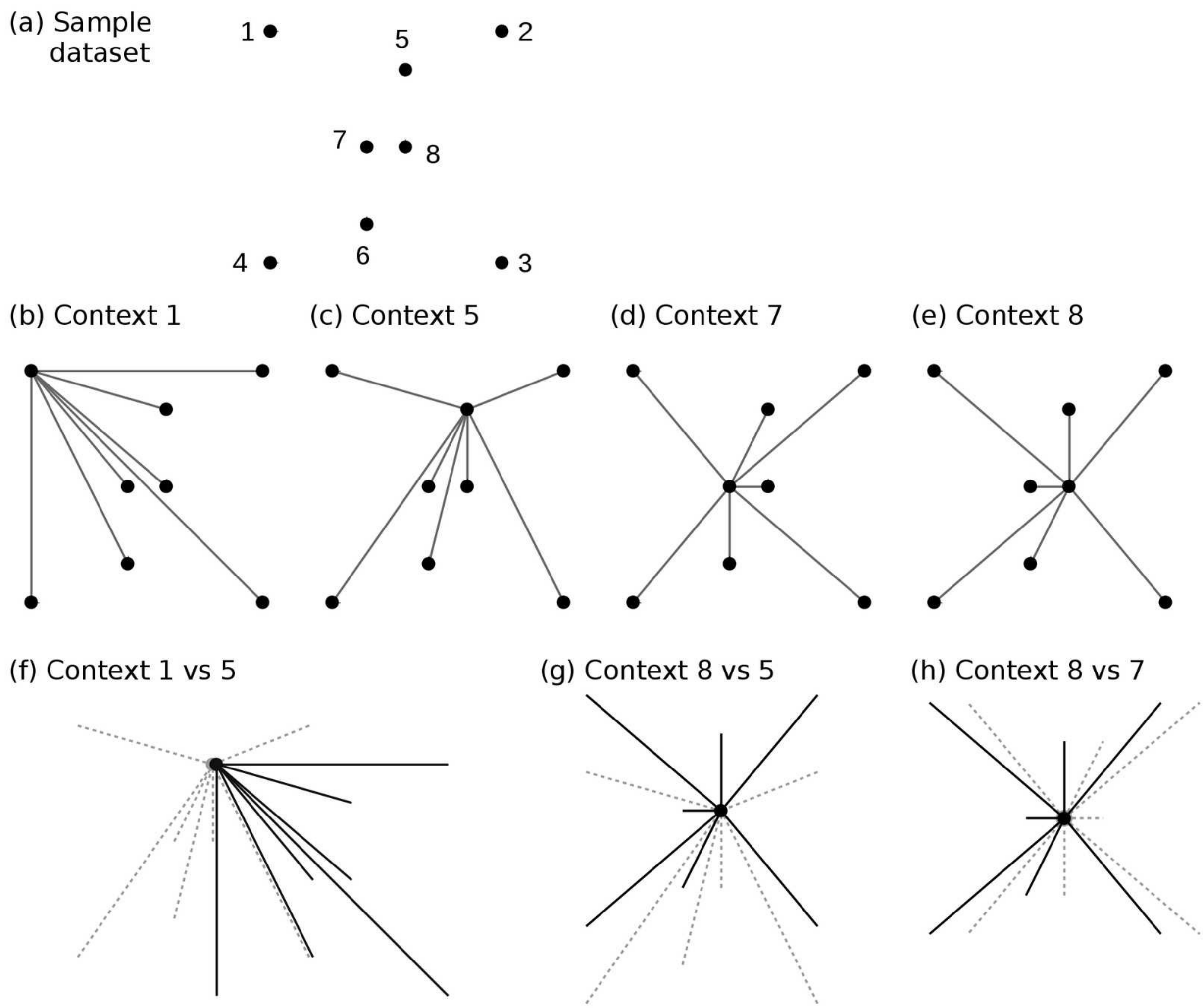


Figure 3.12. Examples of the geographic context classification. (a) Sample dataset. (b-e) Context for points 1, 5, 7 and 8. (f) Context of point 1 (solid black line) compared with context of point 5 (dashed grey line). (g) Context of point 8 compared with context of point 5. (h) Context of point 8 compared with context of point 7.

Considering the dataset in Figure 3.12(a) with eight objects. So, we can draw the context for each point using lines that represent the relative position of all other objects, as we can see in Figure 3.12(b-e). If we consider the 1's context with 5's context (Figure 3.12(f)), we can see that these contexts are very different, so we can consider a distinct context (class 3). Considering the contexts of objects 7 and 8 (Figure 3.12(h)) there is some similarity between them, so we have an uncertain context (class 1). But if we consider the context of objects 8 and 5 in Figure 3.12(g) we can see that there are some commonalities and discrepancies, so this is an intermediary context (class 2).

As Zheng and Doermann (2006) pointed out, defining the concept of neighbourhood for a point set is not a trivial task. In order to assess the geographic context of features we propose to use the shape context descriptor developed by Belongie et al. (2002). This descriptor was used in this study as a basis for a similarity measure (see Section 3.1.1.1).

The method to classify the features according its geographic context works as follows: First we calculate the shape context signature (histogram) for all objects in a dataset (centroid for areas). Then we compare the shape context of a feature against all other shape contexts using the cost function, which vary from 0 (similar) to 1 (dissimilar), and we select the smaller cost. After that we have a list of tuples (object1, object2, minimum cost). Using this list we can classify those objects according our three context classes (uncertain, intermediary, distinct).

In order to define the 'cut-off' values whose determine the context class we prepared an on-line quiz with 50 randomly selected samples and we asked for ten GIS experts to answer it. The aim was to identify cut-off values that reflected the human judgement about geographic context. The experts where oriented to indicate the context class (uncertain, intermediary, or distinct) for each sample. The results of this quiz were used to define the cut-off values. In this study, the experts where invited from academia (Universidad de Jaén, Spain) and industry (Brazilian Army Geographic Service).

3.1.3.5. Factor: systematic disturbance

Systematic disturbance are the fifth controlled factor investigated in this study. The aim is identifying the influence of intentional systematic perturbations over some geospatial data matching procedures. In our method we generated controlled systematic disturbances over original data using an affine transformation, which is able to represent translations, rotations, and scaling (dilation or shear). The systematic disturbance method is detailed in the Section 3.1.2.3. This factor has many levels according to the distinct transformations generated.

3.1.3.6. Factor: random disturbance

The last controlled factor in our design of experiment for geospatial data matching is the random disturbance. The aim is to assess the robustness of matching methods in the presence of controlled random perturbations. In our study we propose a new method to disturb geospatial data using vector fields created for a given standard displacement. The random disturbance method is detailed in the Section 3.1.2.4. The levels in this factor are straight related to the predetermined standard displacements.

3.2. Internal matching

Internal matching refers to finding correspondences between internal parts of features, more specifically vertices of lines or polygons. There are few matching methods focused in this actuation level, as we can see in Xavier et al. (2016a). In this study, we are proposing a new method for matching geospatial data at internal level based on the shape context descriptor from Belongie et al. (2002). The proposed method permits to establish the correspondences of vertices from linear or areal data.

The method works as follows: As of other internal matching procedures (Huh et al. 2011, Ruiz et al. 2015) this method requires a previous matching at feature level in order to identify objects' pairs (lines or areas) (Figure 3.13(b)). Any feature matching method can be used, including those that support many-to-many corresponding case (m:n). After determining the features' pairs, the next step is extract the relevant vertices of those objects (Figure 3.13(c)). In this method, we consider *relevant* vertices those that, considering the anterior and posterior vertices, form an angle greater than a given threshold ang_{th} (Figure 3.13(d)).

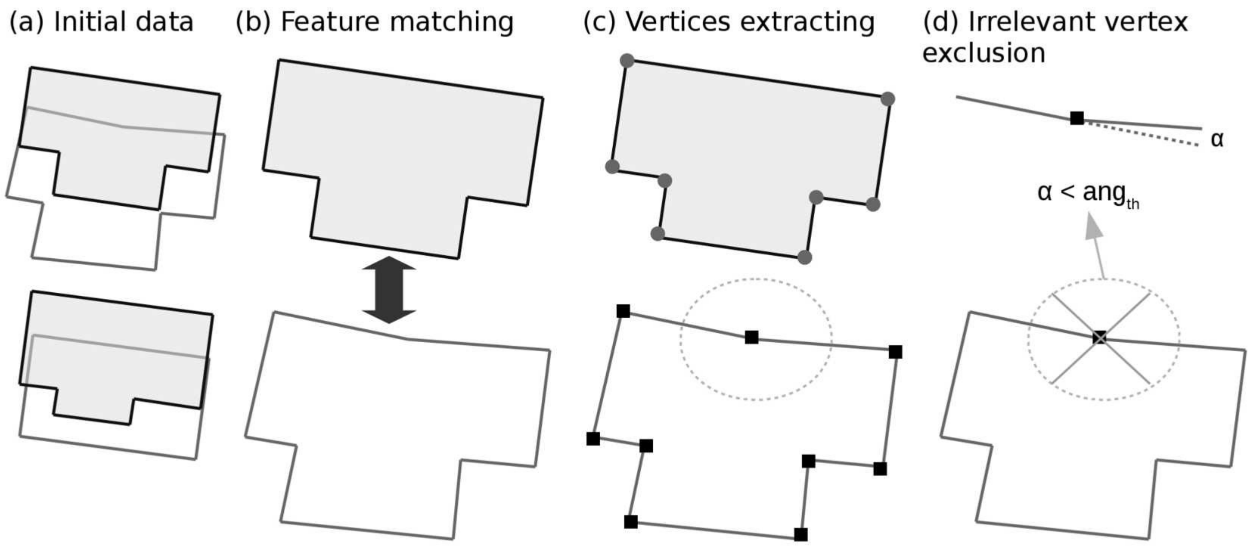


Figure 3.13. Internal matching method. (a) Initial datasets. (b) Feature matching. (c) Extracting relevant vertices. (d) Vertices with an angle below the angular threshold are not used to point matching.

This list of relevant vertices is used to compound a list of points that will be submitted to a point matching procedure based on the geographic context measure described in Section 3.1.1.1. Each point receives an object identifier (OID) based on the source object (line or area) that originated the point. For points from areas, the OID receives the area's OID, followed by the ring number (0 for exterior ring, and 1..n for internal rings), then the number of vertex inside the ring, beginning with 1, like the SFS specification (Herring 2011). Similar procedure is used for points from lines, where OID receives the line's OID followed by the number of vertex inside the line. The separator between values is the point ('.'). For instance, the OID of a point that is the second vertex in the exterior ring of polygon whose OID is '1234' is '1234.0.2'.

Many times the list of points from compound objects (lines or areas) does not contain enough points to form the context signature of each one. So, these point sets are densified in order to reach a pre-determined quantity of points, or a minimum distance between coordinates, which is necessary for the effectiveness of the geographic context measure. Figure 3.14 illustrates how this densification generates sufficient points for the geographic context measure, since this measure uses the number of points in each bin in order to create the context signature of each point.

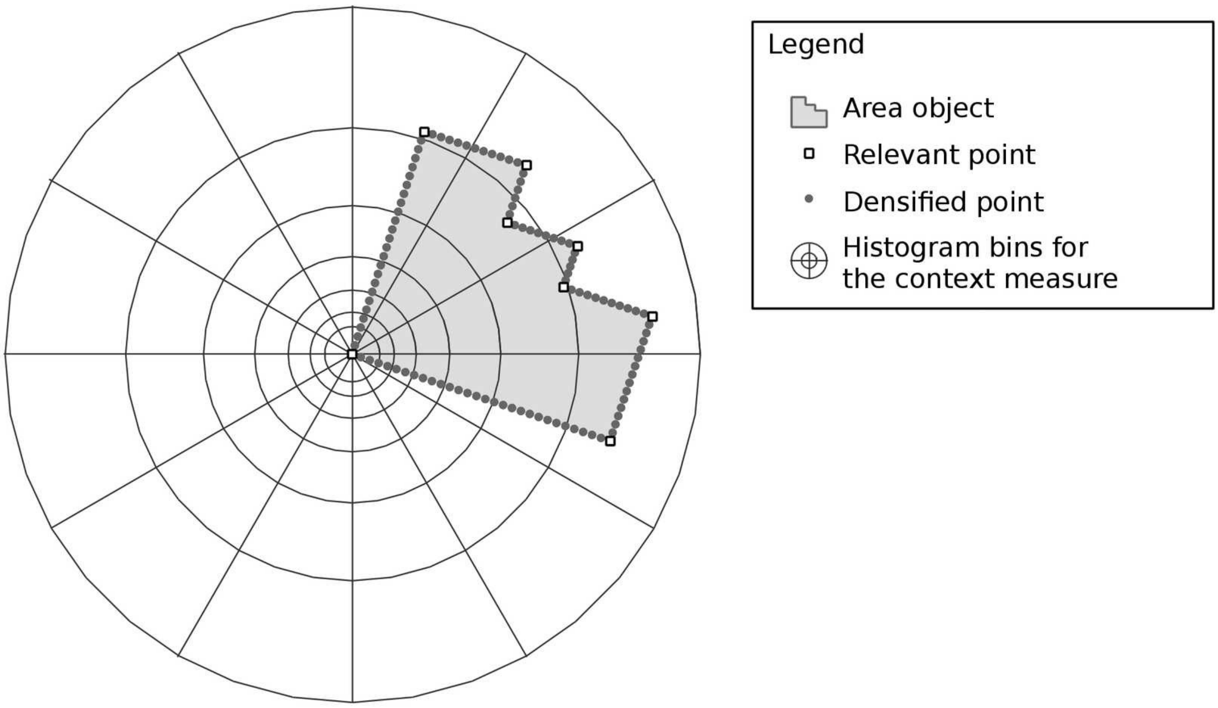


Figure 3.14. Densification of points inside an area feature in order to obtain points to compose the context signature of the relevant points, which are submitted to the internal matching procedure.

With the sets of points representing the features' parts in each dataset, now we can use the geographic context measure to calculate the cost function between point pairs to find the correspondences. In order to increase the precision of this method we shall apply an exclusion criterion after finding the corresponding point pairs. We propose to use the difference between the gradients of point pairs as an exclusion criterion, i.e., if the difference between the gradients reaches a value greater than the threshold $grad_{th}$, it will not be considered a point pair. Figure 3.15 shows an example of how the gradient threshold can be applied to discard two matches with similar geographic context but in fact they do not represent the same points pairs.

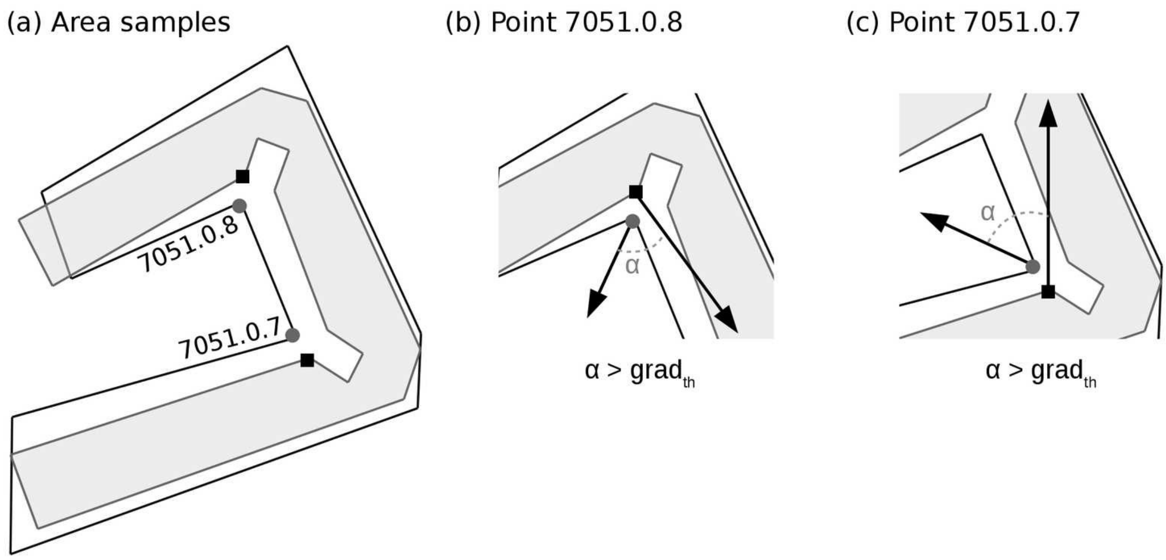


Figure 3.15. Gradient threshold in internal matching. (a) Two area samples. (b-c) Two point pairs previously marked as 'match' are discarded because the angular difference between their gradients is beyond the gradient threshold.

In this framework for quality evaluation using web services we are primarily adopting the quality model described in the Brazilian standard CQDG (DCT 2016a). This quality model provides a point-based method in order to control the positional quality. Thus, an internal matching procedure becomes a crucial element of this framework.

3.3. Quality evaluation module

The quality evaluation tier represents the kernel of this architecture towards the quality assessment of geospatial data using web services. In this tier are implemented the quality evaluation procedures (internal and external), and the quality report that is sent to the requester. In this part of architecture we develop internal and external quality evaluation procedures described in the CQDG standard (DCT 2016a) for products of type vector geospatial datasets. The selected procedures are presented in Table 3.1.

Table 3.1. Quality evaluation procedures considered in the evaluation tier.

ID	Method type	Sampling	Scope	Quality element	Measure
211	Internal	Full inspection	All points	Topological consistency	Rate of invalid points (CQDG:211)
212	Internal	Full inspection	All lines	Topological consistency	Rate of invalid simple lines (CQDG:212)
213	Internal	Full inspection	All areas	Topological consistency	Rate of invalid polygons (CQDG:213)
215	Internal	Full inspection	All areas	Topological consistency	Rate of invalid overlaps (CQDG:215)
101	External	Sampling	Dataset	Commission	Rate of excess items (CQDG:101)
103	External	Sampling	Dataset	Omission	Rate of missing items (CQDG:103)
301	External	Sampling	Dataset	Positional accuracy	Planimetry evaluation (CQDG:301)
302	External	Sampling	Dataset	Positional accuracy	Altimetry evaluation (CQDG:302)

The selected internal procedures refer to topological consistency. An on-line version of evaluation methods in this quality element has been implemented in some data integration projects (e.g. Tagg 2015). The four evaluation procedures considered (ID 211-215) do not require any sampling, so a full inspection is performed. The first three procedures (ID 211-213) refer to the validity of geometries in relation to the Simple Features Standard (SFS) (Herring 2011), with the difference that a line should be a simple line (without self-intersections). The fourth procedure refers to find overlaps between areas in the same layer, which may represent an error according to the data specification.

There are two direct external evaluation procedures (ID 101, 103) that refer to completeness: one procedure for commission and other for omission. Methods of this quality category require an object sampling based on a tesselation of test data according as its scale (4 cm in the test data scale). There are three sampling strategies: isolated lot (ISO 2859-2), lot-by-lot (ISO 2859-1), and full inspection (100%). The ISO strategies are provided in the quality model (DCT 2016a), and the full inspection was added as a third option.

The object sampling is implemented as described in the quality model: The first step is determine the sample size according to the sampling strategy. The second step is to create a tesselation in test data according its scale (Figure 3.16(b)). Then the cells of this tesselation are randomly selected and all objects inside each cell are computed till reach the initial sample size (Figure 3.16(c)). When the test sample done, we use the same cells to find the reference sample in the reference dataset (Figure 3.16(d)). With the sampling done, the method calls the feature matching module and finds the matching pairs between test and reference. For the quality evaluation, it calculates the rate of excess items (commission) and the rate of missing items (omission) (Figure 3.16(e)).

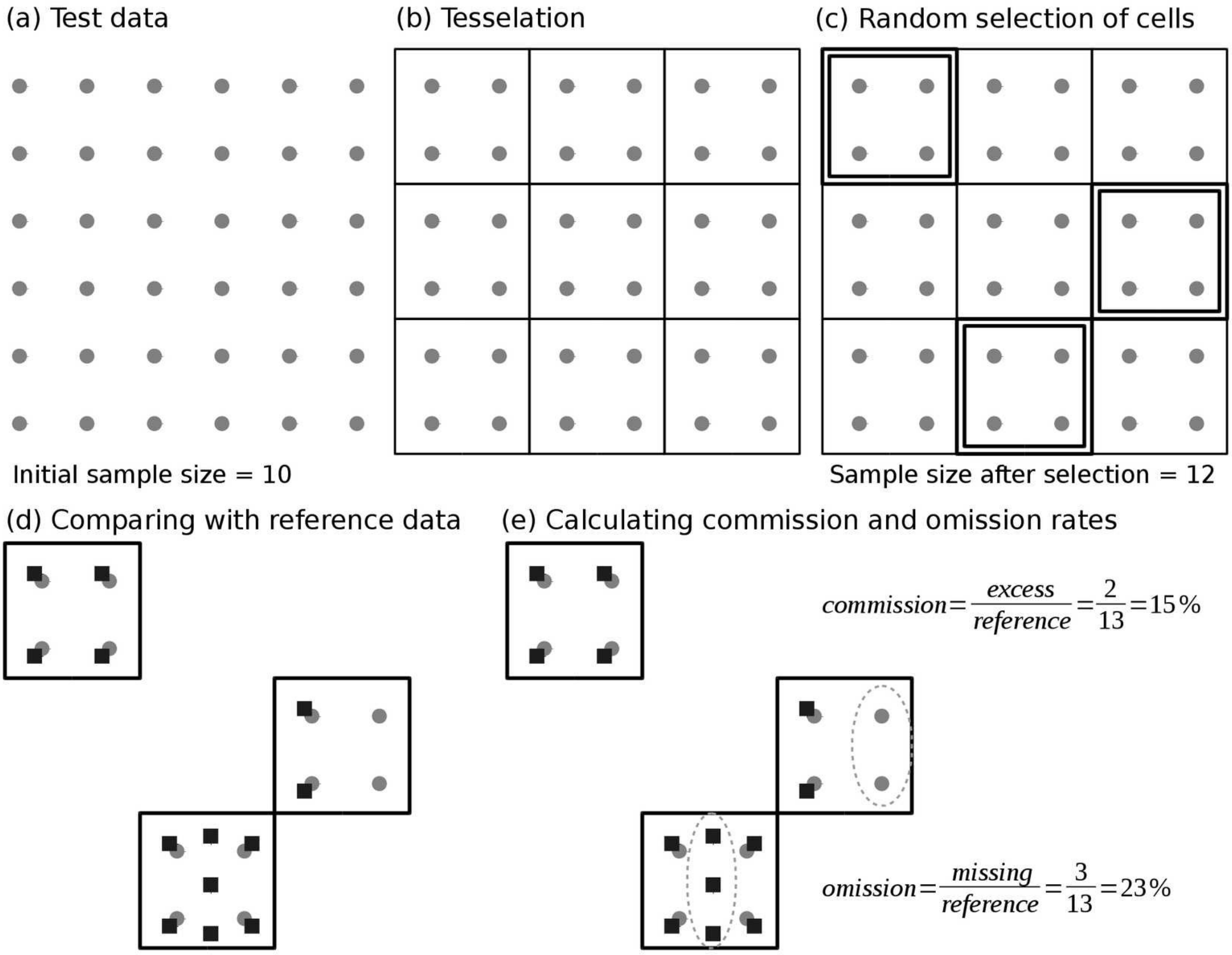


Figure 3.16. Completeness evaluation procedure. (a) Test data. (b) Creating cells over test data using the scale. (c) Cells are randomly selected, all objects are included. (d) Reference data within sample cells are used to compare the test dataset. (e) Quality measures.

The last external evaluation procedures (ID 301, 302) refer to positional accuracy. The quality model (DCT 2016a) provides different methods to assess the planimetric and altimetric quality using points. These methods return a quality category named *PEC* (*Padrão de Exatidão Cartográfica* – cartographic accuracy standard) according to the 90% percentile of errors and the corresponding root mean square. The PEC can assume five values: A, B, C, D and nonconforming (or '0'). As a point-based method, in order to use linear or areal features it is necessary to use the internal matching as described in Section 3.2. The matching should occurs prior to the sampling to avoid selecting unmatched points at sampling phase. These quality methods require a positional sampling procedure that uses a tesselation over test data according to the test data scale, similar to that used in object sampling. The population which is applied the sampling procedure are the cells in this tesselation that have at least one point to be assessed. There are four sampling strategies: isolated lot (ISO 2859-2), lot-by-lot (ISO 2859-1), one-by-cell, and full inspection of points. The ISO strategies are provided in the norm, while one-by-cell means full inspection of cells, i.e., all cells with points should be used. The last strategy, full inspection of points, is applied to consider all available points in the positional accuracy assessment.

Figure 3.17 illustrates an example of how works the quality evaluation for positional accuracy considering an ISO's sampling strategy.

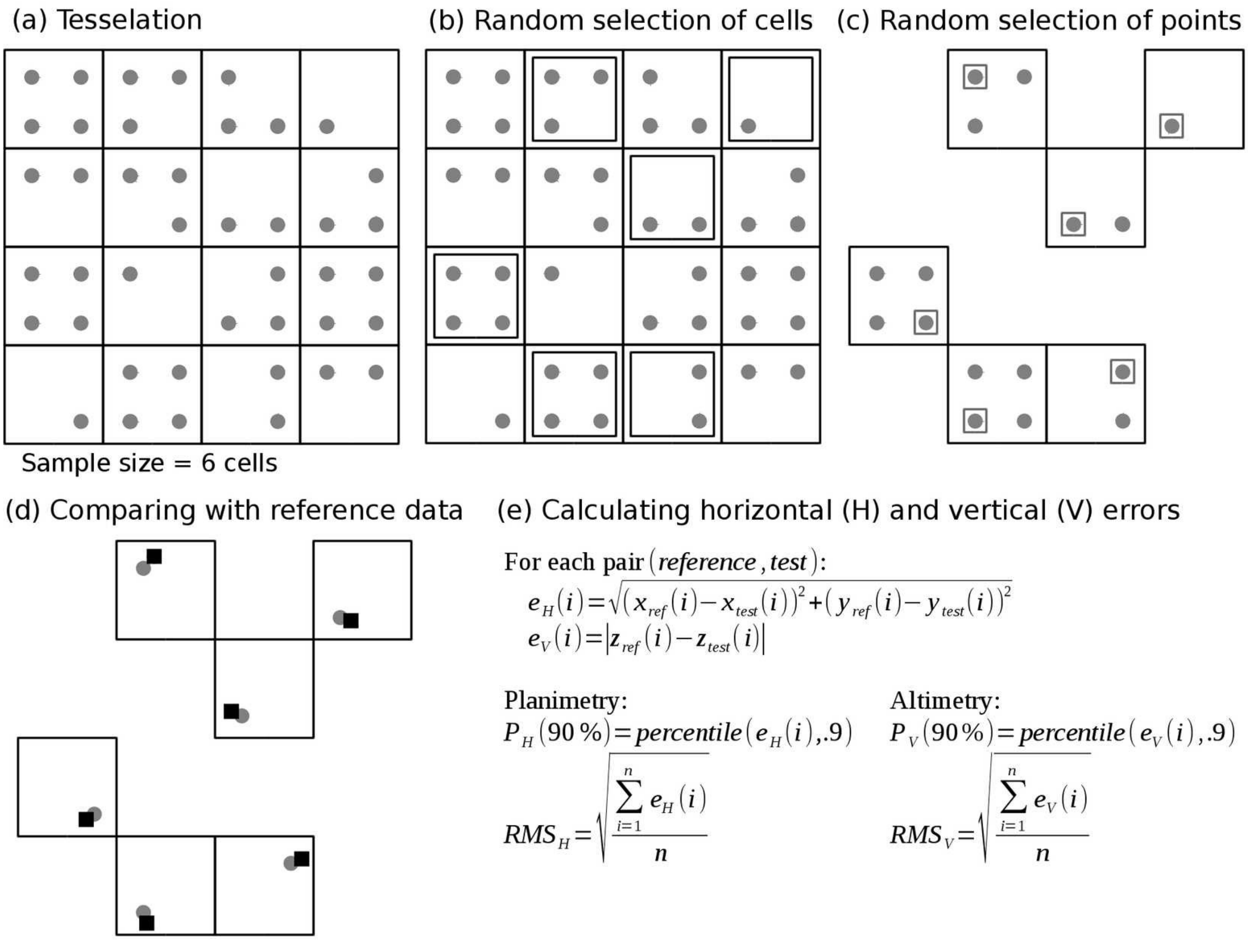


Figure 3.17. Positional accuracy evaluation procedure. (a) Creating cells over test data. (b) Cells are randomly selected according to sample size. (c) In each selected cell, one point is randomly selected as sample. (d) Reference data are used to compare the samples. (e) Quality measures.

Following Figure 3.17(a), the first step is to create a tesselation in test data according its scale, which has a resolution of 4 cm in the scale of assessed data (DCT 2016a). The second step is to determine the sample size according to the sampling strategy taking into account the number of cells with points. This sample size represents the number of cells that will be considered in the evaluation. Then the cells of this tesselation are randomly selected according to the sample size (Figure 3.17(b)). In the following, for each selected cell, the system randomly select one point inside the cell as a sample. At the end we have a list of *sample size* points that will be compared with reference data (Figure 3.17(d)). Finally, the system calculates the values of 90% percentile for planimetric errors and altimetric errors (when available) and the correspondent root mean square. These two values are used to determine the quality category (PEC) of each test dataset.

3.4. WPS tier for quality control

In the proposed architecture of quality control service, the WPS tier plays the role of interface between clients and the evaluation procedures (see Section 3.3). The initial version of the WPS tier was presented in Xavier et al. (2015b). Quality evaluation procedures often involve complex tasks and people from different organizations or departments. Facing this situation we have two design principles: *interoperability* and *simplicity*. The interoperability principle indicates that the WPS tier should follow the WPS specification and schemas in order to permit a standardised way of communication. The simplicity principle leads us to avoid unnecessary issues in the processing itself, so the processing 'part' should be as straight as possible. The WPS tier should manage all communication issues, validation procedures, and client-server tasks. Therefore the designer of the Evaluation tier can only focus on assessment procedures.

In addition to the classes described in the WPS specification we propose the creation of three new interfaces: AbstractProcess, AbstractComplexData, and AbstractExecuteResponse.

AbstractProcess is an interface that all concrete process should implement in order to permits its use under the architecture. The interface is represented in Figure 3.18 using the UML notation. The abstract class has one attribute: the description of the process using the WPS semantic. The interface has two methods: *getLanguages* and *getDescription*; and two pure virtual methods: *execute* and *createDescription*. The *getLanguages* method is used for GetCapabilities operation, and the *getDescription* is used in all operations to return a summarized description for the process (in GetCapabilities), or a more complete description for the DescribeProcess operation and the Execute operation.

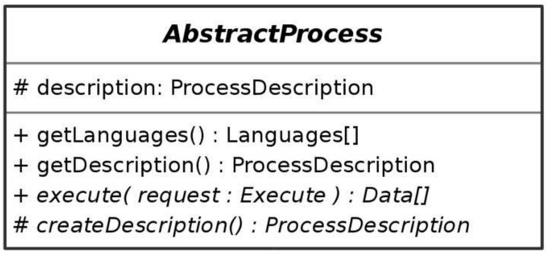


Figure 3.18. The AbstractProcess interface (Xavier et al. 2015b).

A concrete process should implement *createDescription* and *execute* methods. The *createDescription* returns a full description of the process, which can be hard-coded in the implementation, or can be read from a configuration file, like the deegree WPS (Kiehle et al. 2007). The *execute* method effectively runs the processing what the implementation was designed to do. It is possible to note that the execute method does not return an ExecuteResponse object but an array of *Data* objects. The goal is to avoid that the processing handles the final response, but just run its job and returns the processed data. In this architecture we are using the design pattern Abstract Factory (Gamma et al. 1995) in order to manage the processes in a server. So, the processes should be registered into a 'factory' prior to be used.

AbstractComplexData is the interface that implements the access to specific data drivers, like ESRI Shapefiles (ESRI 1998), Geography Markup Language (GML) (Portele 2007), or imagery in GeoTIFF (Ritter and Ruth 1995). AbstractExecuteResponse is the interface for response to an Execute operation request. This interface has two concrete implementations: ExecuteResponse and RawDataResponse. This is necessary because the final response of a processing task may be or a standard ExecuteResponse either a raw data response, in some predefined format, if the client requests in this way. This is other reason because the execute method in AbstractProcess interface returns an array of Data instead of an ExecuteResponse.

When the WPS server receives an Execute request it acts as shown in the sequence diagram in Figure 3.19. First the Server calls the ProcessFactory that instantiates the correct process using the process identifier (ID) informed by the client. Then Server requests to the Process its description. Process instantiates (or read) its description and return it to the Server. Server sends the Execute request to the Description in order to validate it. If any problem occurs, Description throws an exception. After the validation procedure, Server calls Process to run the processing task, and Server receives the array of Data objects resulting from the process. Finally, Server uses the returned Data and assembles the final response to the client, which can be a standard XML response (ExecuteResponse) or in other format (RawDataResponse), and send it to the requester.

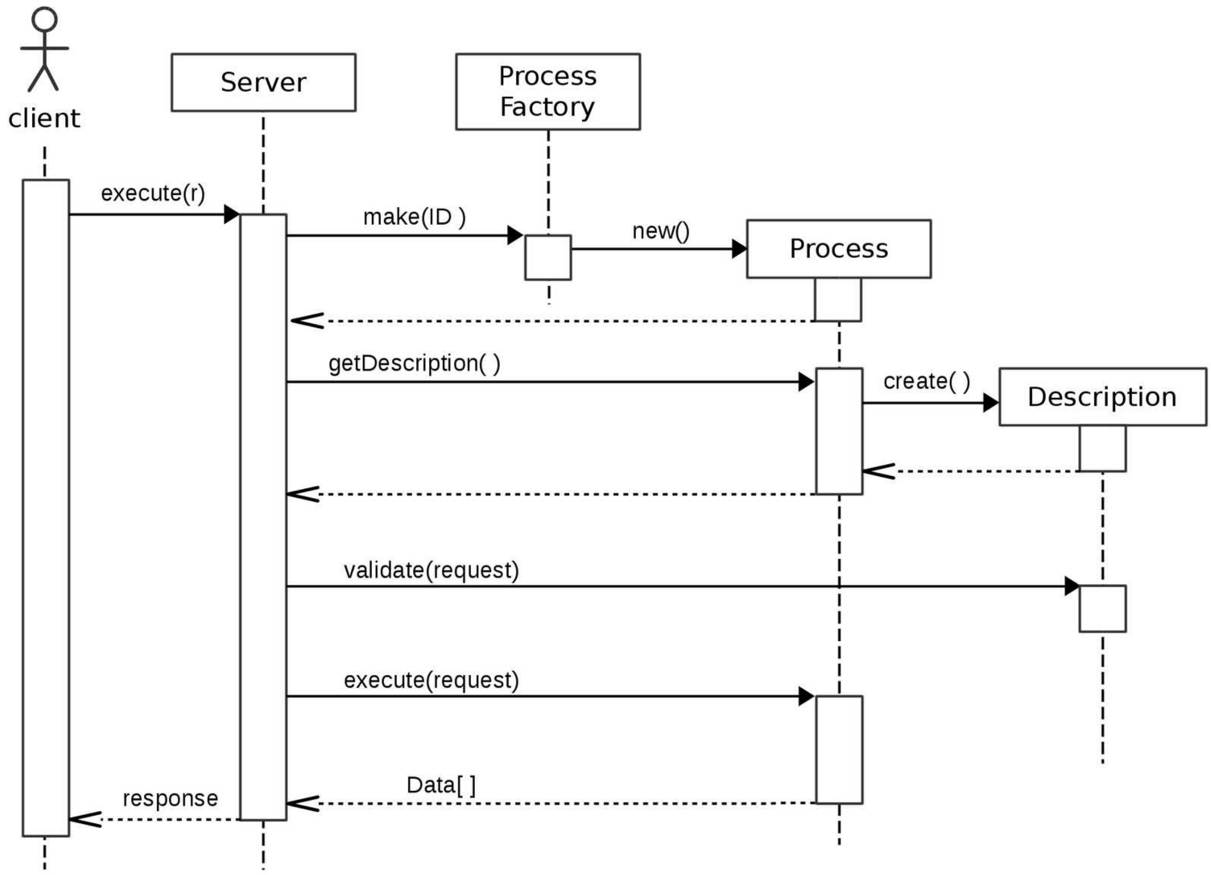


Figure 3.19. Sequence diagram representing an WPS Execute operation.

Quality assessment for geospatial data frequently involves various tasks in a set of processing instructions. Hence it is interesting that the developer of these procedures lays emphasis only in the processing itself, without losing time with other issues. The proposed WPS tier in our quality assessment architecture intends to avoid these losses while guaranteeing the interoperability. One feature of this architecture is the loose coupling between WPS protocol and the process itself.

This last tier concludes the proposed architecture for the quality control service proposed in this study. In the following we test the validity of the approach by means of controlled experiments.

CHAPTER 4
EXPERIMENTS, RESULTS AND DISCUSSION

This chapter presents the experiments executed in order to validate the proposed framework for geospatial data quality evaluation through web services. The essays are designed to assess the proposed framework using both real and synthetic data. As explained in Chapter 3, the proposed architecture is intended to be general for automatic quality assessment for geospatial data, which means that it should be applied independently of datasets or software platform. Hence the first section introduces the material used in the experiments: geospatial data and software. The next sections presents the experiments for each component of the proposed architecture, as illustrates the Figure 4.1.

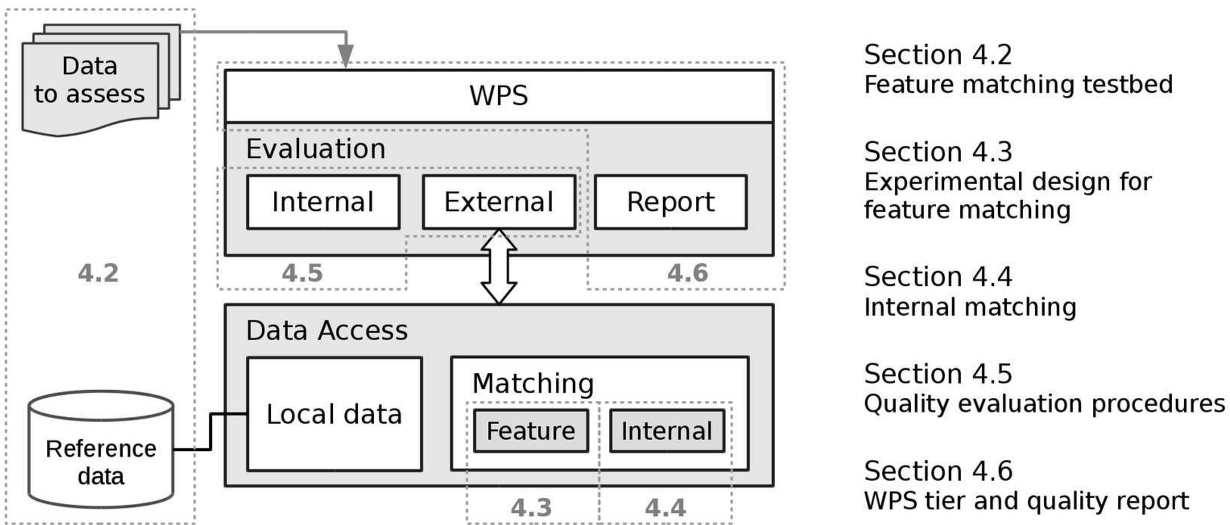


Figure 4.1. Overview of the experiments for each part of the proposed architecture. Dotted gray boxes indicate the section number within this chapter.

From a bottom-up point of view, Section 4.2 presents the results of the creation of feature matching testbed described in Section 3.1.2. This testbed is compound by groups of datasets that are used to test the other parts of the quality control service. As explained before, all external quality evaluation procedures require the matching between features in considered datasets in order to compare them. However, there are many approaches dealing with this issue in the GISciences literature. Hence in Section 4.3 we apply the concepts of a designed experiment for feature matching described in Section 3.1.3 over the testbed created in Section 4.2.

In this study we adopts the Brazilian standard (DCT 2016a) as the quality model. This standard establishes positional accuracy procedures based on points. In order to increase the quantity of points in the positional evaluation, which includes linear and areal objects, we need a matching method able to actuate at the internal level of these objects. In Section 4.4 we present the experiments for checking the internal matching method proposed in Section 3.2 as part of our solution.

Section 4.5 presents the experiments over the kernel of quality control service: the evaluation tier. The experiments in this phase test the capability of automation for the quality control procedures described in Section 3.3. We use the testbed created in Section 4.2 as test and reference data. We use the feature matching methods tested in Section 4.3 to assess the external methods (completeness and positional accuracy). For the point-based positional accuracy procedure we use the internal matching method tested in Section 4.4.

Section 4.6 describes the essays relative to the publication of the quality evaluation procedures developed in Section 4.5 as a Web Processing Service (WPS) following the design principles established in Section 3.4. This section also presents the results about the quality reports generated in the Evaluation tier, in order to show the flexibility of this standardized interface for web services.

Each section ends with a discussion of findings for the corresponding experiments.

4.1. Material

In this research project we use *R* as the statistical computing tool. R is a language and also an environment focused on statistic tools and graphics (R Core Team 2014). Other relevant materials are the geospatial data used to test the quality control service, and the developed software that effectively implements the concepts proposed in this study. Figure 4.2 illustrates where these materials were applied in each part of the architecture. The following subsections detail the specifications for geospatial data and software used in the experiments.

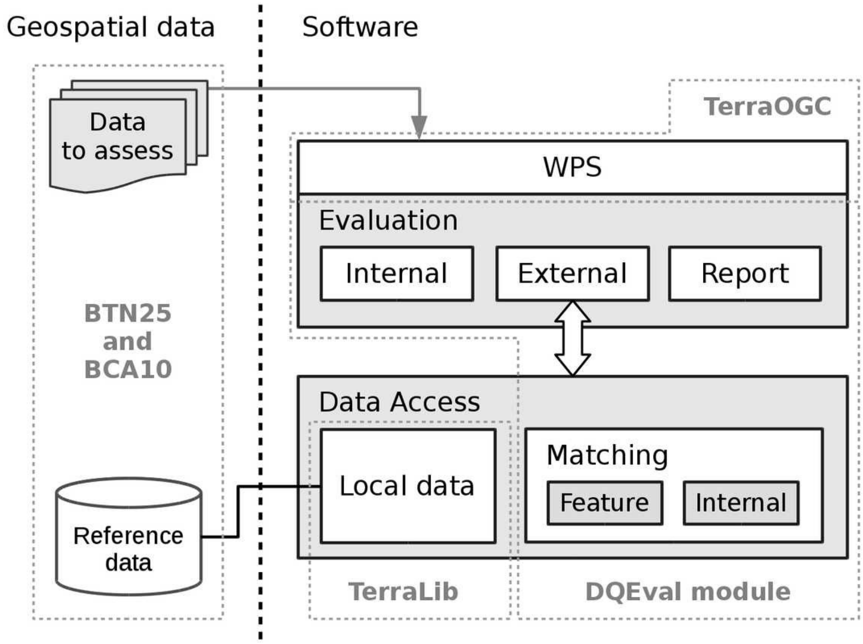


Figure 4.2. Overview of the material used in the experiments and its relation with the architecture of the quality control service.

4.1.1. Geospatial data

In this research project we adopted geospatial datasets built up from mapping data produced by official Spanish mapping agencies for Andalucía, southern Spain. This area was chosen because the Universidade de Jaén is located there, and because there are freely available data covering this area. We used 1:25,000 data from the Base Topográfica Nacional 1:25,000 (BTN25) of national mapping provided by the Instituto Geográfico Nacional of Spain (IGN 2015). We used 1:10,000 data from the Base Cartográfica de Andalucía 1:10,000 (BCA10) of regional mapping provided by the Instituto de Estadística y Cartografia de Andalucía (ICEA 2015). Figure 4.2 shows that these data are decoupled from the solution itself, in order to maintain the generalness of solution.

For checking the matching pairs we used regional orthoimagery available at the web services of REDIAM (Red de Información Ambiental de Andalucía) (Junta de Andalucía 2017, Ortiz et al. 2010).

We selected different landscapes: coast and mountain, rural and urban. The following mapping sheets 1:25,000 were used to define the study: 0896-3, 0896-4, 1003-4, 0999-1, 0999-2, 0999-3, and 0999-4. Figure 4.3 shows an overview of selected mapping sheets over two Andalusian provinces: Huelva and Sevilla.

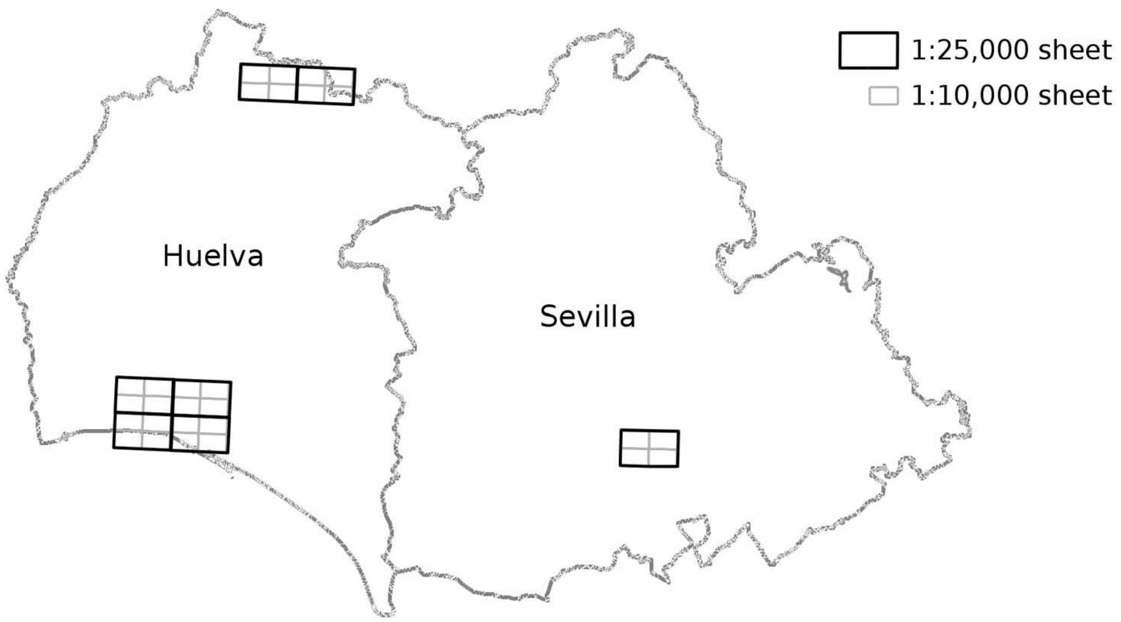


Figure 4.3. Selected mapping sheets for the experiment. There are seven 1:25,000 sheets and their 28 correspondences in 1:10,000. Huelva and Sevilla are provinces in Southern Spain.

4.1.2. Developed software

All software developed in this research project is based on the TerraLib library. TerraLib is an open-source GIS library developed by the Brazilian National Institute for Space Research (INPE) (Câmara et al. 2008). TerraLib code was developed using the C++ programming language, so C++ is used for all developments in this project. We used the TerraLib version 4.2.2 that can be found at the TerraLib repository (DPI 2013). TerraLib implements drivers for access many database management systems and data files, so it can be found in the Data Access tier of the proposed architecture (see Figure 4.2).

Inside TerraLib there is a subprojects named TerraOGC – a framework for Web-GIS development that has been used in web services research (Xavier and Meyer 2013) and development (Xavier et al. 2014). TerraOGC contains modules for many OGC specifications, like WMS, WFS, WCS, and GML.

For this research project the existing WPS module was improved in order to accommodate the design principles described in Section 3.4. As a part of WPS process was created a data quality processing module (DQEval) which contains most of the code related to this project. All developed code are released as a free software under the GNU General Public License version 3. It can be found on-line at its repository (DPI 2017).

4.2. Feature matching testbed

The first experiment deals with the creation of the feature matching testbed. This testbed is composed by four groups of datasets generated using the methodology described in the Section 3.1.2. We believe that this testbed is a valuable tool to be shared with other researches in the GIScience area, so we have submitted it to a public repository of scientific data (Xavier et al. 2017).

All geoespatial data are supplied in the ESRI Shapefile format (ESRI 1998). The projection system is UTM zone 28 North with datum WGS-84 (EPSG:32628). The list of matching pairs, composed of object identifiers (OID), is in plain text.

The following subsections detail the results for each group of datasets: (1) initial datasets: original mapping data; (2) morphology modified: synthetic datasets created with emphasis in some specific morphology class for lines or areas; (3) systematic disturbance: synthetic datasets created from affine transformations; and (4) random disturbance: synthetic datasets created over the influence of randomly generate displacement vector fields. The last subsection discusses the results of this experiment.

4.2.1. Initial datasets

The initial group is composed of six datasets: SP, SL, and SA for test data (scale 1:25,000), and BP, BL, and BA for reference data (scale 1:10,000), for point, line and area, respectively. Additionally, this group also contains the list of matching pairs in plain text. Each record is in the form: [OIDt][,OIDt]*:[OIDr][,OIDr]*, where OIDt is the OID in test data, and OIDr is the OID in reference data. For instance, the matching record '2009:3203,90386' means that the object 2009 in test data is corresponding to the objects 3203 and 90386 in reference data.

The BTN25 data were used to compound the test datasets (SP, SL and SA), and the BCA10 data were used to compound the reference datasets (BP, BL, and BA). Test and reference data are divided into nine regions. Each region was originated from the following mapping sheets 1:25,000: (S1) 0896-3, (S2) 0896-4, (S3) 1003-4, (S4) 0999-1 east, (S5) 0999-2 west and 0999-4 east, (S6) 0999-1 west, (S7) 0999-3, (S8) 0999-4 west, and (S9) 0999-2 east. We performed the manual matching between features for each region. This process is illustrated at Figure 4.4.

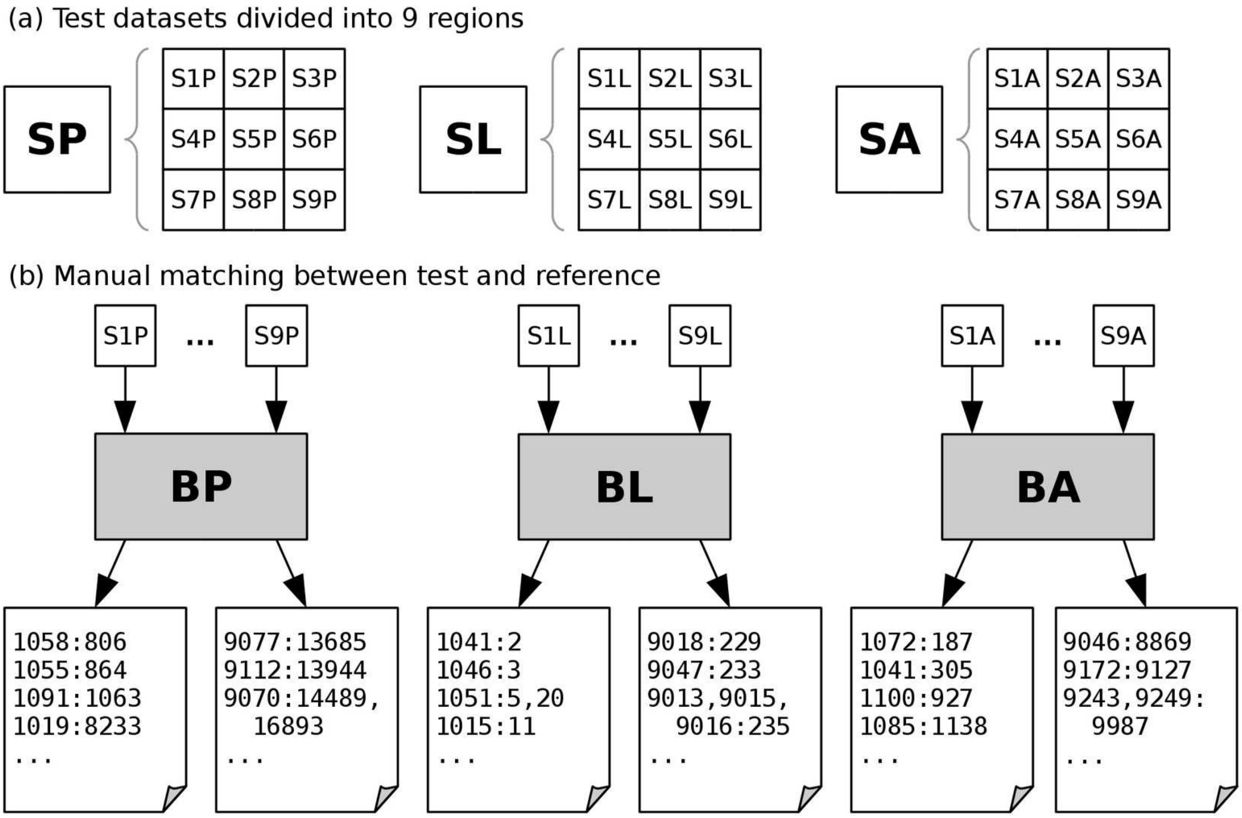


Figure 4.4. Manual matching process. (a) Test datasets are divided into nine regions. (b) Manual matching results in a list of matching pairs.

At the end of this procedure we had 27 sets (nine regions by three geometries) and their correspondences represented in Table 4.1.

In Table 4.1, the column *Geometry* refers to the type of geometry for the dataset. The column *Region* refers to the name of each region. The column *Size* represents the number of objects in each region. The *Matching pairs* columns indicate the number of matching pairs when comparing each test region (S#[PLA]) with each reference region (B#[PLA]). Due to the presence of multiple corresponding case (1:n and m:n) the number of matching pairs differs from the size of test dataset. For instance, the matching '100, 200:101, 102, 103' represents six matching pairs: 100:101, 100:102, 100:103, 200:101, 200:102, and 200:103.

Table 4.1. Regions considered in the initial dataset group and their sizes (Xavier et al. 2017).

Geometry	Region	Size	Matching pairs 1:1	1:n	m:n
Point	S1P	112	91	2	0
	S2P	115	100	6	0
	S3P	126	115	2	0
	S4P	119	98	4	0
	S5P	120	108	2	0
	S6P	120	94	4	0
	S7P	184	170	6	0
	S8P	201	188	0	0
	S9P	118	106	18	0
Line	S1L	133	50	158	29
	S2L	135	61	148	14
	S3L	140	67	128	34
	S4L	138	51	160	64
	S5L	135	37	109	92
	S6L	140	30	233	89
	S7L	140	57	120	54
	S8L	143	56	56	153
	S9L	151	87	93	0
Area	S1A	123	101	19	0
	S2A	140	87	81	0
	S3A	140	122	18	0
	S4A	150	120	33	0
	S5A	202	172	63	0
	S6A	160	107	44	0
	S7A	268	215	55	0
	S8A	253	202	61	8
	S9A	250	159	130	59

4.2.2. Morphology modified datasets

Using the concepts described in the Section 3.1.2.2 we generated five synthetic datasets for lines – one for each morphology class: (1) very smooth, (2) smooth, (3) sinuous, with stable directionality, (4) sinuous, with variable directionality, and (5) very sinuous. Table 4.2 presents the object count in the original and modified datasets (CL1-CL5) for lines.

Table 4.2. Number of objects in each dataset according to the line morphology classification (Xavier et al. 2017).

Morphology class	Dataset					
	Original	CL1	CL2	CL3	CL4	CL5
1	909	1206	564	571	909	909
2	193	8	642	49	28	193
3	112	0	37	626	165	16
4	29	29	0	8	152	96
5	12	12	12	1	1	41

Regarding the morphology classification for areal features: (1) very simple, (2) simple, (3) complex, and (4) very complex, we generated three synthetic datasets for areas, one for each morphology class except for class 4. Table 4.3 presents the object count in the original and modified datasets (CL1-CL3) for areas.

Table 4.3. Number of objects in each dataset according to the area morphology classification (Xavier et al. 2017).

Morphology class	Datasets			
	Original	CL1	CL2	CL3
1	1108	1630	1	0
2	202	0	1630	0
3	325	5	4	1635
4	51	51	51	51

4.2.3. Systematic disturbance datasets

Using the proposed methodology (Section 3.1.2.3) we elected four standard displacements in order to generate the systematic disturbances: 5, 12.5, 25, and 50 meters. These values were chosen taking into account the Brazilian standard for geospatial data quality (DCT 2016a). According to this standard, the maximum positional error accepted for 1:25,000 data vary from 7.0 to 25 meters. So we have chosen one value below this limit (5 m), two values inside (12.5 and 25 m), and another value above (50 m). These four standard displacements have originated 7425 different combinations for each geometry (point, line, and area).

The following tables bring the different configurations for translations (Table 4.4), rotations (Table 4.5), and scalings (dilation/shear) (Table 4.6) used to generate the synthetic perturbed datasets. Rotation and scaling transformations depend on the size of diagonal of considered datasets, so the amplitude of these transformations varies for each type of geometry.

Table 4.4. Translations generated for the four standard displacements (Xavier et al. 2017).

Configuration	Translation in X (m)	Translation in Y (m)	Configuration	Translation in X (m)	Translation in Y (m)
t0	0	0			
t1	5	0	t17	25	0
t2	3.53553	3.53553	t18	17.6777	17.6777
t3	0	5	t19	0	25
t4	-3.53553	3.53553	t20	-17.6777	17.6777
t5	-5	0	t21	-25	0
t6	-3.53553	-3.53553	t22	-17.6777	-17.6777
t7	0	-5	t23	0	-25
t8	3.53553	-3.53553	t24	17.6777	-17.6777
t9	12.5	0	t25	50	0
t10	8.83883	8.83883	t26	35.3553	35.3553
t11	0	12.5	t27	0	50
t12	-8.83883	8.83883	t28	-35.3553	35.3553
t13	-12.5	0	t29	-50	0
t14	-8.83883	-8.83883	t30	-35.3553	-35.3553
t15	0	-12.5	t31	0	-50
t16	8.83883	-8.83883	t32	35.3553	-35.3553

Table 4.5. Rotations generated for the four standard displacements. There are distinct angle values for each type of geometry (Xavier et al. 2017).

Configuration	**Pivot**	**Angle for Point data (rad)**	**Angle for Line data (rad)**	**Angle for Area data (rad)**
r0	lower left	0	0	0
r1	lower left	0.000591397	0.000521513	0.000574903
r2	centre	0.000591397	0.000521513	0.000574903
r3	upper right	0.000591397	0.000521513	0.000574903
r4	lower left	-0.000591397	-0.000521513	-0.000574903
r5	centre	-0.000591397	-0.000521513	-0.000574903
r6	upper right	-0.000591397	-0.000521513	-0.000574903
r7	lower left	0.00147849	0.00130378	0.00143726
r8	centre	0.00147849	0.00130378	0.00143726
r9	upper right	0.00147849	0.00130378	0.00143726
r10	lower left	-0.00147849	-0.00130378	-0.00143726
r11	centre	-0.00147849	-0.00130378	-0.00143726
r12	upper right	-0.00147849	-0.00130378	-0.00143726
r13	lower left	0.00295698	0.00260756	0.00287451
r14	centre	0.00295698	0.00260756	0.00287451
r15	upper right	0.00295698	0.00260756	0.00287451
r16	lower left	-0.00295698	-0.00260756	-0.00287451
r17	centre	-0.00295698	-0.00260756	-0.00287451
r18	upper right	-0.00295698	-0.00260756	-0.00287451
r19	lower left	0.0059139	0.00521509	0.00574897
r20	centre	0.0059139	0.00521509	0.00574897
r21	upper right	0.0059139	0.00521509	0.00574897
r22	lower left	-0.0059139	-0.00521509	-0.00574897
r23	centre	-0.0059139	-0.00521509	-0.00574897
r24	upper right	-0.0059139	-0.00521509	-0.00574897

Table 4.6. Scalings generated for the four standard displacements. There are distinct values of scaling factor for each type of geometry (Xavier et al. 2017).

Configuration	Scaling factor for Point data	Scaling factor for Line data	Scaling factor for Area data
s0	1	1	1
s1	1.00059	1.00052	1.00057
s2	0.999409	0.999478	0.999425
s3	1.00148	1.0013	1.00144
s4	0.998522	0.998696	0.998563
s5	1.00296	1.00261	1.00287
s6	0.997043	0.997392	0.997125
s7	1.00591	1.00522	1.00575
s8	0.994086	0.994785	0.994251

Rotations and scalings are considered for each region (S1-S9) in each dataset.

The names of the datasets in this group identify the configuration used by combining the geometry type (SP, SL or SA), translation (t0 to t32), rotation (r0 to r24), and scaling (s0 to s8). For instance, the dataset called 'SL_t19_r11_s7' represents test data of line type (SL) which were translated 25 m to the east, rotating -0.00130378 rads using the centre as pivot, and dilating using the factor 1.00522.

4.2.4. Random disturbance datasets

Using the proposed methodology (Section 3.1.2.4) we chose the same four standard displacements used in systematic disturbance: 5, 12.5, 25, and 50 meters. We adopted a field resolution of 4 km which represents 16 cm in our 1:25,000 test data. The last parameter is the sigma, which defines the internal variation of amplitude. Here we chose a value of ±10%. Hence we randomly generated 100 vector fields for each standard displacement, which results in 400 vector fields. Each vector field configures a geometric transformation that was applied for each type of geometry: point, line, and area.

The names of the datasets in this group identify the configuration used by combining the geometry type (SP, SL or SA), the standard displacement followed by the word 'random', and a count the represents the number of the vector field (001 to 100). For instance, the dataset called 'SA_random12.5_083' represents test data of area type (SA) which was influenced by a vector field with a standard displacement of 12.5 m, and it is the 83rd field in this configuration.

4.2.5. Discussion of the feature matching testbed

In our previous study (Xavier et al. 2016a) we indicated that the matching research community should create a benchmark dataset for testing new matching methods or measures for geospatial vector data within a homogeneous framework. This matching testbed would undoubtedly be a useful tool for comparing different measures and methods, because we identified that the results may change outside the initial test site. In this context, the main aim of this testbed is to provide a comparing framework for geospatial data matching approaches, which can be useful for the research community or GIS software developers.

As of the source code developed in this research project was available in the internet as a free software, the same philosophy was applied to the geospatial data, since the matching testbed was made accessible through a public repository in the internet (Xavier et al. 2017). The provided matching testbed intend to be as complete as possible by providing datasets in the three geometric primitives: point, line, and area; and trying to contains all corresponding cases (1:1, 1:n, and m:n), despite of there is none m:n pair in point data.

Regarding the morphology modified datasets, there are few objects in some morphology classes (e.g., class 5 lines are less than 1% of total). This limitation cause a reduced number of cases – including in the intentionally modified lines, what sometimes caused that some areas had insufficient objects in a given class to perform a complete analysis.

Systematic disturbances try to emulate possible distortions in data that can be caused by several external factors, within some limits: the standard displacements. The random disturbances follows this same line: it emulates possible distortions in data, but in this case originated from uncontrolled random factors.

The value of the matching testbed can be summarized in four items: (1) these datasets can be used as benchmark data for other studies investigating geospatial data matching at the feature level; (2) the development of new similarity measures can benefit from these datasets as comparing sets used to calculate the new 'distances' between objects; (3) data quality studies focused on positional quality or completeness can use the datasets in order to develop new quality evaluation procedures by adopting two corresponding datasets: one as the test data, and the other as the reference; and (4) there are disturbed data that may permit assessing the robustness of investigated matching techniques in the presence of controlled perturbations.

4.3. Experimental design for feature matching

The designed experiment for the matching of geospatial features uses the concepts described in Section 3.1.3 over the matching testbed described in the Section 4.2. The following subsections detail the experiments executed in order to assess the influence of several factors over four controlled variables: (1) precision, (2) recall, (3) F-measure, and (4) time. The main objective of this experiment is to identify which matching methods are suitable to the quality control service, one of the aims of this study.

The main tool used in this experimental phase is the ANOVA. For each essay we present one table which summarizes the ANOVA results for this essay by presenting the variables and factors in the analysis. Table 4.7 illustrates an example for a single factor analysis, and Table 4.8 presents an example for multi-factor analysis.

Table 4.7. Sample table for single-factor ANOVA results.

Variable	**ANOVA**	**Best mean / treatment**	**Equivalent to**
precision	✗	99% / level 2	Level 3-5
recall	✓	97%	Treatment does not matter
F-measure	✗	98% / level1	Level 4
time	✓	20 ms	Treatment does not matter

Table 4.7 is a sample table to inform single-factor ANOVA results. The first column, *Variable*, indicates which variables were considered in the essay. Some essays does not bring the *time* variable. The *ANOVA* column indicates whether the ANOVA's F test was rejected for the considered variables or not. We adopts two symbols: ✓ for accepted, and ✗ for rejected. When the F test is accepted (✓), it means that the null hypothesis was accepted: there is no effect of the considered treatments over the controlled variable. In the other hand, when the ANOVA is rejected (✗) this means that at least one treatment affects the controlled variable. The column *Best mean / treatment* indicates the best results found for this variable in the essay. Best here refers to higher quality values (precision/recall/F-measure) and lower computational cost (time). When the ANOVA F test is rejected, and the treatment matters, we should find the equivalences using a multiple comparisons method. In this study we adopts the Tukey HSD for this purpose with a 5% of significance level. The column *Equivalent to* indicates which other treatments are equivalent to the 'best' treatment (column *Best mean / treatment*).

Table 4.8. Sample table for multi-factor ANOVA results.

	ANOVA			**Best mean / treatment**	
Variable	**F1**	**F2**	**F1:F2**	**F1**	**F2**
precision	✗	✓	✓	97% / Level 1	Level does not matter
recall	✗	✓	✗	98% / Level 2	Level does not matter
F-measure	✗	✓	✓	98% / Level 1	Level does not matter

Notes: briefly described the factors F1 and F2.

Table 4.8 works similar to Table 4.7, but for multi-factor ANOVA results. The main difference is that in multi-factor we assess more than one factor, and the interaction between the factors (e.g. column *F1:F2*). The columns *Variable*, *ANOVA*, and *Best mean/treatment* have the same meaning, but the last two are divided in factors.

This experiment is divided according to the geometric primitive: point, line, and area, in this order. Each type of geometry has its own essays, or configurations. The designed experiment for feature matching is composed by 20 essays: points (P1-P5), lines (L1-L6), and areas (A1-A9).

4.3.1. Point matching

The experiment regarding point features has five configurations, or essays. The two first configurations aims to check the performance of different matching methods using two measures: Euclidean distance (configuration P1) and context distance (P2). The third configuration (P3) assesses the influence of context over matching performance. The last two configurations evaluate the robustness of matching procedures against perturbed versions of original datasets: systematic disturbance (P4) and random disturbance (P5).

The next subsections details the corresponding essays.

4.3.1.1. Matching methods for point features

In the configuration P1 (Euclidean distance) we have 14 treatments: seven thresholds (5-35 m) by two selection criteria (both nearest and closer). The first threshold (5 m) is below the maximum positional error considered, and 35 m is the maximum positional error expected for the considered scales (DCT 2016a). We combined threshold and selection criteria into a single factor (method). The results for the four variables are shown in Table 4.9.

Table 4.9. ANOVA results for the configuration P1 (Euclidean distance).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	99.9% / both 5 m	Both 10-35 m; closer 5-10 m
recall	✗	100% / closer 20-35 m	Both 10-35 m; closer 10-15 m
F-measure	✗	98.3% / both 10 m	Both 15-35 m; closer 10-15 m
time	✓	15.44 ms	Treatment does not matter

Table 4.9 presents that the ANOVA's F test was rejected for the matching quality variables, which means that the null hypothesis for the equality of means for these variables was reject, so we can conclude that the treatment matters for these variables. If there are differences among treatments, we can looking for the 'best' treatment and then its equivalences using the Tukey HSD method, those are represented in the *Equivalent to* column. By other hand, the time variable was not influenced by the selected treatments.

If we split the combined factor (method) in two factors (threshold and criteria), it is possible to note that the results for F-measure using the both nearest criterion are less sensitive to the threshold variation than the closer criterion. Figure 4.5 illustrates the influence of thresholds over the F-measure for these two different criteria.

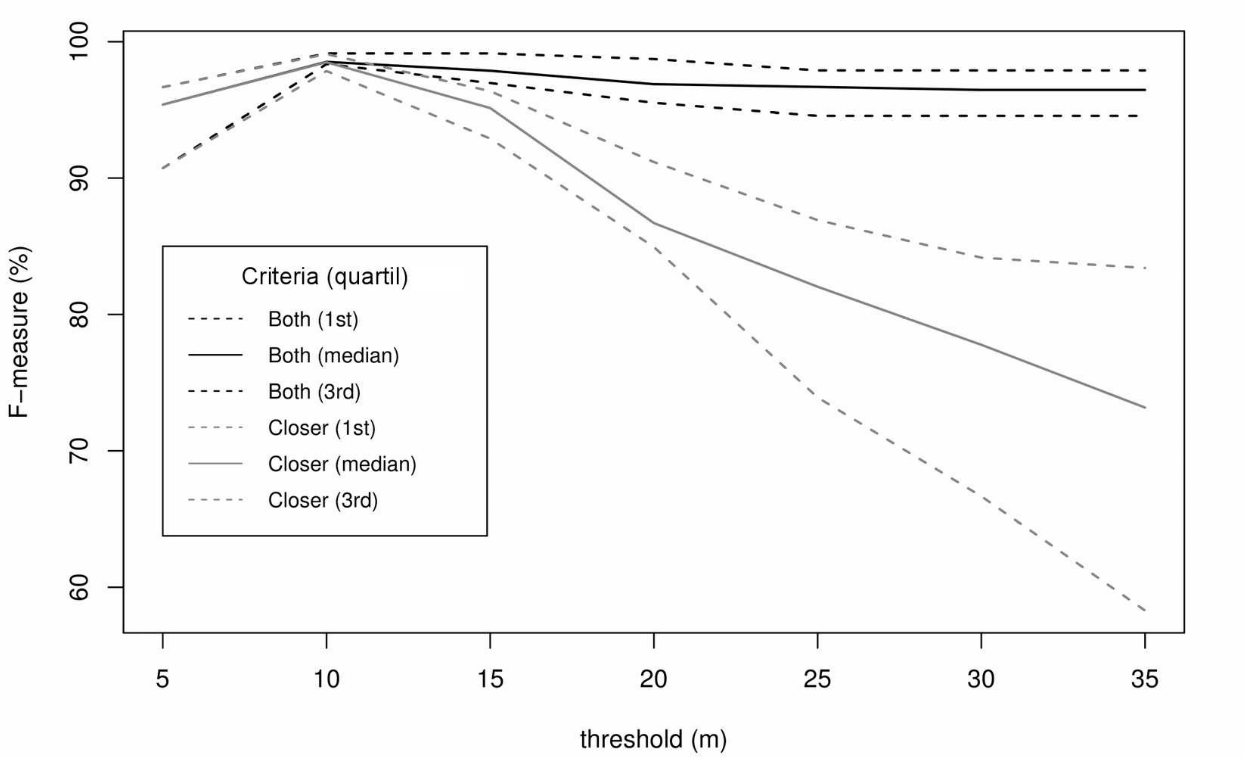


Figure 4.5. F-measure in function of threshold and selection criteria for the essay P1.

The next configuration (P2) investigates the context distance as similarity measure for point features. We combined a set of thresholds (0.3, 0.5, 0.9) and the two criteria: both nearest and closer. The both nearest criterion with the largest threshold (0.9) presented the best results, with a F-measure around 56.9%. However, this solution presented a high computational cost with a mean running time of 1134 ms, which is 70 times greater than the mean time when using Euclidean distance (see Table 4.9). The closer criteria presented worst F-measure than the both nearest criteria in all thresholds. Table 4.10 summarizes the results of ANOVA for this configuration.

Table 4.10. ANOVA results for the configuration P2 (geographic context).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	62.6% / both 0.9	Both 0.3-0.5
recall	✗	95.3% / closer 0.9	Closer 0.5
F-measure	✗	56.9% / both 0.9	Both 0.3-0.5
time	✓	1130 ms	Treatment does not matter

From these two first essays (P1 and P2) we selected three matching methods that will be used in the following configurations: (1) Euclidean distance, closer criterion, 10 m threshold (10C); Euclidean distance, both nearest, 25 m (25B); and geographic context distance, both nearest, 0.9 (CTT). These methods will be referenced as factor F1 and their corresponding acronyms (10C, 25B, and CTT).

4.3.1.2. Geographic context influence

In the configuration P3 we evaluate whether the context around a point feature influences the matching performance. We have nine treatments for two factors: three matching methods (F1) versus three geographic context classes (F2) (see Section 3.1.3.4). The geographic context factor has three levels: (1) uncertain, (2) intermediary, and (3) distinct. Considering that the variable time could not be assessed because the procedure run for the entire dataset (all classes at time), the results for the remained three variables are shown in Table 4.11.

Table 4.11. ANOVA results for the configuration P3 (context class).

	ANOVA			Best mean / treatment	
Variable	F1	F2	F1:F2	F1	F2
precision	✗	✓	✓	97.7% / 10C	Level does not matter
recall	✗	✓	✓	98.6% / 10C	Level does not matter
F-measure	✗	✓	✓	98.1% / 10C	Level does not matter

Notes: factor F1 is the matching methods, factor F2 is geographic context.
10C means Euclidean distance, closer criterion, 10 m threshold (Section 4.3.1.1)

Figure 4.12 presents that the ANOVA's F test was rejected in the factor F1 (matching method), which means that the null hypothesis for the equality of means was rejected, so we can conclude that the treatment matters for these variables. However, when considering the second factor F2 (geographic context), there is no evidence to discard the null hypothesis, so this factor does not influence the controlled variables. In the same hand the interaction between factors (F1:F2) does not matter for these variables.

The level that had suffered most influence from context classes was the geographic context measure (CTT). In Figure 4.6 it is possible to note that context class 1 (uncertain context) presented an average F-measure less that other classes for the method that used the geographic context measure. This occurred because in the uncertain context the geographic context measure compares similar contexts, which can lead to a mismatch.

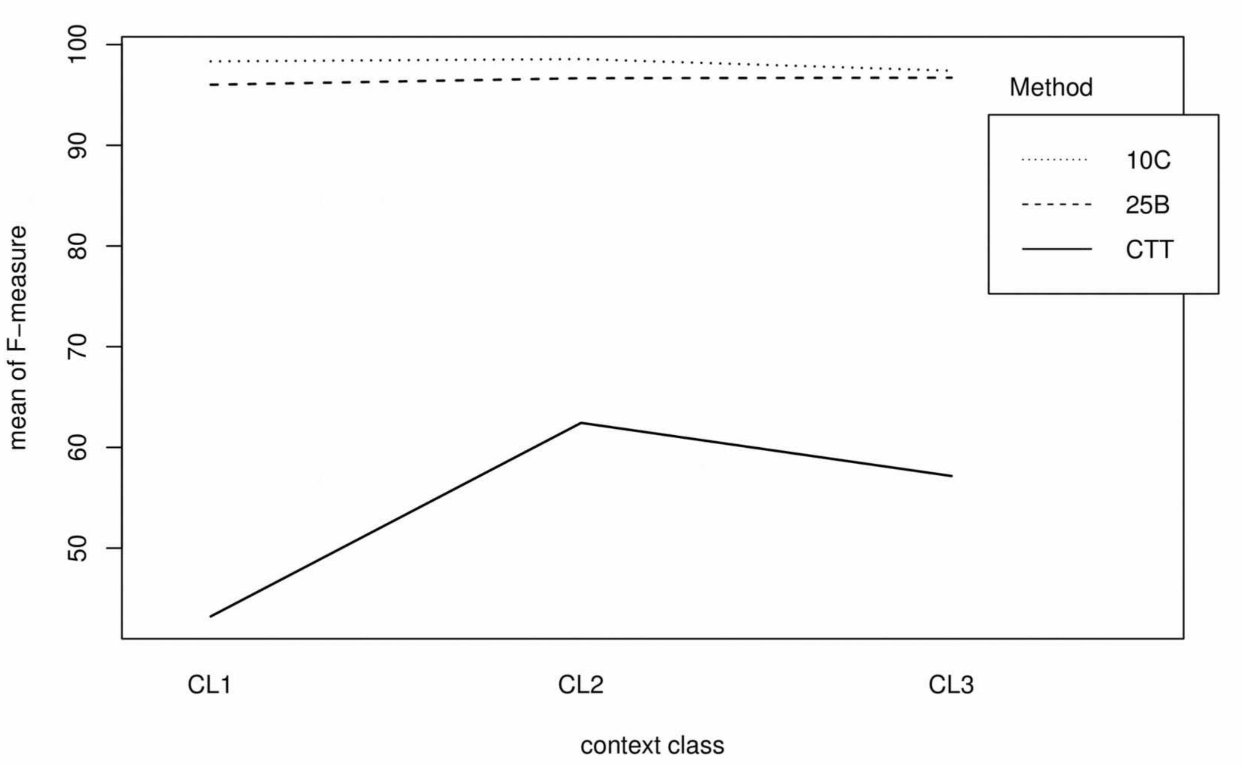


Figure 4.6. Average F-measure in function of context class and matching method in essay P3.

4.3.1.3. Disturbance influence

The essays for assess the influence of disturbance over data does not consider the time variable. The next essay refers to systematic perturbations applied over point data (configuration P4). In this configuration we have the matching methods as blocking factor and each type of systematic disturbance is considered one factor: translation (F3), rotation (F4), and scaling (F5) (see Table 4.12).

Table 4.12. ANOVA results for the configuration P4 (systematic disturbance).

			ANOVA			**Average value (%) relative to no changing**	
Variable	**Method**	**F3**	**F4**	**F5**	**F3**	**F4**	**F5**
precision	10C	✗	✗	✗	35	78	73
	25B	✗	✗	✗	54	84	78
	CTT	✓	✓	✗	100	100	100
recall	10C	✗	✗	✗	27	59	54
	25B	✗	✗	✗	50	78	72
	CTT	✓	✓	✗	100	100	100
F-measure	10C	✗	✗	✗	28	64	58
	25B	✗	✗	✗	51	80	74
	CTT	✓	✓	✗	100	100	100

Notes: matching methods are described in Section 4.3.1.1

From Table 4.12 we can observe that systematic disturbance have affected the controlled variables for each matching method with different acuteness. The *Average value (%) relative to no changing* column shows the average value of each variable in relation to the 'no disturb' configuration. Translations and rotations influenced the results obtained from matching methods based on Euclidean distance (10C and 25B). However these disturbs had no detectable effect over the variables for the matching method based on geographic context distance. Only those perturbations of type scaling (F5) influenced the quality variables for all matching methods.

Figure 4.7 shows how the different translations t0-t32 (see Section 4.2.3) influences the average F-measure according to the matching method. Despite of the method 25B (Euclidean distance, both nearest, 25 m) has obtained the greater average value in the first translations, their results are getting worse when the translation amplitude raises. By other hand, the matching method based on a geographic context distance (CTT) maintained an almost constant performance for F-measure along all assessed translations, which help to corroborate the ANOVA result presented in Table 4.12.

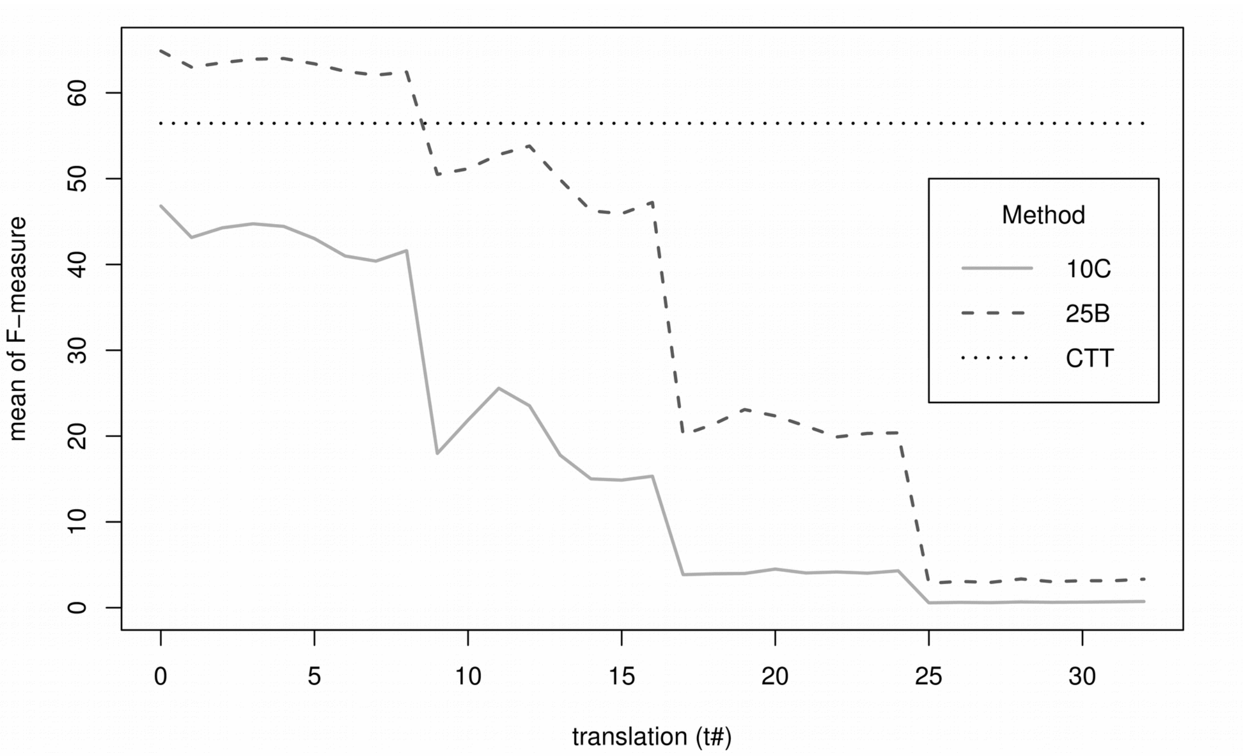


Figure 4.7. Average F-measure for each matching method in function of translation configuration (t0-t32).

The last essay (P5) for point data refers to random perturbations. In this configuration we have four treatments: the four standard displacements (5, 12.5, 25, and 50 m) that generated the vector fields to perturb original data (100 vector fields for each one, see Section 4.2.4). The results of ANOVA are presented in Table 4.13. The column *Average* *value by displacement (%)* shows the mean value for each variable in each standard displacement value.

Table 4.13. ANOVA results for the configuration P5 (random disturbance).

			Average value by displacement (%)			
Variable	**Method**	**ANOVA**	5 m	12.5 m	25 m	50 m
precision	10C	✗	97.23	92.46	65.24	32.31
	25B	✗	95.66	92.43	76.42	45.48
	CTT	✓	62.59	62.58	62.44	61.37
recall	10C	✗	97.63	80.78	33.15	8.98
	25B	✗	97.54	93.41	74.74	34.33
	CTT	✓	52.78	52.59	52.07	50.63
F-measure	10C	✗	97.39	85.43	42.14	13.21
	25B	✗	96.54	92.86	75.47	38.92
	CTT	✓	56.51	56.43	56.07	54.78

Notes: matching methods are described in Section 4.3.1.1

From Table 4.13, ANOVA shows that random disturbance factor influenced the performance Euclidean-based methods, but it did not influenced the context-based method. Once more, the geographic context measure exhibited that despite does not present the best performance, this measure is less sensitive to disturbs than those methods based on the Euclidean distance. Figure 4.8 illustrates this results by presenting the variations in the average F-measure relative to the random disturb for each standard displacement applied.

4.3.2. Line matching

For line matching we prepared a designed experiment with six essays. The first three essays (L1-L3) are focused on try different matching methods by combining similarity measures and thresholds. All essays for linear data uses the selection criteria closer that supports m:n relations, since this kind of corresponding case is very common in the considered datasets (see Table 4.1). The fourth configuration (L4) investigates the influence of line morphology over the performance of matching methods. Finally, the last two essays assess the robustness of linear matching procedures against perturbed versions of original datasets: systematic disturbance (L5) and random disturbance (L6).

The next subsections details the corresponding essays.

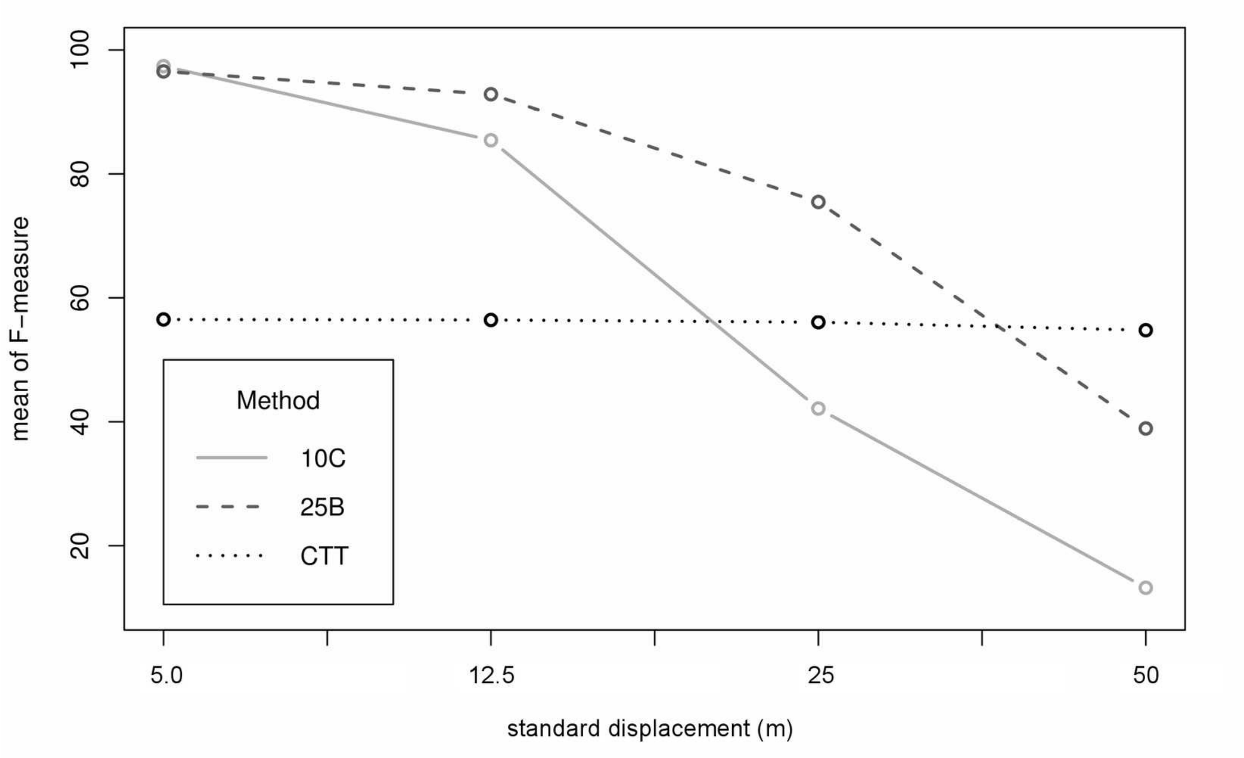


Figure 4.8. Average F-measure for each matching method in function of standard displacement for random disturbs in essay P5.

4.3.2.1. Matching methods for line features

The first essay for line matching (configuration L1) intends to check the similarity measures that performs better using a set of thresholds. There are four measures:

- Hausdorff distance (HD);
- Discrete Fréchet distance (DFre);
- Partial discrete Fréchet distance – lines densified using threshold (PFre); and
- Short-line median Hausdorff distance (SMHD).

The seven thresholds are the same used for the configuration P1 (5 – 35 m). The results for the essay L1 are presented in Table 4.14.

Table 4.14. ANOVA results for the configuration L1 (distances x thresholds).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	100% / HD 5 m	HD 10-25 m; DFre 10-25; PFre 5-10 m; SMHD 5m
recall	✗	86.31% / SMHD 35m	PFre 20-35 m; SMHD 5-35 m
F-measure	✗	84.13% / SMHD 10 m	PFre 15 m; SMHD 5-15 m
time	✗	112 ms / DFre 5 m	HD, DFre, SMHD: all

From Table 4.14 we can notice that ANOVA was rejected for all variables, which would be expected since we have 28 treatments. However it is possible to identify which treatments achieves the better results, and then apply the Tukey HSD in order to find which treatments are equivalent for those better cases. Regarding precision, the HD using a 5 m threshold reached the 100% average precision, but in a cost of an average recall of only 5%. This means that the measure performs well, but in a limited range of objects. The better recall rates were reached for those methods with the largest threshold (35 m). In this case the SMHD measure stands out from the other by reaching a high recall rate including using a small threshold (5 m). The F-measure is a variable that shows a balanced result between precision and recall. In this case, those measures that uses partial values (SMHD and partial discrete Fréchet) obtained the best results. Regarding time (computational cost), those methods that required a previous densification (PFre) achieved the worst results because they have to perform the densify operation prior to the distance computation itself.

Analysing Figure 4.9 it is possible to observe that the measures SMHD and PFre performed betted than other measures for all thresholds. Taking into account all results in this essay (L1) we selected two treatments for further investigation: SMHD 10 m and PFre 15 m.

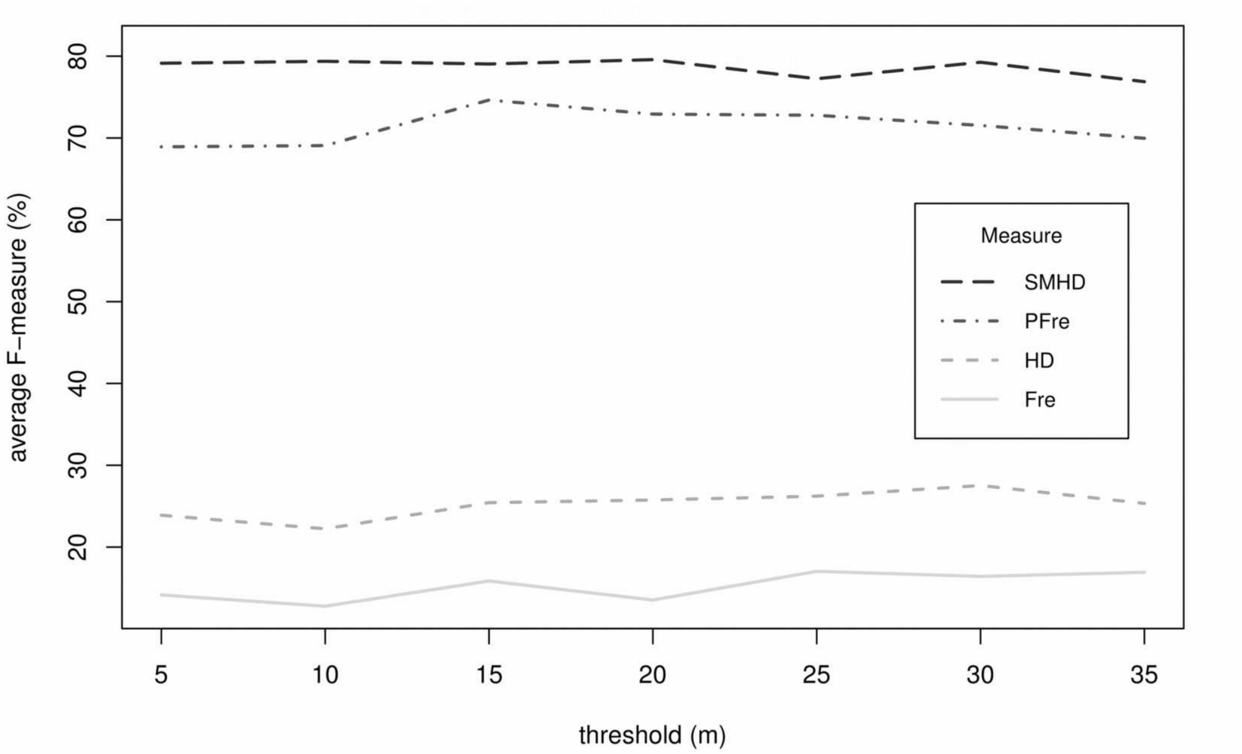


Figure 4.9. Average F-measure for each line measure in function of threshold in essay L1.

The literature review (Section 2.3) indicated that the combination of measures is a common practice in matching methods. So, in the second essay (L2) we investigate the combination of two distance measures (SHMD and PFre) with the partial orientation distance (Section 3.1.1.2). The orientation was used as an exclusion criterion in the study of Kang et al. (2015). We have 12 treatments in this essay: two methods (SMHD 10 m and PFre 15 m) over six thresholds for partial orientation (0.1, 0.2, 0.3, 0.4, 0.5, 4 rad). Effectively, the last orientation threshold (4 rad) means no threshold. Table 4.15 presents th results for the essay L2.

Table 4.15. ANOVA results for the configuration L2 (partial orientation).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	96.05% / SMHD 0.1 rad	SMHD 0.2-0.4 rad; PFre 0.1 rad
recall	✗	85.30% / SMHD 0.5 rad	SMHD 0.1-4 rad
F-measure	✗	88.96% / SMHD 0.4 rad	SMHD 0.1-4 rad
time	✗	132 ms / SMHD 0.1 rad	SMHD 0.1-0.5 rad

Notes: SMHD means a matching method using the SMHD distance and 10 m threshold. PFre means a matching method using the partial discrete of Fréchet distance and 15 m threshold. Radians values are related to the partial orientation distance.

The ANOVA tests presented in Table 4.15 indicate that using the new measure (partial orientation) as an excluding criterion influenced all controlled variables, since precision to time. However, the p-value analysis indicate that orientation influenced more precision and time (p-value $\ll$ 0.01) than recall and F-measure (p-value in [0.011, 0.036]). SMHD measure presented better results when compared with PFre measure as of in the essay L1. Taking into consideration all results from essay L2, we note that applying a partial orientation of 0.4 rad increased the average F-measure for SMHD and PFre from 84.1% to 88.9%, and from 77.9% to 82.2%, respectively. Therefore, we included two methods for further investigation: SMHD 10 m, partial orientation 0.4 rad, and PFre 15 m, partial orientation 0.4 rad.

The last set of measures considered in this experiment refers to buffer-based measures: the single buffer method (SBM) and the double buffer method (DBM). In this essay we adopt one factor – the matching method, that is the combination of measure, buffer length, and similarity threshold. We have 50 treatments: two measures (SBM and DBM) by five buffer lengths (10, 20, 30, 40, 50 m) by five thresholds (0.5, 0.6, 0.7, 0.8, 0.9). With many treatments, the ANOVA test was rejected for all variables (see Table 4.16).

Table 4.16. ANOVA results for the configuration L3 (buffer methods).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	100% / DBM 10 m 0.7-0.8; DBM 20-40 m 0.9	None
recall	✗	50.02% / SBM 40 m, 0.5	SBM 10-50 m 0.5
F-measure	✗	58.91% / SBM 10 m 0.5	SBM 20 m 0.5, 10 m 0.6-0.7; DBM 50 m 0.6, 20-50 m 0.5
time	✗	1268 ms / SBM 50 m 0.7	All SBM

By comparing these results (Table 4.16) with linear methods (Table 4.14) it is possible to note that buffer-based measures have high computational cost (~ 10x more time) for a worst quality performance (best F-measure decays from 84% to 59%).

After analysing the first three essays (L1-L3) we selected five linear matching methods in order to investigate further. (1) SMHD 10 m, (2) PFre 15 m, (3) SMHD 10 m, partial orientation 0.4 rad (SMHDo), and PFre 15 m, partial orientation 0.4 rad (SMHDo), and (5) SBM, buffer length 10 m, threshold 0.5 (SBM). These methods will be referenced as factor F1 in the following essays.

4.3.2.2. Line morphology influence

In the configuration L4 we evaluate whether the line morphology influences the performance of the selected matching methods. This is a two-factor essay where the first factor (F1) uses the matching methods, and the second factor (F2) represents the five morphology classes for lines (see Section 4.2.2). In this experiment the time could not be assessed. The results for the other three controlled variables are presented in Table 4.17.

Table 4.17. ANOVA results for the configuration L4 (line morphology).

	ANOVA			Best mean / treatment	
Variable	**F1**	**F2**	**F1:F2**	**F1**	**F2**
precision	✗	✓	✓	92.1% / SMHDo	Level does not matter
recall	✗	✗	✗	77.6% / SMHD	63.3% / CL2
F-measure	✗	✗	✗	83.2% / SMHDo	70.9% / CL2

Notes: factor F1 is the matching methods, factor F2 is line morphology.
CL2 means class 2 lines - smooth (see Section 4.2.2).

The ANOVA for the essay L4 indicates that the matching method has effect of all variables, as shown in the last essays (L1-L3). However, the morphology class of assessed lines did not influence the precision variable, as was detected in other variables (recall and F-measure). Regarding the F-measure, the class 2 lines (smooth) reached an average value compared to the class 5 lines (very sinuous) (69%). The analysis of Tukey HSD for F-measure indicated that the results for classes CL2-CL5 are equivalent.

4.3.2.3. Disturbance influence

The essays for assess the influence of disturbance over data does not consider the time variable. The configuration L5 refers to systematic perturbations applied over linear data. In this configuration the matching methods (F1) work as a blocking factor and each type of systematic disturbance is considered one factor: translation (F3), rotation (F4), and scaling (F5). The results for ANOVA are shown in Table 4.18.

Table 4.18. ANOVA results for the configuration L5 (systematic disturbance).

Variable	Method	ANOVA F3	ANOVA F4	ANOVA F5	Average value (%) relative to no changing F3	F4	F5
precision	SMHD	✗	✗	✗	68	77	74
	PFre	✗	✗	✗	70	82	77
	SMHDo	✗	✗	✗	75	84	82
	PFreo	✗	✗	✗	74	86	82
	SBM	✗	✗	✗	71	83	79
recall	SMHD	✗	✗	✗	47	62	55
	PFre	✗	✗	✗	47	62	54
	SMHDo	✗	✗	✗	46	61	54
	PFreo	✗	✗	✗	47	62	53
	SBM	✗	✗	✗	45	60	54
F-measure	SMHD	✗	✗	✗	52	66	60
	PFre	✗	✗	✗	51	67	59
	SMHDo	✗	✗	✗	52	67	61
	PFreo	✗	✗	✗	51	67	59
	SBM	✗	✗	✗	49	65	58

Notes: matching methods are described in Section 4.3.2.1.

The ANOVA results in Table 4.18 indicates that all types of systematic disturbance (translation, rotation, and scaling) influenced the quality variables (precision, recall, and F-measure). All considered matching methods were influenced by systematic disturbance. The variable most influenced was recall. Figure 4.10 shows as the average precision and the average recall suffered the effects of translation disturbs (t0-t32). Average recall values are always below the average precision.

The last essay (L6) for line matching refers to random perturbations. In this configuration we have one factor (random disturbance) with four levels: vector fields generated by standard displacements: 5, 12.5, 25 and 50 m (100 vector fields for each one, see Section 4.2.4). Time is not considered in this essay. The results of ANOVA are presented in Table 4.19.

Considering the random disturbance, the results of essay L6 are similar to those equivalent for point data (essay P5, Section 4.3.1.3), which a random disturbance with less amplitude (5 m) reached the best quality results (precision/recall/F-measure). All variables in all matching methods suffered the influence of random disturbance, highlighting the recall. Figure 4.11 illustrates as the standard displacement of the vector field increases, the average F-measure falls.

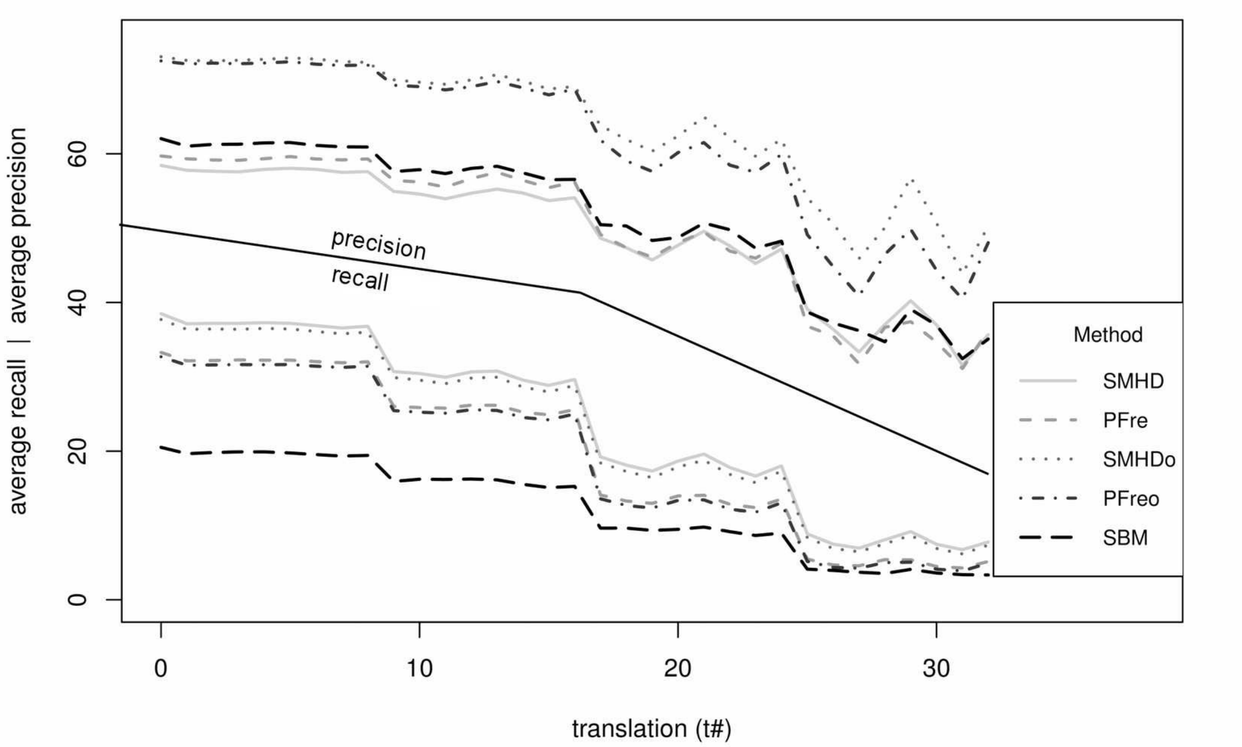


Figure 4.10. Average recall and average precision in for each line matching method in function of translation configuration (t0-t32) in the essay L5.

Table 4.19. ANOVA results for the configuration L6 (random disturbance).

			Average value by displacement (%)			
Variable	**Method**	**ANOVA**	5 m	12.5 m	25 m	50 m
precision	SMHD	✗	81.95	75.86	66.66	55.77
	PFre	✗	81.45	77.77	70.05	57.91
	SMHDo	✗	93.24	89.45	82.09	71.43
	PFreo	✗	92.03	89.99	83.54	71.28
	SBM	✗	83.14	80.27	71.27	57.87
recall	SMHD	✗	84.57	78.97	53.85	26.95
	PFre	✗	74.92	71.39	48.66	21.05
	SMHDo	✗	84.22	78.70	52.99	25.92
	PFreo	✗	74.47	70.96	47.91	20.32
	SBM	✗	46.82	43.61	29.10	13.96
F-measure	SMHD	✗	92.95	76.99	58.90	35.52
	PFre	✗	77.31	73.80	56.62	30.03
	SMHDo	✗	88.18	83.37	63.76	37.35
	PFreo	✗	81.80	78.92	60.21	30.89
	SBM	✗	58.95	55.51	40.34	21.75

Notes: matching methods are described in Section 4.3.2.1.

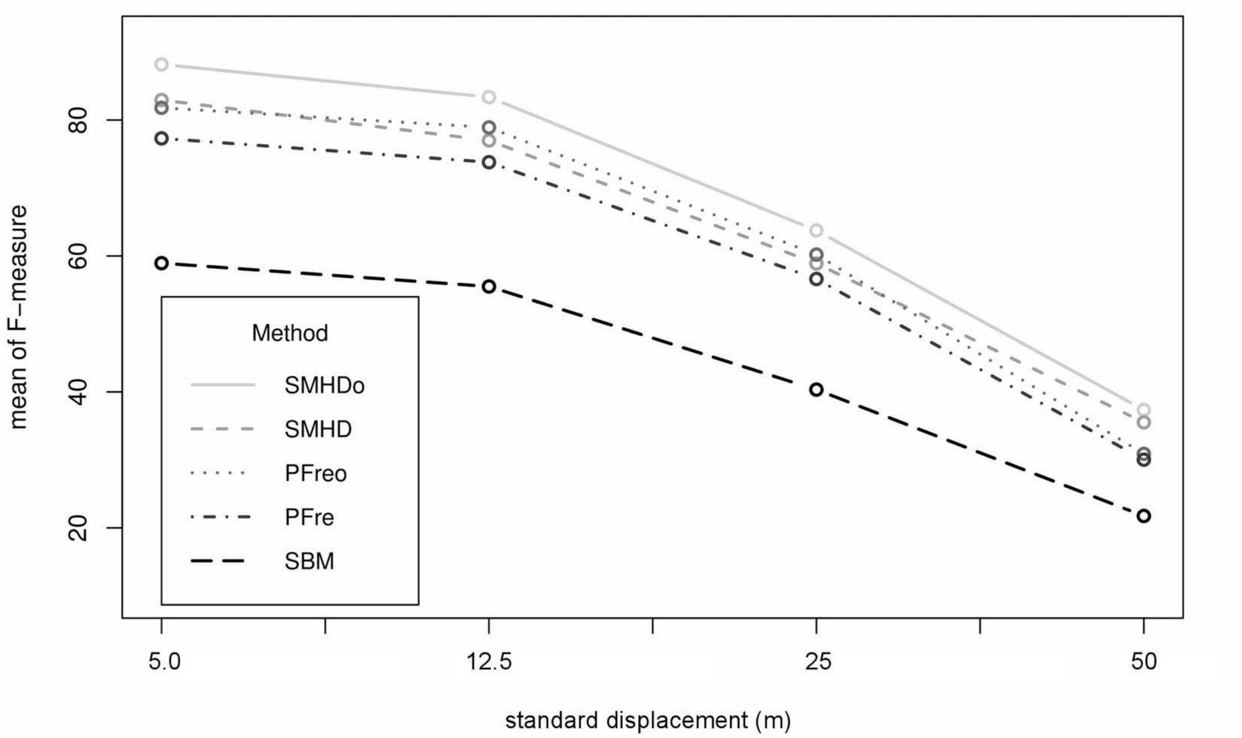


Figure 4.11. Average F-measure for each line matching method in function of standard displacement for random disturbance (essay L6).

4.3.3. Area matching

The designed experiment for area matching has nine essays. The first five essays (A1-A5) investigate the capacity of different similarity measures in matching methods and a corresponding set of thresholds. The configuration A6 finds out the influence of area morphology over the performance of matching methods. Then configuration A7 intend to check if there is a significant influence of geographic context over the controlled variables. In the end, the last two essays rates the robustness of area matching procedures against perturbed versions of original datasets: systematic disturbs (A8) and random disturbs (A9).

The next subsections details the corresponding essays.

4.3.3.1. Matching methods for area features

In the first essay towards area matching methods (A1) we investigate the performance of overlap area measure using different selection criteria and thresholds. This is a single factor essay (method) with 12 treatments: two selection criteria (both nearest and closer) and six area thresholds (5, 10, 20, 30, 40, and 50 m²). The results for the four variables are shown in Table 4.20.

Table 4.20. ANOVA results for the configuration A1 (overlap area).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	99.91% / both 20 m²	Both 5-50 m²; closer 20-50 m²
recall	✗	98.49% / closer 5 m²	Closer 10-20 m²
F-measure	✗	98.44% / closer 5 m²	Closer 10-20 m²
time	✓	93.44 ms	Treatment does not matter

From Table 4.20, the ANOVA test was rejected for the matching quality variables (precision/recall/F-measure), which means that the null hypothesis for the equality of means was reject, so we can conclude that the treatment matters for these variables. If there are differences among treatments, we can find the its equivalences using the Tukey HSD method. By other hand, the time variable was not influenced by the selected treatments.

In the essay A2 we investigate different configurations for the overlap rate measure. We have a single factor (matching method) that combines two variables: two selection criteria (both nearest and closer) and five similarity thresholds (0.1, 0.2, 0.3, 0.4, and 0.5). This measure works similarly to the overlap area (essay A1), but this one quantifies the overlap rate between two areal objects (Section 3.1.3.1). Table 4.21 drawn the ANOVA results for this configuration.

Table 4.21. ANOVA results for the configuration A2 (overlap rate).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	99.95% / both 0.3	Both 0.1-0.5; closer 0.3-0.5
recall	✗	98.34% / closer 0.1	Closer 0.2-0.4
F-measure	✗	98.70% / closer 0.2	Closer 0.3-0.4
time	✓	91.78 ms	Treatment does not matter

Once more the null hypothesis was rejected for the quality variables, which indicates the relevance of this factor over these variables, what could not be observed for the computational cost (variable time). Comparing the results of essays A1 and A2 for overlap area and overlap rate measure, it is possible to note that the closer criterion performed better than the both nearest because there is many-to-many pairs which could be resolved by this criterion. Figure 4.12 illustrates the box-plot graph for the F-measure in the essay A2. In this figure it is possible to note that the closer criterion performs better than the both nearest criterion for the same thresholds, since that is able to deal with m:n pairs.

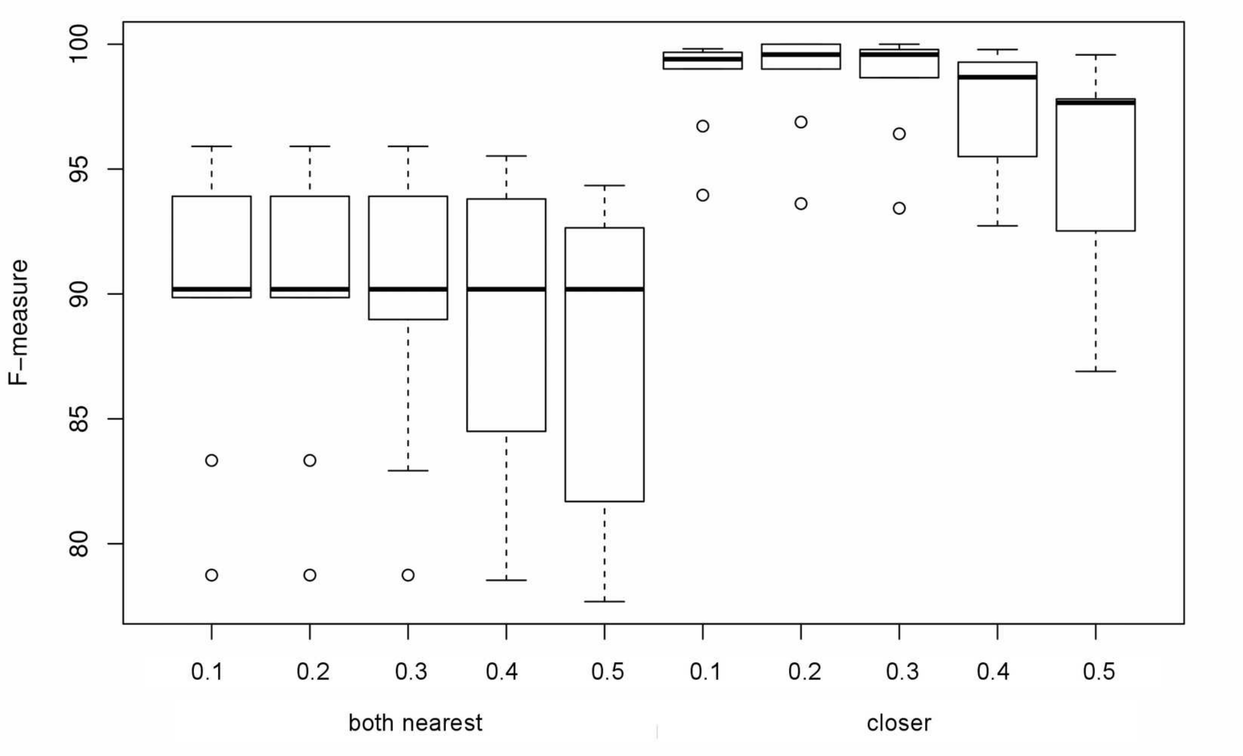


Figure 4.12. Box-plot F-measure in function of two selection criteria (both nearest and closer) by five thresholds (0.1-0.5) for the overlap rate measure (essay A2).

The third matching method tested was developed by Ruiz-Lendínez (2012) in a study focused on quality assessment. The author combined a set of area measures using a genetic algorithm (GA) in order to evaluate the similarity between building objects. In this essay we investigate the GA's method by testing two selection criteria (both nearest and closer) and five similarity thresholds (0.5, 0.6, 0.7, 0.8, and 0.9). The results for the essay A3 are presented in Table 4.22.

Table 4.22. ANOVA results for the configuration A3 (GA).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	100% / closer 0.9	Both 0.5-0.9, closer 0.7-0.8
recall	✗	97.22% / closer 0.5	Closer 0.6
F-measure	✗	93.01% / closer 0.6	Both 0.5-0.6; Closer 0.5
time	✓	48.67 ms	Treatment does not matter

The original study of Ruiz-Lendínez (2012) adopted a selection criteria similar (but not equal) to the both nearest criterion. In that study the author used a threshold equal to 0.8 in order to reach a good precision. In fact, in this essay the 0.8 threshold reached an average precision of 99.8%, and the Tukey HSD indicated this value is equivalent to a 100% precision, which corroborates that decision. However, when using the closer criterion it is possible to increase the global performance (F-measure) because it raises the recall rates.

The next configuration (A4) assess the geographic context distance as similarity measure for area features. As the measure considers points, we use the centroid of areas as points. In this essay we have six treatments: two selection criteria (both nearest and closer) by three thresholds (0.3, 0.5, and 0.9), the same treatments for the equivalent essay for point features (essay P2, Section 4.3.1.1). Table 4.23 presents the results for the configuration A4.

Table 4.23. ANOVA results for the configuration A4 (context).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	54.61% / both 0.9	Both 0.3-0.5
recall	✗	84.32% / closer 0.9	Closer 0.5
F-measure	✗	39.67% / both 0.9	Both 0.3-0.5
time	✓	1891 ms	Treatment does not matter

From Table 4.23 is possible to note that the treatment matters for the quality variables, but the time is not affected. If we compare the best performance values for F-measure and time with the other essays (A1-A3), the geographic context measure performs worst in quality (less than 50%) and computational cost (20x more time).

In the last essay for area matching methods we examine the performance of line distance methods as similarity measures for areal features. Considering the performance in the line matching experiment (Section 4.3.2.1), we selected two line measures: short-line median Hausdorff distance (SMHD) and partial discrete Fréchet distance (PFre). We have 10 treatments: two measures (closer criterion) by five thresholds (10, 20, 30, 40, and 50 m). These measures were applied in the exterior ring of assessed polygonal objects. The results are summarized in Table 4.24.

Table 4.24. ANOVA results for the configuration A5 (line measures).

Variable	ANOVA	Best mean / treatment	Equivalent to
precision	✗	89.15% / PFre 10 m	None
recall	✗	98,63% / SMHD 50 m	SMHD 10-40 m
F-measure	✗	89.18% / SMHD 10 m	None
time	✗	51.9 ms / SMHD 10 m	SMHD 20-50 m

ANOVA tests revealed that the treatment matters for all controlled variables. Despite of PFre 10 m had reached the best precision, its recall is below 30%. Regarding recall rates, SMHD configurations had the best results, the same for time consumption.

Taking into consideration the results of these five essays (A1-A5), we selected five matching methods for further investigation in other essays:

- Overlap area, closer criterion, 10 m² threshold;
- Overlap rate, closer criterion, 0.3 (30%) threshold;
- Set of measures by genetic algorithm (GA), closer criterion, 0.6 threshold;

- Geographic context measure, both nearest criterion, 0.9 threshold; and
- Line measure SMHD, closer criterion, 10 m threshold.

These methods will be referenced as factor 1 (F1) in the following essays.

4.3.3.2. Area morphology influence

The essay A6 was designed to evaluate the influence of area morphology over the selected matching methods. We have a two factors: matching methods (F1) and morphology classes (F2). There are 20 treatments: five matching methods against four morphology classes for areas (see Section 4.2.2). There are only two observations for the last morphology class (CL4), while there are nine for each other (the nine regions). This occurred because there are few class 4 buildings in the dataset. In this essay the time could not be assessed. The results for the other three controlled variables are presented in Table 4.25.

Table 4.25. ANOVA results for the configuration A6 (morphology).

	ANOVA			Best mean / treatment	
Variable	**F1**	**F2**	**F1:F2**	**F1**	**F2**
precision	✗	✓	✓	99.46% / Rate	Level does not matter
recall	✗	✓	✓	97.35% / SMHD	Level does not matter
F-measure	✗	✓	✓	97.92% / Area	Level does not matter

Notes: factor F1 is the matching methods (Section 4.3.3.1), factor F2 is the area morphology.

As verified in the other configurations (A1-A5), the matching method influences the results for all controlled variables. However, the ANOVA test indicated that there is no evidence to reject the null hypothesis considering the morphology class factor. So, we can conclude that this factor has no significant influence over the controlled variables for this set of matching methods.

4.3.3.3. Geographic context influence

Other factor that may influence the performance of matching methods is the geographic context. The configuration A7 was designed to assess whether the context around an area feature influences this performance. This essay has 15 treatments for two factors: five matching methods (F1) against three geographic context classes (F3) (see Section 3.1.3.4). Similar to the point essay (P3, Section 4.3.1.2), the geographic context factor has three levels: (1) uncertain, (2) intermediary, and (3) distinct. The results for the three quality variables are shown in Table 4.26.

Table 4.26. ANOVA results for the configuration A7 (context class).

	ANOVA			Best mean / treatment	
Variable	**F1**	**F3**	**F1:F3**	**F1**	**F3**
precision	✗	✓	✓	99.6% / Rate	Level does not matter
recall	✗	✓	✓	97.74% / SMHD	Level does not matter
F-measure	✗	✓	✓	97.69% / Rate	Level does not matter

Notes: factor F1 is the matching methods (Section 4.3.3.1), factor F3 is the geographic context class.

Similarly to what occurred when assessing this factor for point features (P3), there is no evidence to discard the null hypothesis for the factor F3 (geographic context), from which we can conclude that this factor does not influence the controlled variables. The matching method most influenced by the context class was that based on the geographic context measure. Figure 4.13 shows as the average recall rates decreases for the context method (Ctt) when the context class goes from uncertain (CL1) to distinct (CL3).

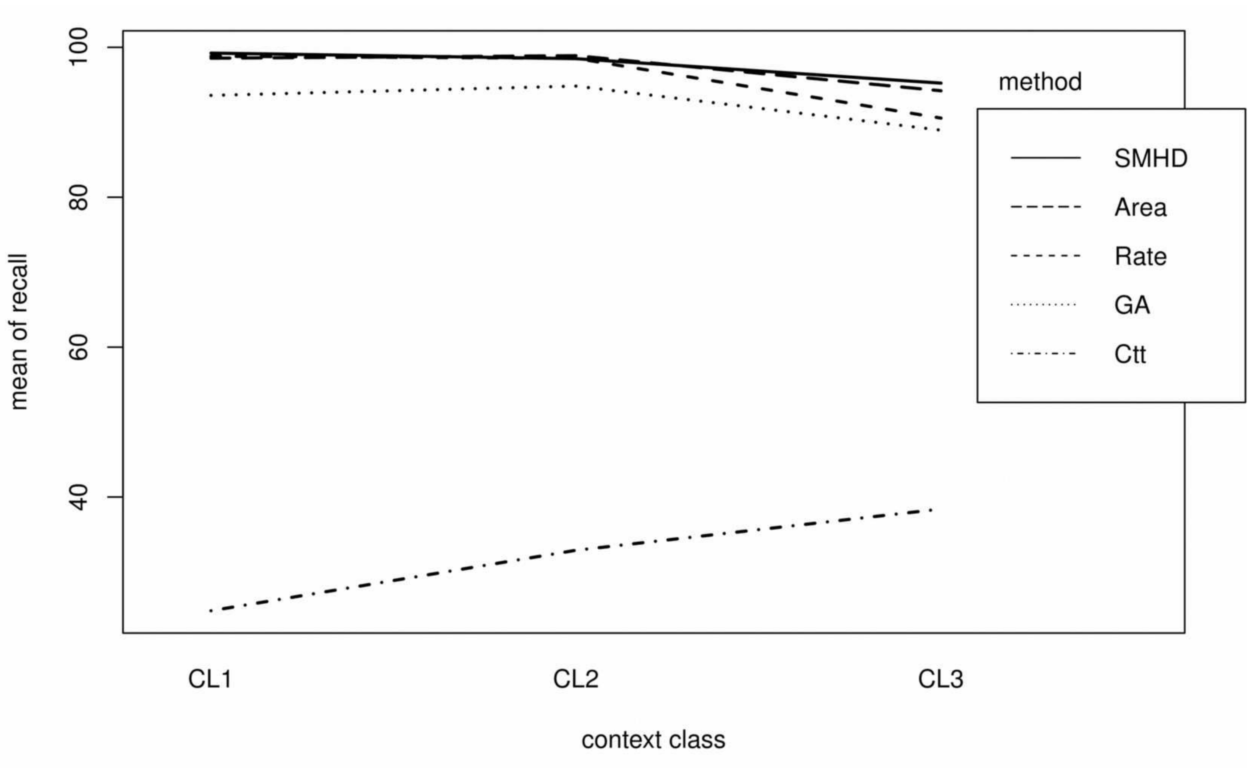


Figure 4.13. Average recall rate in function of the geographic context class and matching measures in the essay A7.

4.3.3.4. Disturbance influence

In the last two essays for area matching we assess the robustness of the selected matching measures testing their performance against synthetic perturbed versions of original data (see Section 4.3.3 and Section 4.3.4). The controlled variable time is not considered in these essays.

The essay A8 refers to systematic perturbations applied over area data. In this configuration we have the matching methods as blocking factor and each type of systematic disturb is considered one factor: translation (F4), rotation (F5), and scaling (F6). Table 4.27 presents the results for systematic perturbations over three variables.

Table 4.27. ANOVA results for the configurationA8 (systematic disturbance).

		ANOVA			Average value (%) relative to no changing		
Variable	**Method**	**F4**	**F5**	**F6**	**F4**	**F5**	**F6**
precision	Area	✗	✗	✗	55	81	79
	Rate	✗	✗	✗	55	83	81
	GA	✗	✗	✗	59	85	83
	Ctt	✓	✓	✗	100	102	102
	SMHD	✗	✗	✗	55	83	80
recall	Area	✗	✗	✗	48	74	70
	Rate	✗	✗	✗	35	65	60
	GA	✗	✗	✗	39	67	63
	Ctt	✓	✓	✗	102	103	103
	SMHD	✗	✗	✗	51	78	74
F-measure	Area	✗	✗	✗	50	76	73
	Rate	✗	✗	✗	41	71	67
	GA	✗	✗	✗	46	73	70
	Ctt	✓	✓	✗	101	103	102
	SMHD	✗	✗	✗	52	80	76

Notes: matching methods are described in Section 4.3.3.1.

The last column in Table 4.27 (*Average value (%) relative to no changing*) represents the relation between the average mean when varying just one factor (F4-F6) in relation to the no disturbed version (t0, r0, s0). From these results we can note that systematic disturbance affected the controlled variables with different amplitudes. Regarding ANOVA tests, scaling perturbations (F6) were important for all matching methods. The other perturbations, translation and rotation (F4 and F5) influenced all matching methods except that based on the geographic context measure. Figure 4.14 illustrates as the average F-measure suffers the effects of different translations. It is possible to note that the results for the geographic context method (Ctt) remains almost the same.

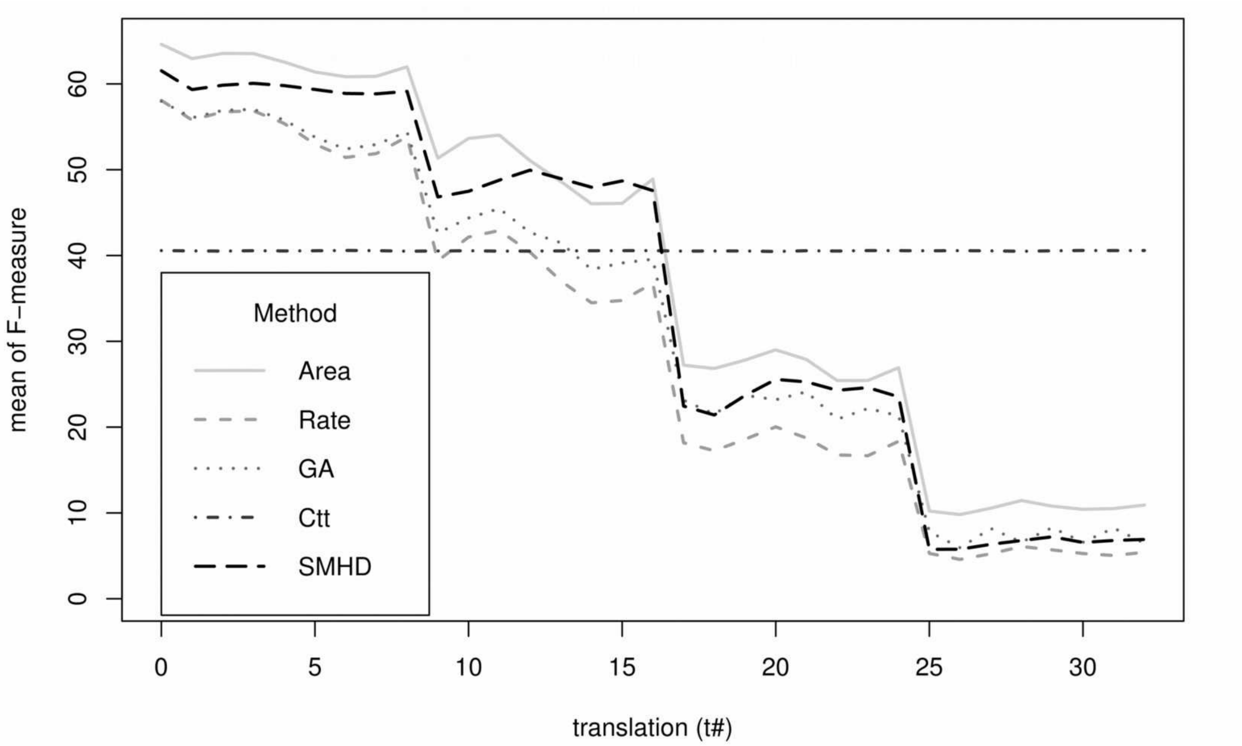


Figure 4.14. Average F-measure for each matching method in function of translation configuration (t0-t32) in the essay A8.

In the last essay (A9) we examine the influence of random disturbance over area matching methods. This one-factor essay has four treatments: the standard displacements used to generate the vector fields (see Section 4.2.4). The results were grouped by matching method in order to assess the influence of factor over each one. The results are drawn in Table 4.28.

Considering the average value of variables according to the standard displacement (column *Average value by displacement (%)*) it is possible to note that recall was more affected than precision. Similarly to that occurred with systematic displacements, the ANOVA tests revealed that the matching method based on the geographic context measure (Ctt) was not affected by the random disturbance factor. All other matching methods experienced some influence when the original data was submitted to the displacement vector field. Figure 4.15 illustrates as the context method, despite of it did not reach the best results, remained almost constant when facing the random disturbs.

Table 4.28. ANOVA results for the configuration A9 (random disturbance).

Variable	Method	ANOVA	Average value by displacement (%) 5 m	12.5 m	25 m	50 m
precision	Area	✗	98.06	92.44	74.13	45.79
	Rate	✗	99.27	95.94	80.06	48.19
	GA	✗	92.46	88.81	75.67	50.00
	Ctt	✓	55.59	55.54	55.29	54.20
	SMHD	✗	81.29	77.15	64.44	39.91
recall	Area	✗	95.88	85.76	61.94	34.36
	Rate	✗	93.53	74.71	44.15	20.46
	GA	✗	89.49	73.67	47.59	24.52
	Ctt	✓	33.46	33.41	33.18	32.19
	SMHD	✗	97.60	94.47	71.05	34.09
F-measure	Area	✗	96.93	88.71	66.47	37.80
	Rate	✗	96.24	83.40	55.73	27.84
	GA	✗	90.76	79.90	57.13	31.64
	Ctt	✓	40.79	40.69	40.42	39.33
	SMHD	✗	88.56	84.79	67.29	36.20

Notes: matching methods are described in Section 4.3.3.1.

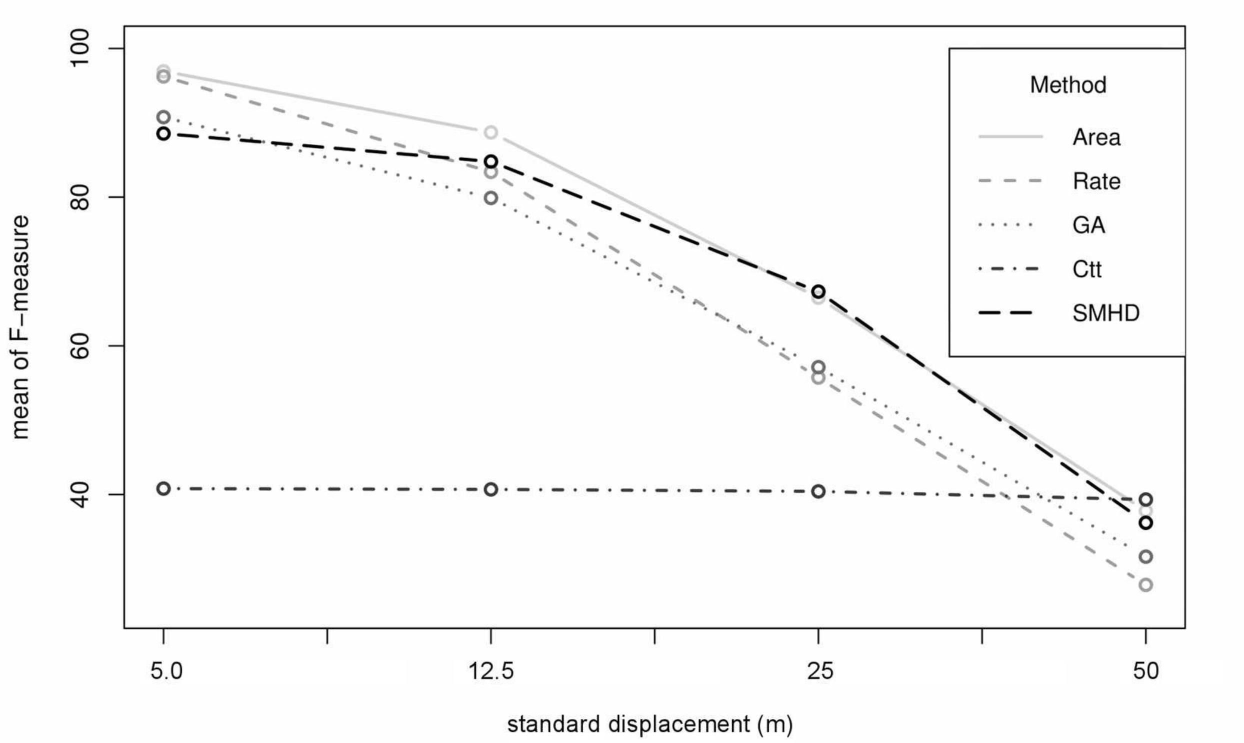


Figure 4.15. Average F-measure for each area matching method in function of standard displacement for random disturbs (essay A9).

4.3.4. Discussion of the feature matching results

The designed experiments for matching methods intended to assess the influence of diverse factors over four controlled variables (precision, recall, F-measure, and time). Each geometric primitive (point, line, area) had its own experiment and factors.

Regarding point matching methods, the experiment revealed that those methods based on Euclidean distance achieved the best quality (precision, recall, F-measure) and the best computational cost (time). However, Euclidean-based methods showed up most sensitive to systematic or random disturbances than the method based on geographic context distance. Some methods used the both nearest criterion and reached good results because the point datasets has predominantly 1:1 pairs, fact that might not be guaranteed when assessing unknown datasets. Regarding the context of point features, ANOVA test indicated that their geographic context does not influence the performance of assessed methods.

Most of matching methods studies are focused on linear data (Xavier et al. 2016a). The six essays of our line matching experiment showed that one linear measure from the 'max-min' family (Hausdorff-based) and other from the 'dog-man' family (Fréchet-based) have comparable performance. The best quality and time performance was obtained by the short-line median Hausdorff distance (SMHD), when combined with the partial orientation measure, using the selection criteria closer. However, none of assessed matching methods demonstrated be robust in presence of systematic or random disturbance. Regarding the morphology class of line features, ANOVA test revealed that this factor does not affect the precision, but it influences the recall and F-measure, with the best results for class CL2 (smooth).

The area matching experiment had nine essays. The results exposed that the overlap area measure presented the best overall performance. This straight measure reached an average F-measure above 98%, as of the overlap rate measure. However, overlap rate showed up more sensitive to disturbance than overlap area. The set of measures combined by means of a genetic algorithm (GA), when using the closer criterion, reached an average F-measure better than the SMHD-based method (93% against 89%), with a similar computational cost (half of overlap methods). Regarding the morphology class of area features, ANOVA test indicated that this factor does not matter for the controlled variables using the assessed matching methods. Regarding the geographic context of area features, ANOVA set forth that this factor has not significant influence over controlled variables. Finally, ANOVA revealed that the context-based method is less sensitive to systematic or random perturbations than the other methods, similarly to what happened in the point matching experiment. It is possible that the presence of diverse multiple corresponding pairs (1:n and m:n) in area datasets had affected the performance of context method, which used the both nearest criterion (limited to 1:1 pairs).

4.4. Internal matching

The experiments for the internal matching method test method described in Section 3.2 using the datasets created in Section 4.2 for finding correspondences between parts of features. In this study we are adopting a quality model of the Brazilian standard (DCT 2016a). This standard describes as positional quality procedure a point-based method originally developed in 1984 (DCT 2016a). Therefore, we develop this internal matching method in order to increase the quantity of points for the positional quality assessment, since it allows to use area and line features at the quality process.

There are two experiments in this phase: one for area features and other for line features, in this order. The last subsection brings the results discussion.

4.4.1. Area features

Regarding the internal matching using areas' parts we have two regions: S1A (700 parts, or vertices) and S7A (2109 vertices). We manually matched the internal parts of assessed polygons taking into account just the one-to-one (1:1) corresponding case. This restriction was considered because this internal matching method is designed to positional quality evaluation, which requires single point pairs (1:1).

In this experiment we used a straight method based on the Euclidean distance, already used for point matching at feature level (see Section 4.3.1.1). In this method the parts (vertices) were extracted from original data using the same restrictions of the new proposed internal method (Section 3.2). We compared this Euclidean-based method with six configurations of the new internal method based on geographic context (Section 3.2). The Euclidean-based method works without any previous feature matching, however the context-based method requires feature matching prior to internal matching itself. In this case we used the overlap area measure, closer criterion, and 10 m² threshold, as used in other experiments (see Section 4.3.3.1). This experiment has seven matching methods, all using the both nearest criterion:

- Euclidean distance, 25 m threshold;
- Geographic context distance, three thresholds: 0.25, 0.50, and 0.90;
- Geographic context distance, three thresholds: 0.25, 0.50, and 0.90, and a difference between the gradients of point pairs below a threshold of $\pi/12$ rad.

The results for these methods over the controlled variables (precision, recall, F-measure, and time) are presented in Table 4.29.

Table 4.29. Results for internal matching over areal features.

			CTT			CTTg		
Variable	**Region**	**25B**	0.25	0.50	0.90	0.25	0.50	0.90
precision	S1A	79.04	91.41	82.84	82.30	95.11	87.79	87.16
	S7A	66.51	88.22	83.16	82.78	91.58	88.28	88.11
recall	S1A	91.22	75.16	89.94	96.57	74.95	89.29	95.93
	S7A	67.55	90.99	96.39	97.18	89.50	94.44	95.22
F-measure	S1A	84.69	82.49	86.24	88.87	83.84	88.53	91.33
	S7A	67.03	89.58	89.29	89.40	90.53	91.26	91.53
time	S1A	253	252	256	255	252	252	253
(ms)	S7A	768	987	968	987	952	947	972

Notes: 25B means Euclidean distance, 25 m threshold. CTT means geographic context distance (Section 3.2), with corresponding thresholds. CTTg means the same measure, but also using an exclusion threshold based on gradient difference (π/12 rad).

From Table 4.29 it is possible to observe that the Euclidean-based method, despite of its simplicity, it reached an average precision above 50%, with a simple and fast algorithm. Regarding the context-based configurations of the new proposed method, it obtained the best quality results (precision/recall/F-measure), with fair computation cost (similar to the Euclidean-based method) considering that it needs to run a feature matching prior to the internal matching calculus. Analysing the use or not of the gradient threshold, it is possible to note that the global performance (F-measure) is almost the same, but with a slight gain in precision. The similarity threshold can be adjusted in order to reach the best relation precision/recall. Figure 4.16 illustrates an example for the region S7A. In this case, a similarity threshold in [0.6, 0.9] reached a good relation between precision and recall.

In order to asses the robustness of internal methods, we performed an essay using the two regions (S1A and S7A) and the corresponding random disturbs datasets (see Section 4.2.4). There are 100 version of each standard displacement (5, 12.5, 25 and 50 m). Figure 4.17 shows an overview of the results for the variable F-measure. It is possible to note that the Euclidean-base method (25B) is more sensitive to random disturbs than the context-based methods (CTT and CTTg). We added the feature matching performance in the line graph in order to show that the quality of context-based internal matching depends on the quality of feature matching.

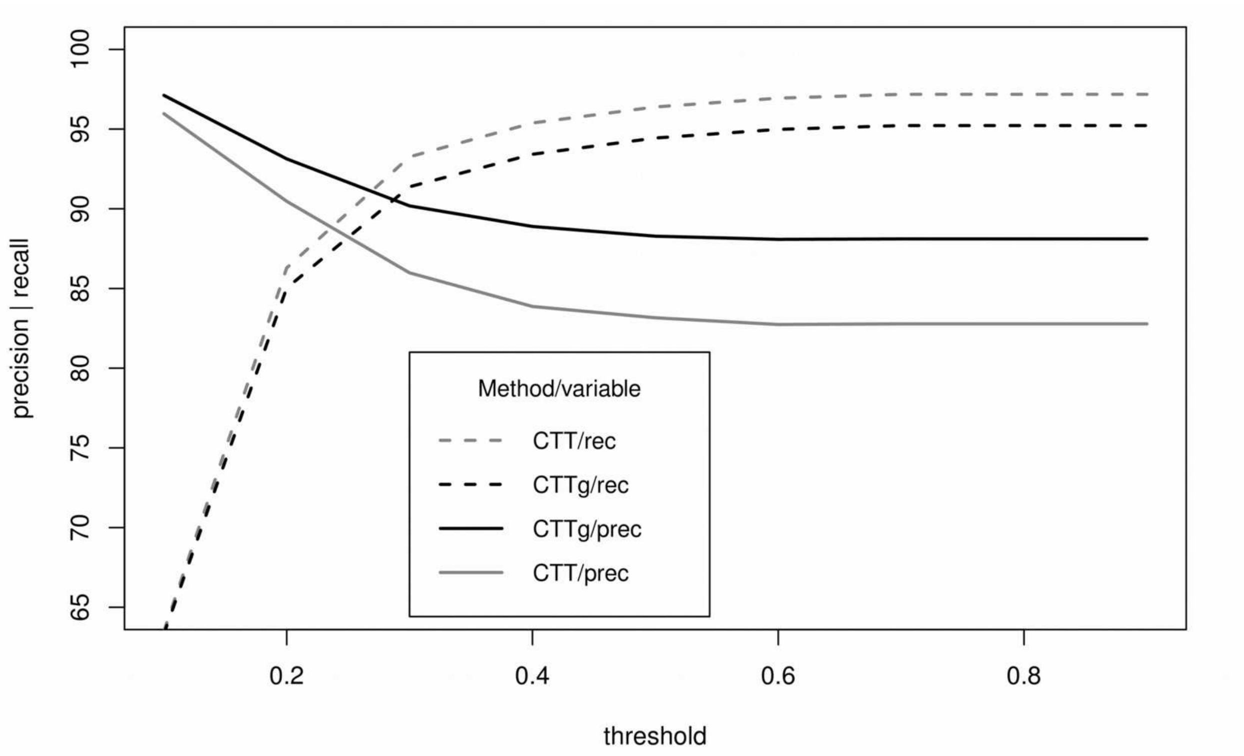


Figure 4.16. Precision and recall in function of the similarity threshold for context-based internal matching for the region S7A.

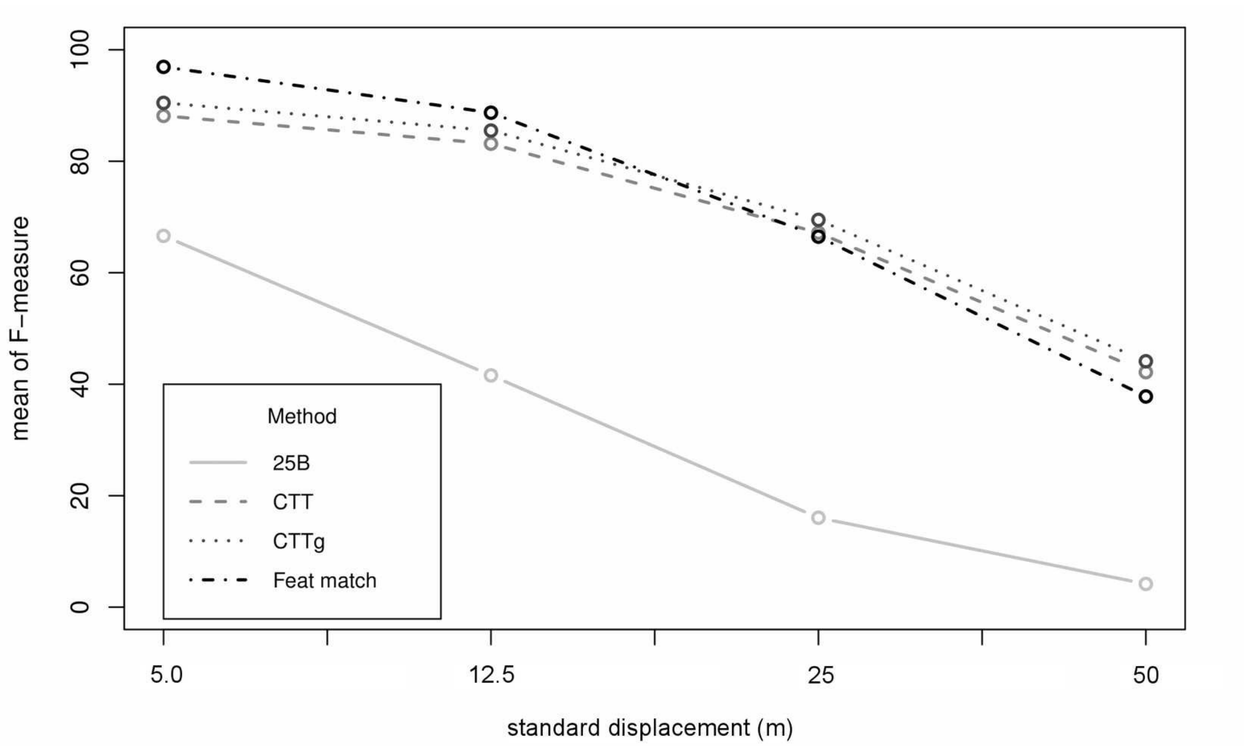


Figure 4.17. Average F-measure for each internal matching method in function of standard displacement for random disturbs. The fourth line (Feat match) refers to the feature matching method required by the context-based methods.

4.4.2. Line features

The internal matching using lines' parts is composed by data from only one region (S1L), with 266 vertices. As of in the areas experiment (Section 4.4.1) we manually matched the internal parts of assessed lines taking into account just the one-to-one (1:1) corresponding case. We used the straight method based on Euclidean distance and the two versions of context-based method (without and with gradients). The context-based method requires feature matching prior to internal matching itself. In this case we used the SMHD measure, closer criterion, and 10 m threshold, partial orientation 0.4 rad, as used in other experiments (see Section 4.3.2.1). This experiment has seven matching methods, all using the both nearest criterion:

- Euclidean distance, 25 m threshold;
- Geographic context distance, three thresholds: 0.25, 0.50, and 0.90;
- Geographic context distance, three thresholds: 0.25, 0.50, and 0.90, and a difference between the gradients of point pairs below a threshold of π/12 rad.

Table 4.30 presents the results for these methods over the controlled variables (precision, recall, F-measure, and time).

Table 4.30. Results for internal matching over linear features.

		CTT			**CTTg**		
Variable	**25B**	0.25	0.50	0.90	0.25	0.50	0.90
precision	32.97	44.44	45.21	42.42	64.29	59.18	58.18
recall	16.57	11.05	18.23	23.20	9.94	16.02	17.68
F-measure	22.06	17.70	25.98	30.00	17.22	25.21	27.12
time (ms)	98	499	519	513	505	506	505

Notes: 25B means Euclidean distance, 25 m threshold. CTT means geographic context distance (Section 3.2), with corresponding thresholds. CTTg means the same measure, but also using an exclusion threshold based on gradient difference (π/12 rad).

Once more the Euclidean-based measure (25B) reached performance values (F-measure) comparable with some configurations of the context-based methods. If we taking into consideration the average rates reached in the area experiment (Section 4.4.1), something above 80%, we can conclude that the line method requires improvements, even if we are using a limited experiment. An interesting result in this experiment is that we can note the precision improvement when using the gradient checking.

4.4.3. Discussion of the internal matching results

The internal matching experiments were prepared in order to check the validity of the new context-based method proposed in Section 3.2. There were two experiments: one for areal data and other for linear data.

Regarding areal data, the experiment revealed that the context-based method reached an average performance (F-measure) above 80%, and till 90% in some configurations. The overall precision grew when using the difference between the gradient of vertices, with a slight loss in recall rates. The context-based internal matching method requires a previous matching between component features (areas in this case). The results using disturbed datasets (random disturbance, Section 4.2.4) revealed that the internal matching performance decreases when the feature matching performance decreases, which evidenced the dependency of a high-performance method at the feature level. The current implementation of the new internal matching method has two highlighted gains when compared with other equivalent methods (Fan et al. 2014, Ruiz-Lendínez et al. 2015): (1) this method are able to deal with many-to-many area pairs and find their corresponding parts; and (2) this method can work with polygon holes, what can increasing the quantity of corresponding parts.

We tested the context-based method using linear data in one experiment limited to just one region. The results exposed that the context-based method did not achieve results equivalent to those in the area experiment. The method obtained precisions above 50%, which might be valid in some applications. However, we believe that this internal matching for linear data needs some improvements in order to be used in real-case issues.

4.5. Quality evaluation procedures

In this section we test the validity of the quality evaluation procedures described in the Evalution tier (Section 3.3) of proposed architecture. We used the datasets prepared in the matching testbed (Section 4.2). The main objective of this phase is to identify the degree of automation for the quality evaluation procedures so it can be implemented in the proposed quality control service.

The experiments are divided according to the quality element in consideration: topological consistency, completeness, and positional quality. The next subsections details each one.

4.5.1. Topological consistency

In the first experiment for quality evaluation procedures we verify the implementation of the quality element topological consistency. We implemented three quality procedures for line and area data: invalid simple lines (212), invalid polygons (213), and invalid overlaps (215) (details in Section 3.3). All procedures are internal and with full inspection. We took the initial datasets of the matching testbed (SL, SA, BL, BA, see Section 4.2.1) and submitted them to the quality evaluation procedures. The results are shown in Table 4.31.

Table 4.31. Results for topological consistency evaluation in all line and area datasets.

Dataset	Size	Quality procedure result 212	213	215	Observations
SL	1255	1/1255	-	-	Error on 4026
SA	1686	-	0	0	No errors found
BL	6496	0	-	-	No errors found
BA	19041	-	0	0	No errors found

From Table 4.31 we can identify that just one topological error was found for the considered datasets. The error at line object 4026 was an invalid internal overlap, which dismisses the simple line characteristic. However, the line remains valid against the OGC specification (Herring 2011). These results confirmed that these procedures are working as planned, once we tested these same datasets using other GIS tools (Xavier et al. 2017).

4.5.2. Completeness

The experiment for the quality element completeness allowed assessing the implementation of commission (101) and omission (103) procedures (see Section 3.3). There are three essays in this experiment: the K1 essay verifies if the implementation works as planned for a full inspection; the K2 essay assess the influence of a matching method over the quality results; and the K3 essay assess the influence of random disturbance over the quality results.

In the first essay (K1) we checked the validity of implementation against a *rule of thumb* for the initial datasets of matching testbed (Section 4.2.1), divided into their 27 regions: three geometric primitives (point, line, area) by nine regions. We adopted as the universe of discourse de 1:10,000 data (BP, BL, BA) and as test data those of 1:25,000 scale (SP, SL, SA). The rule of thumb, or *manual* procedure, that is considered the 'truth', was created from the manually matched pairs, which provided the commission and omission cases in assessed datasets. It is important to note that some test datasets (S#[PLA]) were determined using random selection in the original datasets (see Section 3.1.2.1). So, it is possible to note some large omission rates in some regions.

In order to compare our implementation, we adopted the following matching methods for each kind of geometry

- Point: Euclidean distance, closer criterion, 10 m threshold (Section 4.3.1.1);
- Line: SMHD with 10 m threshold, closer criterion, partial orientation less than 0.4 rad (Section 4.3.2.1); and
- Area: overlap area measure, closer criterion, 10 m² threshold (Section 4.3.3.1).

We ran the automatic quality procedures for those 27 regions using full inspection. The results comparing the manual (rule of thumb) and the automatic (full inspection) procedure are presented in Table 4.32.

Table 4.32. Comparison between manual and automatic completeness procedures (essay K1).

	Manual					Automatic	
Region	**Com**	**Om**	**Ref size**	**Com rate**	**Om rate**	**Com rate**	**Om rate**
S1P	19	443	535	3.55	82.80	3.93	82.80
S2P	12	1231	1337	0.90	92.07	0.90	91.92
S3P	10	1028	1145	0.87	89.78	0.79	89.17
S4P	19	2000	2102	0.90	95.15	0.86	95.05
S5P	11	1827	1937	0.57	94.32	0.52	94.32
S6P	24	2312	2410	1.00	95.93	0.95	95.81
S7P	10	122	297	3.37	41.08	3.70	40.74
S8P	13	148	336	3.87	44.05	3.87	43.75
S9P	4	1290	1414	0.28	91.23	0.35	91.87
S1L	12	552	765	1.57	72.16	1.83	72.68
S2L	10	470	681	1.47	69.02	1.47	69.02
S3L	11	448	654	1.68	68.50	1.68	68.96
S4L	7	718	944	0.74	76.06	0.95	76.17
S5L	15	800	958	1.57	83.51	1.98	84.13
S6L	13	1227	1499	0.87	81.85	1.80	82.12
S7L	22	557	756	2.91	73.68	3.17	74.21
S8L	24	732	877	2.74	83.47	2.39	83.58
S9L	10	294	454	2.20	64.76	2.42	64.76
S1A	7	526	639	1.10	82.32	1.41	82.63
S2A	10	1605	1763	0.57	91.04	0.62	91.04
S3A	11	1344	1484	0.74	90.57	0.94	90.84
S4A	11	2309	2458	0.45	93.94	0.45	93.90
S5A	6	1385	1617	0.37	85.65	0.37	85.59
S6A	32	3212	3359	0.95	95.62	0.98	95.68
S7A	18	167	426	4.23	39.20	3.99	38.97
S8A	7	551	801	0.87	68.79	0.87	69.16
S9A	2	1789	2079	0.10	86.05	0.10	86.05

Notes: 'Com' and 'Om' means commission cases and omission cases, respectively. 'Ref size' means the reference size which the rates refer to. 'Com rate' and 'Om rate' refers to commission and omission rates (in %), respectively.

From Table 4.32 is possible to note that the implementation of automatic completeness assessment worked similarly to the manual procedure for a full inspection. The following line graphs illustrates this result. Figure 4.18 shows the values for commission rates in the two methods: manual (rule of thumb) and automatic (full inspection). In this chart is possible to identify some noticeable variations in commission rates between considered methods. This occurred because there are few commission cases in some assessed datasets (e.g., 13 in S6L), so a variation of few objects (e.g., 5) causes proportionally more changes in commission rates. Figure 4.19 brings the same analysis but for omission rates.

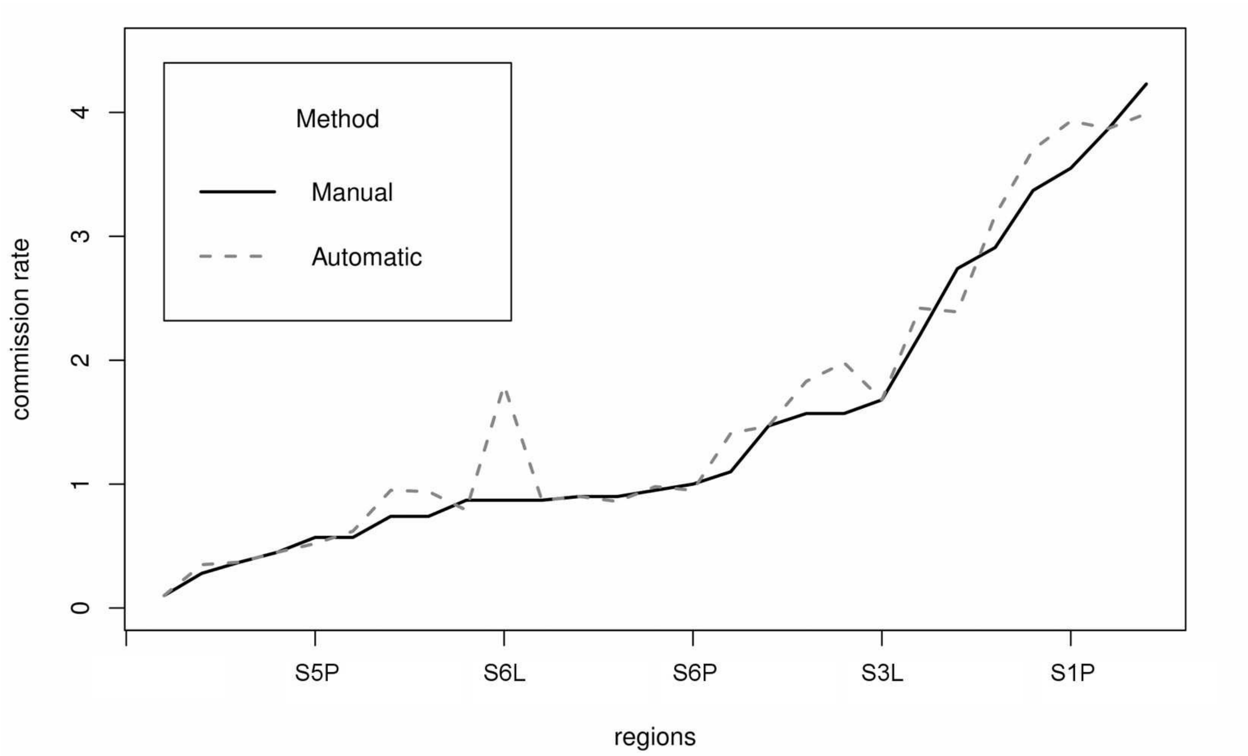


Figure 4.18. Commission rates in 27 regions using manual (rule of thumb) and automatic (full inspection) methods (essay K1).

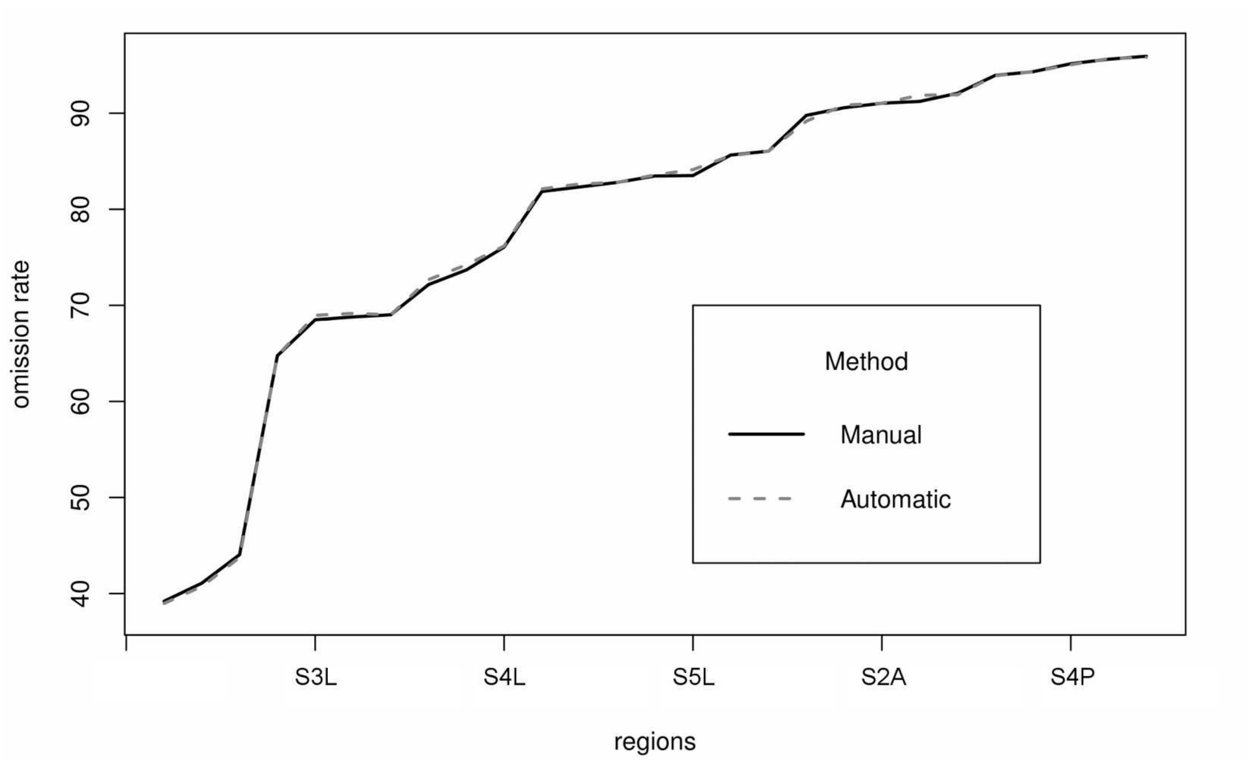


Figure 4.19. Omission rates in 27 regions using manual (rule of thumb) and automatic (full inspection) methods (essay K1).

In the next essay (K2) we try to detect the influence of matching methods in quality results for completeness. For this purpose we used the original linear datasets, omission rates, and three matching methods (Section 4.3.2.1): (1) SMHD with 10 m threshold, closer criterion, partial orientation less than 0.4 rad (used in essay K1); (2) Partial discrete Fréchet distance (PFre), closer criterion, 15 m threshold; and (3) single-buffer method, buffer length 10 m, threshold 0.5. The chart in Figure 4.20 illustrates the results. It is possible to note that the differences among the performance of matching methods influenced the quality result of the automatic evaluation.

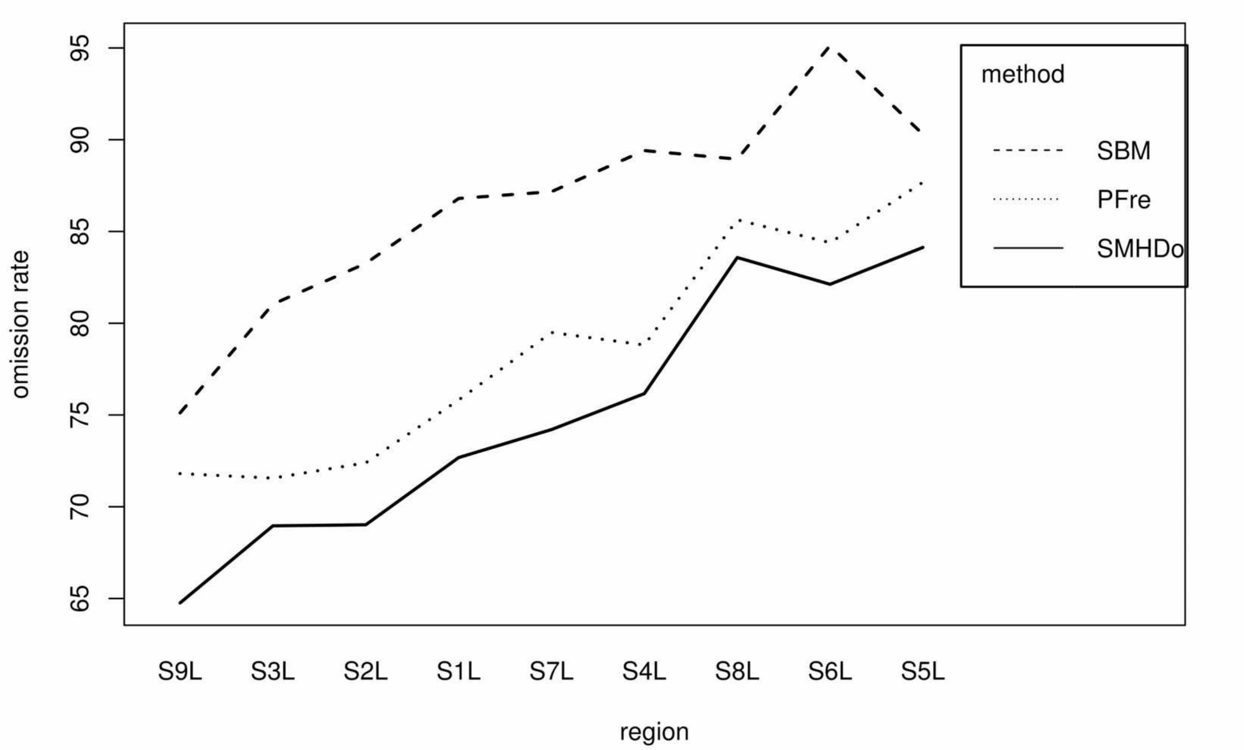


Figure 4.20. Omission rates in line regions for different matching methods (essay K2). SMHDo method is closest to the manual procedure.

The third essay in the completeness experiment (K3) investigates the influence of random disturbance over the quality results. In this essay we used point data and the quality element commission. As shown in Section 4.2.4, there are four standard displacements: 5, 12.5, 25 and 50 meters, and 100 different vector fields for each displacement. These perturbed point datasets were submitted to the automatic completeness evaluation that were using the same matching method of the essay K1 (Euclidean distance, closer criterion, 10 m threshold). The results are illustrated in the graph of Figure 4.21. The no-disturb result (None) was added to comparison purposes. From this chart we can notice that when the amplitude of disturbance is bellow the positional quality limit (5 m), the quality results remains approximately the same. However, it is possible to see some larger commission rates, mainly for the agglomerated areas (S7P and S8P).

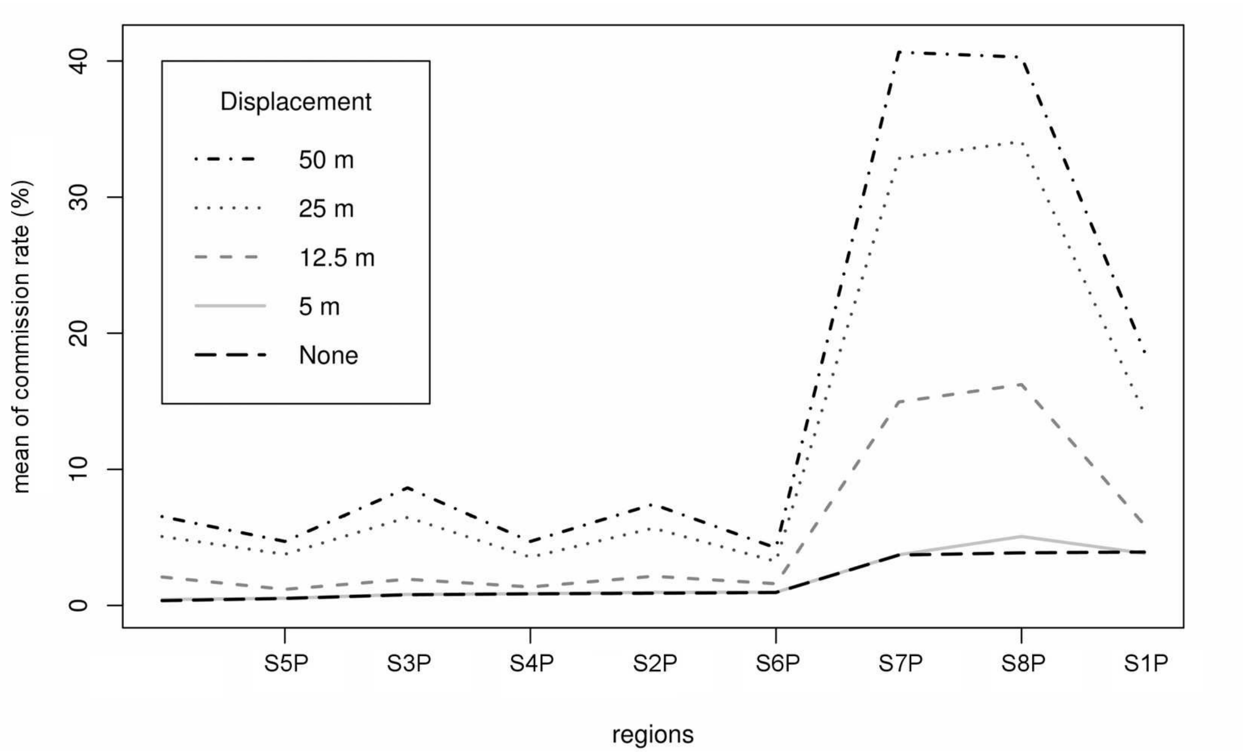


Figure 4.21. Average commission rate for point regions in function of standard displacement for random disturbance (essay K3). The last line (None) refers to no-disturb data.

4.5.3. Positional accuracy

The experiment for the positional accuracy aims to assessing the implementation of planimetry quality (301) procedure (see Section 3.3) by means of essay Q1.

The essay Q1 aims to check the performance of the automatic positional accuracy evaluation against a rule of thumb for point and area initial datasets of matching testbed (Section 4.2.1), in 11 regions: S1P-S19, S1A, and S7A. These regions were selected because the positional accuracy method is based on point calculus (see Section 3.3), so we have the manually matched all point regions, and two area regions (S1A and S7A) for the internal matching experiment (Section 4.4.1). We adopted as the universe of discourse the 1:10,000 data (BP, BA) and as test data those of 1:25,000 scale (SP, SA). The rule of thumb, or *manual* procedure, that is considered the 'truth', was created from the manually matched pairs and then we applied the planimetry evaluation described in CQDG (DCT 2016a) for a full inspection. This procedures returns a quality category (named *PEC*) according to the 90% percentile of planimetric errors (ER90) and the corresponding root mean square (RMS).

In order to compare our implementation we adopted the same matching methods of the completeness experiment (see Section 4.5.2): (P) Euclidean distance and (A) overlap area. For internal matching we adopted the context-based method, both nearest criterion, 0.9 threshold, and gradient difference below π/12 rad (more details in Section

4.4.1). We ran the automatic quality procedures for 11 regions using full inspection. The results comparing the manual (rule of thumb) and the automatic (full inspection) procedure are drawn in Table 4.33.

Table 4.33. Comparison between manual and automatic positional accuracy procedures (essay Q1).

			Manual			Automatic		
Region	**Cells**	**Points**	**PEC**	**ER90**	**RMS**	**PEC**	**ER90**	**RMS**
S1P	59	93	A	3.303	3.033	A	3.288	2.63
S2P	44	103	A	5.721	3.695	A	5.721	3.695
S3P	30	116	A	3.050	2.305	A	3.661	2.99
S4P	23	100	A	3.961	2.807	A	5.064	3.158
S5P	11	109	A	3.393	2.164	A	3.422	2.301
S6P	41	96	A	5.141	3.286	A	5.589	3.559
S7P	3	174	A	4.808	3.710	A	4.811	3.658
S8P	4	188	B	6.723	4.793	B	6.692	4.748
S9P	11	114	B	7.527	5.191	B	6.707	4.445
S1A	57	467	A	4.019	2.673	A	4.734	3.758
S7A	5	1276	B	6.212	4.558	B	6.375	4.911

Notes: *Cells* is the number of cells in each dataset that is considered for evaluation. *Points* is the number of coordinates used in the manual evaluation. *PEC* means the quality category for planimetric assessment.

Table 4.33 revealed that the implementation of automatic positional accuracy assessment for a full inspection worked alike to the manual procedure. The quality category (PEC) was preserved in the 11 cases. Regarding the evaluation of positional errors (ER90 and RMS), the regions presented similar behaviours. Figure 4.22 illustrates this result. The line graph indicates that the automatic procedures provided slightly high error values.

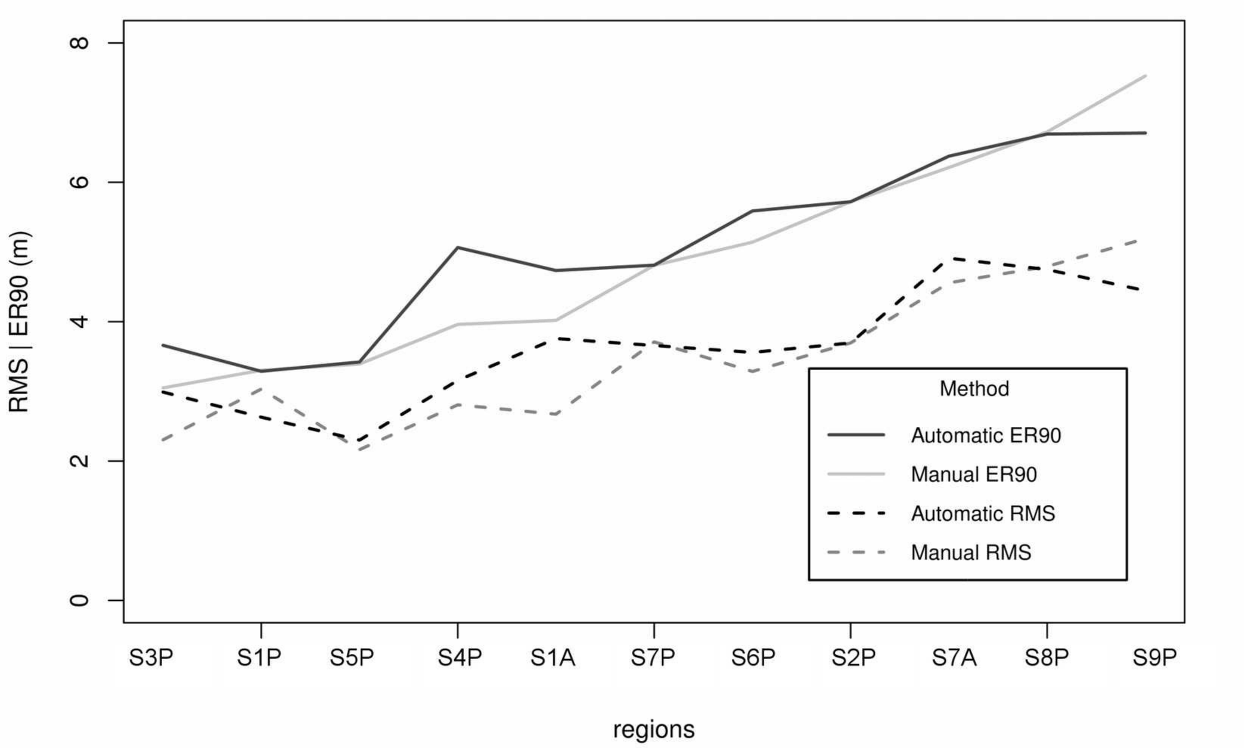


Figure 4.22. Root mean square (RMS) and percentil 90% for horizontal errors (ER90) in 11 regions using manual and automatic methods (essay Q1).

4.5.4. Discussion of the quality evaluation results

In this study we chose some quality evaluation procedures described in the Brazilian standard (DCT 2016a) and we implemented an automatic version of them. The experiments about quality evaluation procedures aimed to assess these implementations in order to checking if these can be applied in the proposed quality evaluation framework. The experiments were designed to test internal (topological consistency) and external (completeness, positional accuracy) quality procedures.

Topological consistency procedures are internal methods, so these do not require any external reference for comparisons. This feature makes this quality element feasible for straight implementation as we can see in Mobasheri (2013) and in the Spanish Army (Tagg 2015). The experiments for this quality element showed that the automation of the three elected procedures worked as provided in the standard.

Regarding the completeness element, we implemented commission and omission procedures to assess test datasets against reference datasets. First we checked the validity of the automatic evaluation against a manual evaluation in 27 regions of our matching testbed (see Section 4.2.1). It allowed us to verify the performance of automatic completeness in all geometric primitives (point, line, area). As the completeness assessment essentially uses the matching pairs to comparison, this factor (the matching performance) becomes a key point in this procedure. The results revealed that the automatic implementation worked satisfactorily for the full inspection.

Taking into consideration the influence of selecting a different matching method, the results in essay K2 revealed that we obtains different quality results for different matching methods in completeness. Regarding the influence of random disturbance over the quality results, the essay K3 revealed that large positional displacements over assessed data influence the quality results of completeness.

In the positional accuracy experiment we verified the automatic implementation of the planimetry procedure of CQDG. In the essay Q1 we checked the performance of a full inspection automatic method against a manually performed method in 11 regions: nine point regions and two area regions. The results revealed that the automatic planimetry evaluation performed similar to the manual procedure, with the quality category (PEC) preserved in all considered regions.

Finally, from these experiments we can conclude that is possible to implement geospatial data quality evaluation procedures with a high degree of automation. So this procedures can be plugged into a WPS server and works in the internet. However, we also noted that the quality of automation is directly linked to the quality of matching methods (at all levels) for those procedures that require an external dataset to comparison, i.e., the external methods. The limitation of performance in the used matching method becomes a limitation of the automatic quality assessment for external procedures.

4.6. WPS tier and quality report

In order to validate our proposal for the WPS tier presented in Chapter 3, we developed two web services for quality assessment of positional accuracy using the American standard (NSSDA, FGDC 1998) and the Spanish standard (UNE 148002, AENOR 2015). The following subsections presents the results of each service.

Section 4.6.1 was first presented in Xavier et al. (2015b) and Section 4.6.2 was presented in Xavier et al. (2015c).

4.6.1. NSSDA service

The core of NSSDA service is the *PointEvaluation* class, an implementation of the *AbstractProcess* interface. Considering that the NSSDA procedure is applied over pairs of homologous points, from a reference and a test site, the first task is to perform a matching between reference and test datasets. For this purpose we adopted a simple solution using the nearest neighbour strategy taking into account only 1:1 matches. Since our objective here is assess the WPS tier, we chose this effortless matching approach for simplification purposes. After the matching, the calculus procedure runs straightforward, and the execution returns a double value (in meters) that represents the result of the NSSDA evaluation.

For this experiment we prepared two datasets of point data in the Shapefile format, with approximately 40 points in each one. Then we created a simple HTML5/Javascript client able to convert the data into a WPS execute request. The simple client was used to encode the Shapefile data into base64 encoding (Josefsson 2006), mount the execute request, send it to server, and receive the response. Figure 4.23 brings an extract of the returned response in XML.


```xml
<wps:ExecuteResponse service="WPS"
  serviceInstance="http://localhost/"
  version="1.0.0"
  xml:lang="en"
  xmlns:wps="http://www.opengis.net/wps/1.0.0"
  xmlns:ows="http://www.opengis.net/ows/1.1"
  xmlns:xlink="http://www.w3.org/1999/xlink">

  <wps:Process wps:processVersion="1.0">
    <ows:Identifier>PointEvaluation</ows:Identifier>
    <ows:Title>Point Evaluation - Data Quality</ows:Title>
  </wps:Process>

  <wps:Status creationTime="2015-03-27T13:59:56Z">
    <wps:ProcessSucceeded>Done</wps:ProcessSucceeded>
  </wps:Status>

  <wps:ProcessOutputs>
    <wps:Output>
      <ows:Identifier>NSSDA</ows:Identifier>
      <ows:Title>NSSDA result as real</ows:Title>
      <wps:Data>
        <wps:LiteralData> 18.3520 </wps:LiteralData>
      </wps:Data>
    </wps:Output>
  </wps:ProcessOutputs>

</wps:ExecuteResponse>
```


Positional accuracy value

Figure 4.23. Response of an *execute* request to the NSSDA service (Xavier et al. 2015b).

The positional accuracy value returned by the server was calculated following the NSSDA methodology. This value represents the horizontal positional accuracy of tested data against the reference data at 95% confidence level.

4.6.2. UNE 148002 service

The UNE 148002 service was implemented over the existing WPS architecture. In order to check if the service respond as expected in the Spanish standard, we prepared an experiment with three essays: (1) point data and Euclidean measures, (2) line data and Hausdorff distance (HD), and (3) line data and single buffer method (SBM). For this experiment was considered the sampling scheme for isolated lots (ISO 2859-2)

For the first essay (point data and Euclidean measure) we used 32 points measured from the building layer of BTN25 as test data. The reference point dataset was created using building instances from BCA10. We picked up 32 points from a central region of the city of Huelva. The maximum distance between homologous points is about six meters. For the point evaluation we considered that the lot size is 10 times greater the measured points, i.e., lot size equal to 320 features. We considered 'faulty data' those points whose reach a Euclidean distance greater than a given tolerance. Using these parameters we ran the service for a set of values for LQ and tolerance, as shown in Table 4.34. For each pair of parameters (LQ, tolerance) the service defines the values of 'Sample size / Acceptance' using ISO 2859-2 (ISO 1985), and then calculates the quality control response, which can be an error, or the standard responses: accepted (Ac) or rejected (Re).

Table 4.34. Results for direct evaluation of point data using Euclidean distance (Xavier et al. 2015c).

LQ (%)	Sample size / Acceptance	Tolerance (m) 4	5	5.17	6
5.0	50 / 0	Error	Error	Error	Error
8.0	32 / 0	Re	Re	Re	Re
12.5	32 / 1	Re	Re	Re	Ac
20	20 / 1	Re	Re	Re	Ac
32	20 / 3	Re	Ac	Ac	Ac

The results shown in Table 4.34 indicate that the UNE 148002 service is working in accordance with the procedures described in the standard for point evaluation. We added to the list of tolerances the resulting value for the NSSDA evaluation (5.17 m), which was calculated using other NSSDA service (Section 4.2.1). For the LQ = 5% the service threw error messages indicating that there was an insufficient sample size for this configuration. In this experiment there was two errors greater than five meters, so they prevented the acceptance of the lot in many configurations, like LQ between 8% and 20% in this tolerance. For the configurations using a sampling size of 20 elements, the service randomly selected the 20 from the 32 given points.

For the line essays we chose an extract with 50 objects of line data in systematic disturbance datasets (see Figure 4.24). For the line assessment we considered the lot size equal to 2000 features.

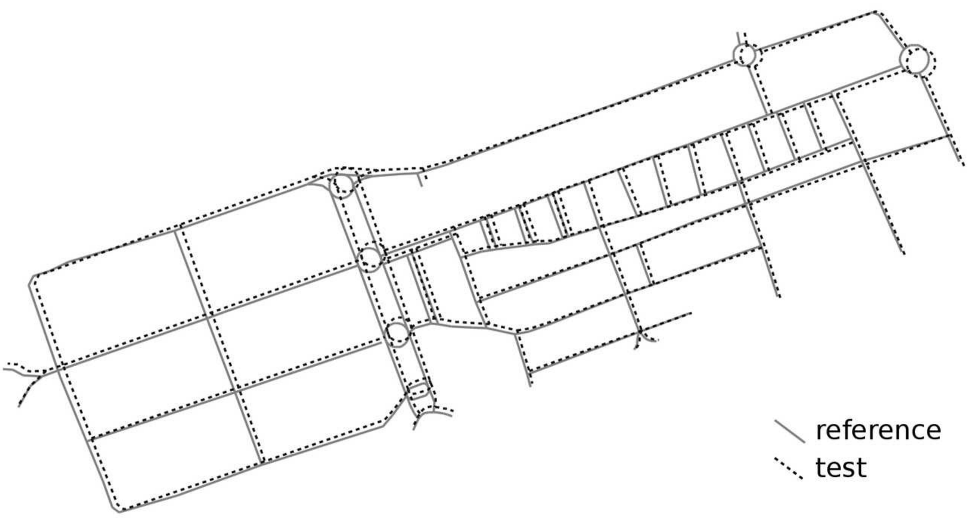


Figure 4.24. Line datasets for the UNE 148002 service experiment.

In the first essay for line data we considered faulty data those pairs that reach a Hausdorff distance greater than a given tolerance, for a certain LQ. Table 4.35 presents the results for some configurations of LQ and tolerance, which the last one (18.2) is greater than the maximum value. The results indicate that the service is working in conformance with the UNE 148002 procedure. Regarding the sampling size, the service worked in its limit, it used all the 50 available features when possible, and it threw an error message when the user required a more restricted limiting quality.

Table 4.35. Results for direct evaluation of line data using HD (Xavier et al. 2015c).

LQ (%)	**Sample size / Acceptance**	**Tolerance (m)**		
		10	15	18.2
8.0	80 / 3	Error	Error	Error
12.5	50 / 3	Re	Re	Ac
20	50 / 5	Re	Ac	Ac
32	50 / 10	Re	Ac	Ac

The second essay for line assessment used the SBM as the indicative of faulty or valid data. This measure require two parameters: a buffer length that is created around the reference line, and an overlap proportion, that represents the proportion of the test feature's length that is located within the buffer. This combination of parameters (buffer length/overlap proportion) was used as the tolerance. The used tolerances were a combination of three buffer lengths (10, 15, 18 m) with three overlap proportions (99.8, 90, 67%). Table 4.36 shows that the UNE 148002 service worked satisfactorily, so it returned responses according to the standard.

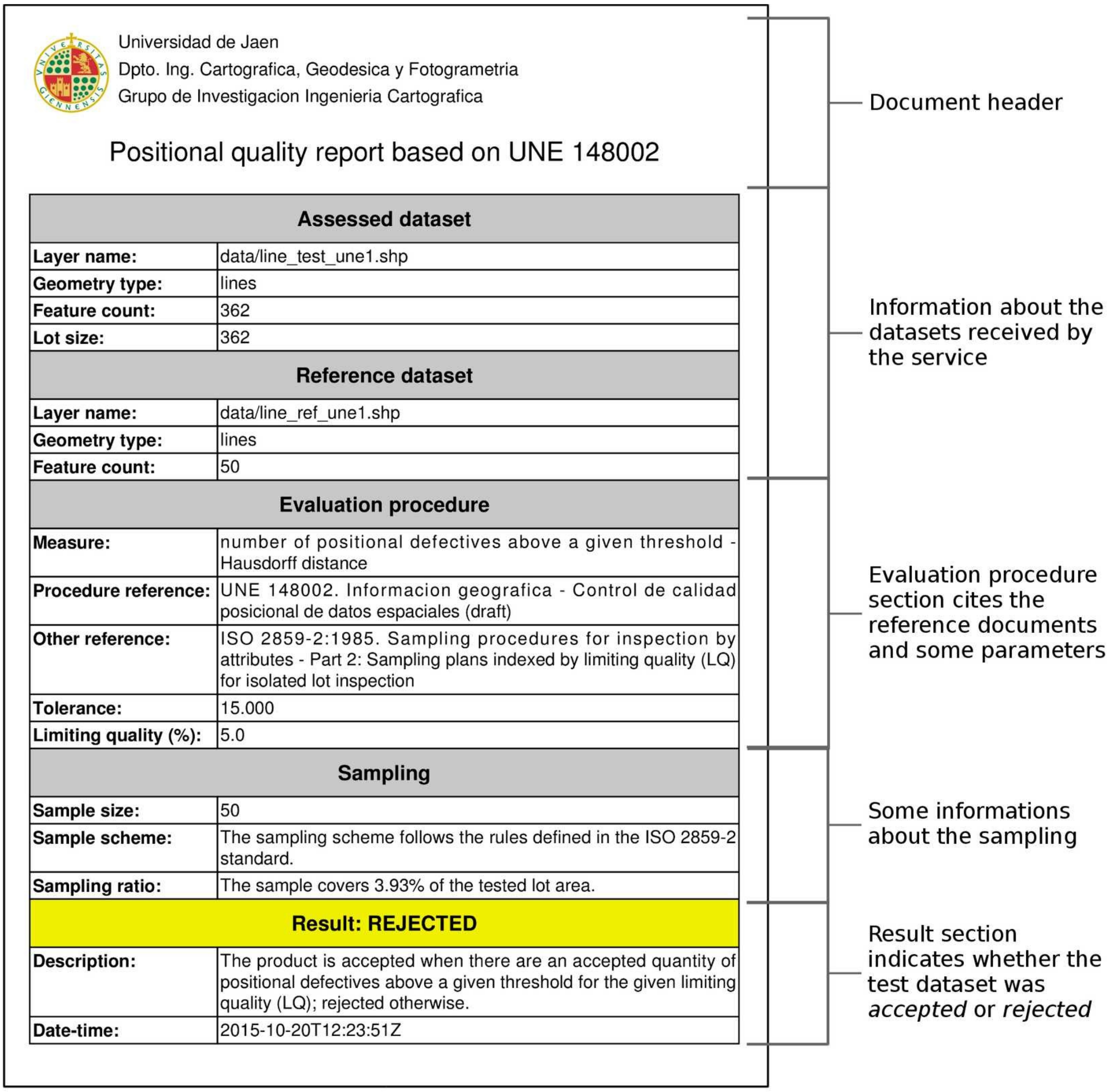
Table 4.36. Results for direct evaluation of line data using SBM (Xavier et al. 2015c).

LQ (%)	**Sample size / Acceptance**	**Buffer length (m) / overlap proportion**						
		10 / 0.998	10 / 0.90	10 / 0.67	15 / 0.998	15 / 0.90	15 / 0.67	18 / any
8.0	80 / 3	Error	Error	Error	Error	Error	Error	Error
12.5	50 / 3	Re	Re	Re	Re	Ac	Ac	Ac
20	50 / 5	Re	Re	Re	Ac	Ac	Ac	Ac
32	50 / 10	Re	Re	Ac	Ac	Ac	Ac	Ac

In order to check the flexibility of WPS for output formats, in the UNE 148002 service was implemented two outputs for the *execute* operation: XML report and PDF report.

The XML report is based on the concepts of the new ISO standards regarding quality evaluation (ISO 19157, ISO 2013) and metadata (ISO 19115-1, ISO 2014). The ISO standard that deals with the XML schema implementation was recently published (ISO/TS 19157-2, ISO 2016). The corresponding XML schemas that can be found at Metadata for Data Quality (MDQ) page (ISO 2015). Based on these schemas, the UNE 148002 service is able to generate an XML report that encodes the quality element assessed (DQ_AbsoluteExternalPositionalAccuracy) as well as some metaquality elements, also described in the draft of Spanish standard. The generated quality report is an XML file that is valid against the XML grammar defined in the MDQ package.

The PDF report is a human-driven report based on the informations generated by the service for the XML report. Despite of the XML report can be read by humans too, as well as by machines, with the PDF report we intend to generate a quality appraisal more suitable for human users, so the UNE 148002 is focused on consumers' needs. Figure 4.25 presents an example of a PDF report generated by the web service.



Universidad de Jaen
Dpto. Ing. Cartografica, Geodesica y Fotogrametria
Grupo de Investigacion Ingenieria Cartografica

Positional quality report based on UNE 148002

Assessed dataset	
Layer name:	data/line_test_une1.shp
Geometry type:	lines
Feature count:	362
Lot size:	362
Reference dataset	
Layer name:	data/line_ref_une1.shp
Geometry type:	lines
Feature count:	50
Evaluation procedure	
Measure:	number of positional defectives above a given threshold - Hausdorff distance
Procedure reference:	UNE 148002. Informacion geografica - Control de calidad posicional de datos espaciales (draft)
Other reference:	ISO 2859-2:1985. Sampling procedures for inspection by attributes - Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection
Tolerance:	15.000
Limiting quality (%):	5.0
Sampling	
Sample size:	50
Sample scheme:	The sampling scheme follows the rules defined in the ISO 2859-2 standard.
Sampling ratio:	The sample covers 3.93% of the tested lot area.
Result: REJECTED	
Description:	The product is accepted when there are an accepted quantity of positional defectives above a given threshold for the given limiting quality (LQ); rejected otherwise.
Date-time:	2015-10-20T12:23:51Z

Figure 4.25. Sample PDF report for the UNE 148002 service (Xavier et al. 2015c).

4.6.3. Discussion of WPS tier results

The development of the WPS tier of this architecture aroused some aspects of the applicability of WPS while a service interface facing quality evaluation. These aspects can be divided into strengths and weakness.

The identified WPS strengths were:

- Multiple inputs and outputs: the WPS interface does not limit anyway the quantity, type, or format for inputs or outputs. This flexibility permits that a quality evaluation service created on top of a WPS framework can be able to generate various interrelated quality outputs, for example:
  - Quality report in PDF using some template of the evaluator;
  - DQ_DataQuality from ISO 19157 (ISO 2013) encoded in XML, or according to the legacy ISO 19115 (ISO 2003b);
  - Some literal value (like the Hausdorff distance) as a part of a quality evaluation chain.
- Ready for service chaining: the specification indicates some options and previous research pointed out its feasibility (Kiehle et al., 2007, Friis-Christensen et al., 2009). The inputs and outputs for data processing can be accessed as on-line resources, for example:
  - The reference data in a positional quality evaluation can be a file available on-line for download;
  - The test data in the same situation can be distributed by means of a Web Feature Service (WFS);
  - Some parameter in a quality evaluation procedure, like the NSSDA result, can be obtained from other WPS server.
- Process extension is relatively easy: any extension to some process can take advantage of the entire framework.
  - These experiments has shown that different quality procedures (NSSDA and UNE 148002) can share the common part of WPS interoperability framework (HTTP connexions, XML processing). The same server can host different quality procedures;
  - There is no limit for geometric representation: point, line or area data can be used;
  - The architecture is flexible enough to allow complex sampling schemes as described in ISO 2859-2, or 100% inspection.

Despite the WPS presents many advantages, in this study we have identified one open issue: there is no direct way to indicate local data in the processing. Sometimes the differential of a service may be its local data, which for some reason (e.g. license, security) cannot be available in the web. This 'reserved' dataset cannot be distributed, but it can be used in processing jobs, like a reference dataset in quality evaluation procedures. The WPS specification permits send data, or reference remote data, without the prevision of local data. The latest version of specification (WPS 2.0) also does not forecast the use of local data, but it uses the concepts of data for value or for reference.

In this study we propose to describe the possible reserved reference dataset into the metadata of the quality procedure that might use it. So, one of the parameters of quality evaluation named 'reference' would be left blank.

CHAPTER 5
CONCLUSIONS

The geomatics industry is living a data overload scenario which are raising new challenges to the authoritative data producers, or NMCAs. Today it is possible to find diverse datasets representing the same geographic extent from many producers: volunteers (e.g. OpenStreetMap), commercial mapping companies, and official mapping agencies (at distinct levels). Each one creates its datasets following its own acquisition rules (and sources), which leads us to the question: 'which one does fit my purposes?', a fitness for use issue (Servigne et al. 2006).

In this context, this study looked for a solution able to answer the key-question: 'what is the quality of this dataset?' Where the geospatial data is created almost automatically (e.g. Coumans 2016), we also need a quality evaluation tool capable to respond in the speed that the data are created. Therefore the main goal of this study was to develop a standardized web service with the capability to assess the quality of geospatial datasets in fully automatic way. In order to reach this goal we developed a framework for automatic evaluation of geospatial data quality, as described in Chapter 3. Then we tested each part of our solution for the quality control service in the experimental phase (Chapter 4).

The obtained results confirmed that the main goal was reached: we have a quality control service that automatically assess the quality of geospatial datasets with results comparable to the manual procedure. However, the proposed solution has its limitations, which are exposed in Section 5.1.

The main goal is a 'lead cable' to our secondary aims. In the following we revisit each secondary aim and its results:

- *A1: Implement a generic WPS server that supports the evaluation service*. In Chapter 3 we presented the architecture of our quality control service as the solution to our main question. Section 3.4 presented the WPS tier, and the results published in Xavier et al. (2015b) endorsed that the WPS server is working well;
- *A2: Develop a generic architecture able to accommodate any quality evaluation procedure over a web environment*. The results of quality control through WPS in the Section 4.6, and the results published in Xavier et al. (2015c) confirmed that this aim was reached;
- *A3: Develop a matching approach able to compare two datasets by combining different techniques found in literature*. We developed a broad experimental design in order to find adequate feature matching methods (Section 4.3). We also developed a new internal matching approach with practical results (Section 4.4);
- *A4: Use a common evaluation framework, if available, or develop a simple quality model as a proof of concept, all based on the new ISO standards*. During the course of this research project the Brazilian Army published the new geospatial data quality standard for national mapping, named CQDG (DCT 2016a). CQDG was adopted in this study as the quality model (Section 2.1); and
- *A5: Execute quality evaluation procedures from quality elements: positional accuracy, completeness, and logical consistency*. The results of automatic quality evaluation for diverse quality elements (topological consistency, commission, omission, positional accuracy) in Section 4.5 are evidences that confirmed this aim was reached.

The contributions of this study are manifold. We can divide them into three categories: (1) data quality evaluation, (2) data simulation, and (3) matching of geospatial data. The contributions relative to data quality evaluation are:

- Full automatic procedures to evaluate the topological consistency, confirming other studies (e.g. Mobasheri 2013);
- Full automatic procedures to evaluate the completeness based on the results of matching methods; and
- Full automatic procedures to evaluate the positional accuracy for the Brazilian standard using a new internal matching method.

The contributions of data simulation are:

- New method for systematic disturbance based on the affine transformation;
- New method for random disturbance based on displacement vector fields;
- New method to change the morphology of lines;

- New morphology classification for building data;
- New method to change the morphology of areas; and
- New classification for geographic context using shape context and GIS experts.

The contributions of matching of geospatial data are:

- New experimental design to test feature matching methods, as embracing as possible;
- New geographic context measure based on the shape context from Belongie et al. (2002);
- New feature matching method using this geographic context measure. This method proved to be less sensible to data disturbance (systematic or random) then other methods; and
- New internal matching method using this geographic context measure. We demonstrated the feasibility of this measure for area data.

In a classical book of cartography, Robinson et al. (1995) argued that 'one of the most difficult tasks for cartographers is to indicate to map readers the quality of data used'. In the web era, or as prefer some authors: the information age (Blinder 2006), we hope that the cartographers might delegate this 'painful' task to the machines.

5.1. Limitations and future work

Despite of the experimental phase had demonstrated the feasibility of our solution towards the quality control service, there are some limitations in this approach that can be solved in future research. The results revealed that the capacity of automatic quality assessment for external methods (those that require an external dataset) is directly related to the performance of the matching methods (at all levels). Then, the performance of the automatic quality control service depends on the performance of used matching methods. Advances in the matching methods research will bring advantages to other related areas, notably map conflation (Ruiz et al. 2011).

Starting from this study we can suggest further work in:

- Development of a temporal model for mitigate the natural time degradation. In Xavier et al (2016b) we developed the first design of a solution to use third-party data as a source for updating a reference database built over three concepts: rules, quarantine database, and conflation. We believe this field might provided interesting results for the GIScience community, and the NMCAs in particular;

- Integration of geodata with non-spatial data using linked data techniques (Kuhn et al. 2014). We believe that the links between related instances might be created using some matching method (Xavier et al. 2016a);
- Development and implementation of networks of WPS servers able to attend a new set of emerging 'users': the autonomous devices from the Internet of Things. Norris (2015) affirmed that in the next five to ten years the internet will be invaded by these devices.

CHAPTER 6
RESUMEN EXTENDIDO

6.1. Introducción

Actualmente hay muchas fuentes de datos geoespaciales disponibles para generar datos de forma casi instantánea. Imágenes de plataformas aéreas o satelitales y la popularización de los vehículos aéreos no tripulados (*drones*), han permitido generar conjuntos de datos geoespaciales de manera inmanejable, lo que algunos autores llaman de tendencia '*big data*' (datos masivos) (Crampton et al. 2013). Los <<Terabytes son muy típicos hoy en día>>, dijeron Traxler y Hesina (2017). Otra fuente de datos importante son los datos generados por voluntarios casi a diario (Neis y Zielstra 2014). Este escenario de sobrecarga de datos trae nuevos retos para los proveedores de datos espaciales oficiales, o Agencias Cartográficas Gubernamentales. Tradicionalmente, estas instituciones crean y gestionan conjuntos de datos oficiales de forma estandarizada. Sin embargo, hoy hay muchos 'productores' de datos que representan los mismos fenómenos, objetos geoespaciales, siguiendo sus propias reglas. Este nuevo escenario puede llevar a los usuarios preguntarse acerca de la calidad de esos datos.

En general, poca, o ninguna, información sobre la calidad de los datos espaciales está disponible, así que creemos que sería interesante un servicio web con la capacidad de evaluar la calidad de un conjunto de datos cualesquiera (conjunto a probar) frente a un conjunto de datos de referencia. El desarrollo de servicios de validación de calidad de datos es un tema atractivo en las agendas de investigación geoespacial y que está siendo desarrollado en proyectos actuales (Kruse 2014).

Una tendencia reciente en la industria geomática es automatizar la mayor parte de la cadena productiva, como podemos ver en los proyectos recientes, por ejemplo, el proyecto 'mapas como servicio' en Irlanda (Coumans 2016) y el uso de drones en cartografía catastral (Ramadhani et al 2016). Es justo asumir que la evaluación de la calidad de los datos también ha de experimentar esta tendencia.

El estado de la cuestión relativo la automatización del control de calidad para datos espaciales ha mostrado avances recientes. El estudio de Donaubauer et al., (2008) propone un servicio web con la capacidad para generar información de calidad de los datos. Su trabajo utilizó estándares bien definidos, con la interfaz *Web Processing Service* (WPS) (Schut 2007) en el proceso de control de calidad, y la norma ISO 19115 (ISO 2003b) para el informe de calidad por medio de metadatos. WPS es una especificación abierta del consorcio *Open Geospatial Consortium* (OGC). A pesar de la simplicidad del procedimiento de calidad, una simple superposición de datos previamente etiquetados con algunos elementos de calidad, este estudio parece ser el primer intento de un servicio de evaluación automática encontrado en la literatura. Otro estudio indicó también que la evaluación de la calidad puede ser ejecutada a través de un WPS (Mobasheri 2013). Más recientemente, Meek et al. (2016) presentaron una solución para la evaluación de la calidad de los datos de voluntarios mediante encadenamiento de servicios y WPS.

En la Universidad de Jaén se desarrolló una investigación exitosa enfocada a la automatización de la evaluación de la exactitud posicional (Ruiz-Lendínez 2012). El autor propuso una solución para la evaluación automática de la exactitud posicional de objetos poligonales utilizando un enfoque de casado apoyado por algoritmos genéticos. Su tesis presentó resultados alentadores y es nuestro punto de partida para la investigación actual.

La libre disponibilidad de información geográfica para los usuarios finales ha planteado cuestiones sobre el costo de mantenimiento para las Agencias Cartográficas Gubernamentales (Carpenter y Snell 2013). Sin embargo, mientras más información está disponible para los usuarios, más se hace necesario evaluar su calidad con el fin de identificar si estos datos se ajustan a los requerimientos de los usuarios. Esto puede ser la oportunidad para que un proveedor de datos fidedignos pueda desempeñar el papel de 'validador', proporcionando informes estandarizados y útiles sobre la calidad de los datos. Otra posibilidad es crear la certificación de calidad para datos geoespaciales, como apuntó Ariza-López (2013).

6.1.1. Hipótesis y objetivos

Nuestra pregunta principal surge de la necesidad de un servicio de evaluación de la calidad en línea: ¿hasta qué punto podemos automatizar la evaluación de la calidad de los datos geoespaciales?

El estado actual de la cuestión de la automatización del control de calidad para datos espaciales muestra algunas investigaciones recientes:

- Es posible generar información de calidad sobre un conjunto de datos espacial mediante la interfaz WPS (Donaubauer et al. 2008), confirmada también por un trabajo posterior (Mobasheri 2013);
- Estudios dentro del proyecto ESDIN describen servicios de evaluación automática de la calidad de datos (Beare et al. 2010);
- Ruiz-Lendínez (2012) propuso y demostró una solución factible para la automatización de la evaluación de la exactitud posicional usando una solución de casado;
- Ariza López (2013) argumentó que la calidad de datos espaciales ha recibido reciente y continuo desarrollo en las normas internacionales, en particular la 19157:2013 ISO (ISO 2013);
- Fan et al. (2014) demostró que es posible evaluar varios elementos de la calidad - compleción, exactitud posicional, exactitud temática y exactitud de la forma - para objetos del tipo edificio utilizando un conjunto de datos de prueba frente una referencia, donde el primer paso fue el casado entre los conjuntos de datos; y
- Brovelli et al. (2017) presentó un nuevo método para realizar las comparaciones entre dados VGI y datos oficiales de carreteras con un alto grado de automatización.

Tomando estos hechos como premisas de trabajo, podemos formular la hipótesis de trabajo H1: es posible un procedimiento de evaluación automática, sin intervención humana, que evalúe un conjunto de datos de prueba contra un conjunto de datos de referencia. Creemos que un servicio de validación de calidad de datos traerá beneficios tanto para los productores y como para los consumidores de datos. Así, por ejemplo, los productores de datos se pueden beneficiar de un conjunto de datos estandarizado para evaluar sus propios suministros externos de datos. Los consumidores pueden obtener informes de calidad usando un protocolo estándar (WPS).

Teniendo en cuenta esta hipótesis, nuestro principal objetivo es desarrollar un servicio web capaz de evaluar la calidad de los conjuntos de datos geoespaciales utilizando la interfaz estandarizada de WPS en modo completamente automático.

6.1.2. Importancia institucional y publicaciones

Este estudio ha sido desarrollado en el grupo de investigación GIIC (Grupo de Investigación en Ingeniería Cartográfica) de la Universidad de Jaén, España. Este grupo de investigación (TEP-164) ha producido estudios relevantes en el área de calidad de datos geoespaciales, entre ellos podemos citar: Ariza-López y Atkinson-

Gordo (2008), Ariza-López et al. (2011), Ariza-López y Mozas-Calvache (2012), Ariza-López y Rodríguez-Avi (2014), Ruiz Lendínez et al. (2016) y Gil de la Vega et al. (2016).

El proyecto de investigación fue apoyado por el Departamento de Ciencia y Tecnología (DCT) del Ejército de Tierra de Brasil, que patrocinó este proyecto en beneficio de su Servicio Geográfico (DSG). DSG lleva la información geoespacial en el ejército brasileño. Según la legislación brasileña (Brasil 1967) DSG es el órgano responsable de generar y mantener los estándares técnicos para la asignación cartografía terrestre en Brasil.

Este proyecto de investigación ha generado, hasta el momento presente, las siguientes publicaciones hasta el momento presente: Xavier et al. (2014), Ariza-López et al. (2015), Xavier et al. (2015a, 2015b, 2015c), y Xavier et al. (2016a, 2016b). Dos trabajos están actualmente bajo consideración de los comités editoriales: Ariza-López et al. (2017) y Xavier et al. (2017).

6.2. Método y material

Para alcanzar nuestro objetivo principal se propone una solución basada en una arquitectura para la evaluación de la calidad a través de servicios web. Esta solución se presenta a continuación. La Sección 6.2.2 presenta el material utilizado en los experimentos para su validación.

6.2.1. Servicio de Control de Calidad

Se propone una arquitectura de tres capas para una plataforma de servicios web centrada en el control de calidad de datos geoespaciales (ver Figure 6.1), a la que llamamos *servicio de control de calidad*. Desde un punto de vista de abajo hacia arriba, el primer nivel, *Acceso a datos*, es utilizado por los métodos de evaluación externa para gestionar datos de referencia: recuperación y casado. El segundo nivel, llamado *Evaluación*, implementa los procedimientos de evaluación de calidad disponibles en el servicio. El último nivel, *WPS*, gestiona las solicitudes de los clientes mediante la interfaz estándar del OGC WPS. Esta arquitectura se presentó en Ariza et al. (2015).

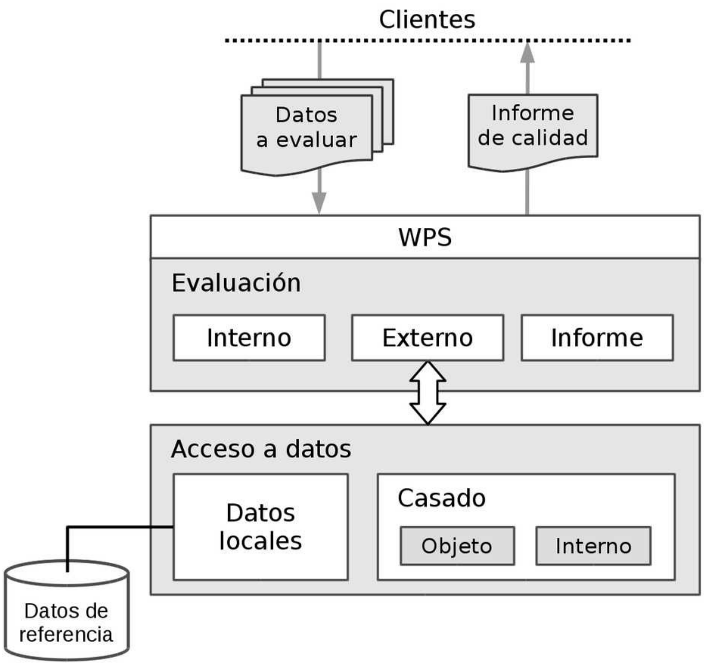


Figure 6.1. Arquitectura propuesta para un servicio de control de calidad.

La capa *Acceso a datos* gestiona a la relación entre los datos de prueba y de referencia. Puesto que los procedimientos de evaluación externa dependen de datos de referencia para la comparación, este nivel proporciona las correspondencias (casados) entre ambos conjuntos de datos para así permitir comparaciones. Hay dos maneras de facilitar los datos de referencia: (1) remoto: tal que quien llama al servicio proporciona los datos de referencia; (2) local: tal que el servicio mismo tiene su propio conjunto de datos de referencia. Esta capa administra el acceso a datos de referencia locales y también ofrece un módulo de *casado* que proporciona las correspondencias entre los datos de prueba y de referencia (locales o remotos). Según lo que se requiera por el método externo, este casado puede ser en el nivel de objeto geográfico, o en el ámbito interno, es decir, teniendo en cuenta los vértices de una geometría.

El casado entre objetos es un requisito de los métodos de evaluación externa para la calidad de datos geoespaciales. En la arquitectura propuesta el módulo de casado desempeña el papel de encontrar las correspondencias entre estos dos conjuntos de datos (referencia y prueba). Estas correspondencias pueden ser a nivel de función (entre objetos), o el nivel interno (entre partes de objetos, por ejemplo, los vértices). Según se analizó en Xavier et al. (2016a), hay una diversidad de soluciones en la literatura para el casado de objetos. Así que decidimos investigar cuáles serían adecuadas a nuestro servicio de control de calidad. Para lograr este objetivo hemos abierto tres frentes de trabajo: (1) el desarrollo de medidas de similitud; (2) preparación de un banco de pruebas de casado; y (3) sobre este banco de pruebas la aplicación de algunos métodos de casado bajo un diseño de experimentos. En cuanto al casado interno, hay pocos métodos en este nivel de actuación, como podemos ver en Xavier et al. (2016a). En este trabajo se propone un nuevo método para casado a nivel interno que está basado en el descriptor de contexto de forma de Belongie et al. (2002).

La capa *Evaluación* contiene las implementaciones de métodos de evaluación directos externos y directos internos. Los métodos directos externos requieren una referencia externa que se maneja en el nivel del Acceso a datos. La capa de Evaluación también contiene el módulo de informe que es responsable de generar el informe de calidad de diferentes maneras: un informe legible, o un informe XML en formato ISO, actual (ISO 2016) o antigo (ISO 2003b). Este nivel representa el núcleo de la arquitectura propuesta para la evaluación de la calidad de datos geoespaciales utilizando servicios web. En este estudio adoptamos la norma brasileña para la calidad de los datos geoespaciales, llamada CQDG (DCT 2016a). Teniendo en cuenta que esta norma proporciona los procedimientos de evaluación de calidad para todos los productos de datos geoespaciales en Brasil, esta norma desempeña el papel de modelo de calidad en este proyecto de investigación. En esta parte de la arquitectura desarrollamos procedimientos de evaluación de la calidad descritos en la norma CQDG para los productos datos geoespaciales vectoriales.

En la arquitectura propuesta, la capa WPS es el punto de contacto con los clientes. Esta capa está encargada de las solicitudes y respuestas usando la interfaz WPS. Los procedimientos de evaluación de la calidad a menudo implican tareas complejas y personas de diferentes organizaciones o departamentos. Ante esta situación tenemos dos principios de diseño: interoperabilidad y simplicidad. El principio de interoperabilidad indica que el nivel WPS debe seguir la especificación WPS y sus esquemas XML con el fin de permitir una forma estandarizada de comunicación. El principio de simplicidad nos lleva a evitar problemas innecesarios en el proceso sí mismo, así que el proceso debe ser tan directo como sea posible. El nivel WPS debe manejar todas las cuestiones de comunicación, procedimientos de validación y tareas de cliente-servidor.

La arquitectura propuesta pretende ser general para la evaluación automática de la calidad y puede aplicarse independientemente de las bases de datos o plataforma de software.

6.2.2. Material

En esta investigación utilizamos *R* como la herramienta informática estadística. R es un lenguaje y un entorno enfocado en herramientas estadísticas y gráficos (R Core Team 2014). Otros materiales relevantes son los datos geoespaciales utilizados para probar el servicio de control de calidad, y el software desarrollado que implementa efectivamente los conceptos propuestos en este estudio.

Se ha trabajado con conjuntos de datos geoespaciales producidos por el IGN y el IECA. Se han utilizado datos de la Base Topográfica Nacional 1:25.000 (BCN25) de la cartografía nacional proporcionados por el Instituto Geográfico Nacional de España

(IGN 2015); y también datos de 1:10.000 de la Base Cartográfica de Andalucía 1:10.000 (BCA10) de la cartografía regional proporcionados por el Instituto de Estadística y Cartografía de Andalucía (ICEA 2015). Hemos seleccionado diferentes paisajes: costa y montaña, rural y urbana. Las siguientes hojas 1:25.000 fueron utilizadas para proporcionar los datos: 0896-3, 0896-4, 1003-4, 0999-1, 0999-2, 0999-3 y 0999-4.

Todo el software desarrollado en este proyecto de investigación se basa en la biblioteca TerraLib. TerraLib es una librería de código abierto desarrollada por el Instituto Brasileño para la Investigación Espacial (INPE) (Câmara et al. 2008), y está disponible en su repositorio (DPI 2013). Dentro de TerraLib existe un subproyecto llamado TerraOGC – una estructura para el desarrollo de Web-GIS que contiene módulos para muchas de las especificaciones del OGC, como GML, WMS, WFS y WCS. Para esta investigación, el módulo WPS existente fue mejorado para atender a los principios de diseño descritos aquí. Como parte de los procesos WPS, se creó un módulo de para procesamiento de la calidad de datos (DQEval) que contiene la mayor parte del código relacionados con este proyecto. Se puede encontrar en línea en su repositorio (DPI 2017).

6.3. Resultados y discusión

Esta sección presenta los experimentos ejecutados con el fin de validar la solución propuesta para la evaluación automática de calidad de datos geoespaciales a través de servicios web. Los ensayos están diseñados para evaluar la solución utilizando datos reales y sintéticos.

El primer experimento se ocupa de la creación de un banco de datos sintéticos de prueba para los métodos de casado. Este banco de pruebas está compuesto por cuatro grupos de conjuntos de datos: (1) grupo inicial: datos originales seleccionados; (2) morfología modificada: datos sintéticos creados con énfasis en alguna clase de morfología específica para las líneas o áreas; (3) perturbación sistemática: datos sintéticos creados a partir de transformaciones afines; y (4) perturbación aleatoria: datos sintéticos creados aleatoriamente sobre la influencia de campos vectoriales de desplazamiento. Creemos que este banco de pruebas es una herramienta valiosa para ser compartida con otros investigadores en el área geomática, así que lo hemos presentado para su publicación a una revista que ofrece un repositorio público de datos científicos para pruebas (Xavier et al. 2017). Los datos generados en este experimento se utilizaron en los experimentos siguientes.

El segundo experimento utiliza los conceptos del diseño de experimentos para comparar un conjunto métodos de casado usando el banco de pruebas desarrollado en el experimento anterior. Este experimento se dividió según las primitivas geométricas: punto, línea y área, en este orden. Cada tipo de geometría tiene sus propios ensayos o configuraciones. El experimento diseñado para métodos de casado está compuesto por

20 ensayos: puntos (P1-P5), líneas (L1-L6) y áreas (A1-A9). Basado en los resultados de estos ensayos fue posible seleccionar algunos métodos de casado con los resultados más adecuados para implementar en nuestra propuesta de servicio de control de calidad.

El tercer experimento testó el nuevo método de casado interno desarrollado en esta investigación. En este estudio estamos adoptando un modelo de calidad de la norma brasileña (DCT 2016a), que describe un procedimiento de calidad posicional basado en puntos. Por lo tanto, hemos desarrollado este método de casado interno con el fin de aumentar la cantidad de puntos para la evaluación de la calidad posicional, ya que permite utilizar objetos del tipo línea o área en el proceso de calidad. Los resultados indicaron que la implementación actual del nuevo método de casado interno tiene dos logros destacados en comparación con otros métodos equivalentes (Fan et al. 2014, Ruiz-Lendínez et al. 2015): (1) este método es capaz de tratar con múltiples pares (m:n) de objetos; y (2) este método puede trabajar con los agujeros de los polígonos lo que puede aumentar la cantidad de partes correspondientes. Los resultados con respecto a partes de líneas no alcanzaron un rendimiento aceptable para la evaluación de la calidad.

En el cuarto experimento probamos la validez de los procedimientos de evaluación de calidad desarrollados en la capa Evaluación de nuestro servicio de control de calidad utilizando los datos generados en el primer experimento. Los experimentos fueron divididos según el elemento de calidad en consideración: consistencia topológica, compleción y exactitud posicional. Los resultados revelaron que los procedimientos de consistencia topológica trabajaban conforme a lo dispuesto en la norma, de manera totalmente automática. En relación a la compleción, verificamos el rendimiento de la evaluación automática en todas las primitivas geométricas (punto, línea, área). Los resultados revelaron que la implementación automática funcionó satisfactoriamente. Sin embargo, hemos identificado que el rendimiento del método de casado seleccionado influye en el rendimiento de la evaluación automática de la calidad. En el ensayo de exactitud posicional verificamos el rendimiento del procedimiento automático para la planimetría en 11 regiones: nueve de puntos y dos de áreas. Los resultados revelaron que el procedimiento de exactitud posicional automático funciona similar al procedimiento manual, resultando la categoría de calidad correcta en todas las regiones consideradas.

En el último experimento comprobamos si la capa WPS era capaz de desempeñar el papel de interoperabilidad entre clientes y procedimientos de evaluación de calidad. En esta fase también se revisaron los posibles reportes de calidad generados. Los resultados indicaron algunos aspectos de la aplicabilidad del WPS como una interfaz de servicio para a la evaluación de la calidad. Podemos destacar: (1) WPS permite múltiples entradas y salidas; (2) WPS está listo para encadenamiento de servicios; y (3) la extensión del proceso es relativamente fácil. Por lo tanto, podemos concluir que la interfaz WPS es una solución factible para implementar el servicio de control de calidad.

6.4. Conclusiones

La industria geomática está viviendo una situación de sobrecarga de datos que plantea nuevos desafíos a los productores de datos cartográficos. Hoy en día es posible encontrar diversos conjuntos de datos que representa la misma extensión geográfica que vienen de productores distintos: voluntarios (por ejemplo, OpenStreetMap), empresas de cartografía y los organismos cartográficos oficiales (en diferentes niveles). Cada uno crea sus bases de datos siguiendo sus propias reglas de adquisición (y fuentes), que nos lleva a la pregunta: '¿cuál base de datos se ajusta a mis propósitos?', una cuestión de aptitud de uso (Servigne et al. 2006).

En este contexto, este estudio se pretendió proporcionar una solución capaz de responder a la pregunta clave: '¿Cuál es la calidad de este conjunto de datos'? Donde los datos espaciales son creados casi automáticamente (por ejemplo, Coumans 2016), también necesitamos una herramienta de evaluación de calidad capaz de responder con la velocidad que se crean esos datos. Por lo tanto, el objetivo principal de este estudio fue desarrollar un servicio web estándar con la capacidad de evaluar la calidad de los conjuntos de datos geoespaciales en modo completamente automático. Para alcanzar este objetivo desarrollamos una solución para la evaluación automática de la calidad de los datos geoespaciales.

Los resultados obtenidos confirmaron que el principal objetivo fue alcanzado: Contamos con un servicio de control de calidad que evalúa automáticamente la calidad de datos geoespaciales con resultados comparables al procedimiento manual. Sin embargo, la solución propuesta tiene sus limitaciones: la capacidad de evaluación de la calidad automática de métodos externos (aquellos que requieren un conjunto de datos externo) está directamente relacionada con el rendimiento de los métodos de casado. Entonces, el rendimiento del servicio de control de calidad automático depende del desempeño de los métodos de casado.

Las contribuciones de este estudio son múltiples. Podemos dividirlas en tres categorías: (1) evaluación de la calidad de datos, (2) generación de datos sintéticos, y (3) casado de datos espaciales. Las contribuciones en relación con la evaluación de la calidad de datos son:

- Procedimiento automático para evaluar la consistencia topológica, confirmando otros estudios (Mobasheri 2013);
- Procedimiento automático para evaluar la compleción basada en los resultados de métodos de casado; y

- Procedimiento automático para evaluar la exactitud posicional para el estándar brasileño usando un nuevo método de casado interno.

Las contribuciones en generación de datos sintéticos son:

- Nuevo método de perturbación sistemática basada en la transformación afín;
- Nuevo método de perturbación aleatoria basada en campos vectoriales de desplazamiento;
- Nuevo método para cambiar la morfología de las líneas;
- Nueva clasificación de la morfología para datos de edificios;
- Nuevo método para cambiar la morfología de las áreas; y
- Nueva clasificación del contexto geográfico usando contexto de la forma y expertos en GIS.

Las contribuciones en métodos de casado son:

- Nuevo diseño experimental para poner a prueba métodos de casado;
- Nueva medida de contexto geográfico basado en el contexto de la forma (Belongie et al. 2002);
- Nuevo método de casado de objetos utilizando la medida de contexto geográfico. Este método demostró ser menos sensible a la perturbación de datos (sistemática o aleatoria) que otros métodos; y
- Nuevo método de casado interno usando esta medida contexto geográfico. Se demostró la viabilidad de esta medida para los datos de áreas.

En un libro clásico de la cartografía, Robinson et al (1995) afirmaron que una de las tareas más difíciles para los cartógrafos es indicar a los usuarios de mapas la calidad de los datos. En la era de la web, o la era de la información (Blinder 2006), esperamos que los cartógrafos pueden delegar esta tarea 'dolorosa' en las máquinas.

REFERENCES

AENOR (2016) *UNE 148002. Información geográfica - Control de calidad posicional de datos espaciales*.

Al-Bakri, M. and Fairbairn, D. (2012) 'Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources', *International Journal of Geographical Information Science*, 26(8), pp. 1437–1456. doi: 10.1080/13658816.2011.636012.

Alba Fernández, M. V. (2013) 'Diseño de experimentos en información geográfica', in Experto Universitario en Gestión de la Información Geográfica (1a Edición). Jaén, España: Universidad de Jaén.

Alt, H. and Godau, M. (1995) 'Computing the Fréchet distance between two polygonal curves', *International Journal of Computational Geometry & Applications*, 5(1–2), pp. 75–91.

Arkin, E. M., Chew, L. P., Huttenlocher, D. P., Kedem, K. and Mitchell, J. S. B. (1991) 'An efficiently computable metric for comparing polygonal shapes', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3), pp. 209–216. Available at: http://doi.ieeecomputersociety.org/10.1109/34.75509.

Ariza López, F. J. and Atkinson Gordo, A. D. (2008) 'Variability of NSSDA estimations', *Journal of Surveying Engineering*, 134(2), pp. 39–44. doi: 10.1061/(ASCE)0733-9453(2008)134:2(39).

Ariza López, F. J. and García Balboa, J. L. (2008) 'Generalization-oriented road line segmentation by means of an artificial neural network applied over a moving window', *Pattern Recognition*, 41, pp. 1610–1626. doi: 10.1016/j.patcog.2007.10.010.

Ariza-López, F. J., Mozas-Calvache, A. T., Ureña-Cámara, M. A., Alba-Fernández, V., García-Balboa, J. L., Rodríguez-Avi, J. and Ruiz-Lendínez, J. J. (2011) 'Influence of sample size on line-based positional assessment methods for road data', *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(5), pp. 708–719. doi: 10.1016/j.isprsjprs.2011.06.003.

Ariza-López, F. J. and Mozas-Calvache, A. T. (2012) 'Comparison of four line-based positional assessment methods by means of synthetic data', *GeoInformatica*, 16(2), pp. 221–243. doi: 10.1007/s10707-011-0130-y.

Ariza-López, F. J. (2013) 'Calidad de la información geográfica: perspectivas de futuro', in *Conferencia Internacional de Geografía y Medio Ambiente*. Ciudad de México.

Ariza-López, F. J. and Rodríguez-Avi, J. (2014) ‘A statistical model inspired by the National Map Accuracy Standard’, *Photogrammetric Engineering & Remote Sensing*, 80(3), pp. 271–281. doi: 10.14358/PERS.80.3.000.

Ariza-López, F. J., Xavier, E. M. A. and Ureña-Cámara, M. A. (2015) ‘Proposal of a web service for positional quality control of spatial data sets’, in *International Workshop on Spatial Data and Map Quality*. Valletta, Malta.

Ariza-López, F. J., Xavier, E. M. A., Chicaiza, E. and Buenaño, X. (2017) ‘Métodos de evaluación de la calidad posicional en Hispanoamérica: análisis de la situación’. In press.

ASPRS (2015) ‘ASPRS positional accuracy standards for digital geospatial data’, *Photogrammetric Engineering & Remote Sensing*, 81(3), pp. A1–A26. doi: 10.14358/PERS.81.3.A1-A26.

Beare, M., Onstein, E., Mäkelä, J., Marttinen, J., Henriksson, R., Jakobsson, A., Tsoulos, L., Williams, F., Meulenaer, L. De, Persson, I. and Kavadas, I. (2010) *ESDIN Quality Final Report - Part A*.

Beaujardière, J. (2006) *OpenGIS® Web Map Server Implementation Specification*. Wayland. Available at: http://portal.opengeospatial.org/files/?artifact_id=14416.

Beeri, C., Kanza, Y., Safra, E. and Sagiv, Y. (2004) ‘Object fusion in geographic information systems’, in *Proceedings of the 30th VLDB Conference*. Toronto, Canada, pp. 816–827.

Belongie, S. and Malik, J. (2000) ‘Matching with shape contexts’, in *IEEE Workshop on Content-Based Access of Image and Video Libraries, CBAIVL 2000*. Hilton Head Island: IEEE, pp. 20–26. doi: 10.1109/IVL.2000.853834.

Belongie, S., Malik, J. and Puzicha, J. (2002) ‘Shape matching and object recognition using shape contexts’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4), pp. 509–522. doi: 10.1109/34.993558.

Blinder, A. S. (2006) ‘Offshoring: the next industrial revolution?’, *Foreign Affairs*, 85(2), pp. 113–128. doi: 10.2307/20031915.

Brasil (1967) ‘Decreto-Lei nº 243, de 28 de fevereiro de 1967’. Brasília: Diário Oficial da União. Available at: http://www.planalto.gov.br/CCIVIL_03/Decreto-Lei/1965-1988/Del0243.htm.

Câmara, G., Vinhas, L., Ferreira, K. R., Queiroz, G. R., Souza, R. C. M., Monteiro, A. M. V., Carvalho, M. T., Casanova, M. A. and Freitas, U. M. (2008) ‘TerraLib: an open source GIS library for large-scale environmental and socio-economic applications’, in Hall, G. B. and Leahy, M. G. (eds) *Open Source Approaches in Spatial Data Handling*. Springer Berlin Heidelberg, pp. 247–270. doi: 10.1007/978-3-540-74831-1_12.

Carpenter, J. and Snell, J. (2013) *Future trends in geospatial information management: the five to ten year vision*. Available at: http://ggim.un.org/docs/meetings/3rd UNCE/UN-GGIM-Future-trends.pdf.

CONCAR (2009) *Perfil de Metadados Geoespaciais do Brasil (Perfil MGB)*. Available at: http://www.inde.gov.br/images/inde/Perfil_MGB_Final_v1_homologado.pdf.

Coumans, F. (2016) 'Ordnance Survey Ireland shows new perspectives', *GIM International*, 30(12), pp. 27–29.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W. and Zook, M. (2013) 'Beyond the geotag: situating "big data" and leveraging the potential of the geoweb', *Cartography and Geographic Information Science*, 40(2), pp. 130–139. doi: 10.1080/15230406.2013.777137.

Crompvoets, J. and Broucker, B. (2015) 'Geospatial information broker. A new role of national mapping agencies', *Micro Macro Mezzo Geo Information*, 4, pp. 1–10.

Deng, M., Li, Z. and Chen, X. (2007) 'Extended Hausdorff distance for spatial objects in GIS', *International Journal of Geographical Information Science*, 21(4), pp. 459–475. doi: 10.1080/13658810601073315.

DCT (2016a) *Norma da especificação técnica para controle de qualidade de dados geoespaciais (ET-CQDG)*. Brasília. Available at: http://www.geoportal.eb.mil.br/images/PDF/ET_CQDG_1a_edicao_2016.pdf.

DCT (2016b) *Norma da especificação técnica para produtos de conjuntos de dados geoespaciais (ET-PCDG)*. Brasília. Available at: http://www.geoportal.eb.mil.br/images/PDF/ET_PCDG_2016_2aEdicao_Aprovada_Publicada_BE_7_16.pdf.

Devogele, T. (2002) 'A new merging process for data integration based on the discrete Fréchet distance', in Richardson, D. E. and van Oosterom, P. (eds) *Advances in spatial data handling*. Springer, pp. 167–181.

Donaubauer, A., Kutzner, T. and Straub, F. (2008) 'Towards a quality aware web processing service', in *Proceedings of the 6th Geographic Information Days*. Münster, Germany, pp. 16–18. Available at: http://www.gi-days.de/archive/2008/downloads/acceptedPapers/Papers/Donaubauer,Kutzner,Straub.pdf.

Douglas, D. H. and Peucker, T. K. (1973) 'Algorithms for the reduction of the number of points required to represent a digitized line or its caricature', *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2), pp. 112–122. doi: 10.3138/FM57-6770-U75U-7727.

ESRI (1998) *ESRI Shapefile Technical Description*. Available at: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf.

Fan, H., Zipf, A., Fu, Q. and Neis, P. (2014) 'Quality assessment for building footprints data on OpenStreetMap', *International Journal of Geographical Information Science*, 28(4), pp. 700–719. doi: 10.1080/13658816.2013.867495.

FEMA (2003) *Guidelines and specifications for flood hazard mapping partners*.

Frank, R. and Ester, M. (2006) 'A quantitative similarity measure for maps', in Riedl, A., Kainz, W., and Elmes, G. A. (eds) *Progress in Spatial Data Handling*. Springer, pp. 435–450.

Friis-Christensen, A., Lucchi, R., Lutz, M. and Ostländer, N. (2009) 'Service chaining architectures for applications implementing distributed geographic information processing', *International Journal of Geographical Information Science*, 23(5), pp. 561–580. doi: 10.1080/13658810802665570.

Frontiera, P., Larson, R. and Radke, J. (2008) 'A comparison of geometric approaches to assessing spatial similarity for GIR', *International Journal of Geographical Information Science*, 22(3), pp. 337–360. doi: 10.1080/13658810701626293.

García Balboa, J. L. (2006) *Automatización de los procesos de segmentación y clasificación de vías de comunicación en generalización cartográfica*. Universidad de Jaén.

Gil de la Vega, P., Ariza-López, F. J. and Mozas-Calvache, A. T. (2016) 'Models for positional accuracy assessment of linear features: 2D and 3D cases', *Survey Review*. Taylor & Francis, 48(350), pp. 347–360. doi: 10.1080/00396265.2015.1113027.

Göesseln, G. Von and Sester, M. (2004) 'Integration of geoscientific data sets and the German digital map using a matching approach', in *Proceedings of XXth ISPRS Congress*. Istanbul, Turkey, pp. 1249–1254.

Goodchild, M. F. (2007) 'Citizens as sensors: the world of volunteered geography', *GeoJournal*, 69(4), pp. 211–221.

Gumiaux, C., Gapais, D. and Brun, J. P. (2003) 'Geostatistics applied to best-fit interpolation of orientation data', *Tectonophysics*, 376(3–4), pp. 241–259. doi: 10.1016/j.tecto.2003.08.008.

Herring, J. R. (2011) *OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*. Wayland. Available at: http://portal.opengeospatial.org/files/?artifact_id=18241.

Huh, Y., Yu, K. and Heo, J. (2011) 'Detecting conjugate-point pairs for map alignment between two polygon datasets', Computers, Environment and Urban Systems. Elsevier Ltd, 35(3), pp. 250–262. doi: 10.1016/j.compenvurbsys.2010.08.001.

ICEA (2015) *Base Cartográfica de Andalucía 1:10.000 - BCA10*. Available at: http://www.juntadeandalucia.es/institutodeestadisticaycartografia/bca/index.htm.

IGN (2015) *Base Topográfica Nacional 1:25.000 (BTN25)*. Available at: http://www.ign.es/ign/layoutIn/actividadesBDGbtn25.do.

ISO (1985) *ISO 2859-2:1985. Sampling procedures for inspection by attributes - Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection*.

ISO (1999) *ISO 2859-1:1999. Sampling procedures for inspection by attributes - Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection*.

ISO (2002) *ISO 19113:2002. Geographic information - Quality principles*.

ISO (2003a) *ISO 19114:2003. Geographic information - Quality evaluation procedures*.

ISO (2003b) *ISO 19115:2003. Geographic information – Metadata*.

ISO (2013) *ISO 19157:2013. Geographic information - Data quality*.

ISO (2014) *ISO 19115-1:2014. Geographic information - Metadata - Part 1: Fundamentals*.

Josefsson, S. (2006) *The base16, base32, and base64 data encodings*. Available at: http://tools.ietf.org/html/rfc4648.

Kim, J. O., Yu, K., Heo, J. and Lee, W. H. (2010) 'A new method for matching objects in two different geospatial datasets based on the geographic context', *Computers & Geosciences*. Elsevier, 36(9), pp. 1115–1122. doi: 10.1016/j.cageo.2010.04.003.

Koukoletsos, T., Haklay, M. and Ellul, C. (2012) 'Assessing data completeness of VGI through an automated matching procedure for linear data', *Transactions in GIS*, 16(4), pp. 477–498. doi: 10.1111/j.1467-9671.2012.01304.x.

Kruse, D. (2014) 'Tools needed for implementing interoperability ELF Geo Tools for schema transformation, data quality validation, generalisation, edge-matching, visualization, change detection, table joining and geo product'. Aalborg, Denmark.

Lacasta, J., Falquet, G., Zarazaga-Soria, F. J. and Nogueras-Iso, J. (2016) 'An automatic method for reporting the quality of thesauri', *Data and Knowledge Engineering*. Elsevier B.V., 104, pp. 1–14. doi: 10.1016/j.datak.2016.05.002.

Li, L. and Goodchild, M. F. (2010) 'Automatically and accurately matching objects in geospatial datasets', in Guilbert, E., Lees, B., and Leung, Y. (eds) *Proceedings of Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science*. Hong Kong, pp. 98–103. Available at: http://www.isprs.org/proceedings/XXXVIII/part2/Papers/51_Paper.pdf.

Li, L. and Goodchild, M. F. (2011) 'An optimisation model for linear feature matching in geographical data conflation', *International Journal of Image and Data Fusion*, 2(4), pp. 309–328. doi: 10.1080/19479832.2011.577458.

MacEachren, A. M. (1985) 'Compactness of geographic shape: comparison and evaluation of measures', *Geografiska Annaler. Series B, Human Geography*, 67(1), pp. 53–67.

Mascret, A., Devogele, T., Le Berre, I. and Hénaff, A. (2006) 'Coastline matching process based on the discrete Fréchet distance', in Riedl, A., Kainz, W., and Elmes, G. A. (eds) *Progress in Spatial Data Handling*. Springer Berlin Heidelberg, pp. 383–400.

Mustière, S. and Devogele, T. (2008) 'Matching networks with different levels of detail', *GeoInformatica*, 12(4), pp. 435–453. doi: 10.1007/s10707-007-0040-1.

Nagel, H. H. and Enkelmann, W. (1986) 'An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(5), pp. 565–593. doi: 10.1109/TPAMI.1986.4767833.

Nebert, D., Whiteside, A. and Vretanos, P. A. (2007) *OpenGIS® Catalogue Services Specification*. Wayland. Available at: http://portal.opengeospatial.org/files/?artifact_id=20555.

Neis, P. and Zielstra, D. (2014) ‘Recent developments and future trends in volunteered geographic information research: the case of OpenStreetMap’, *Future Internet*, 6(1), pp. 76–106. doi: 10.3390/fi6010076.

Mascret, A., Devogele, T., Le Berre, I. and Hénaff, A. (2006) ‘Coastline matching process based on the discrete Fréchet distance’, in Riedl, A., Kainz, W., and Elmes, G. A. (eds) *Progress in Spatial Data Handling*. Springer Berlin Heidelberg, pp. 383–400.

Meek, S., Jackson, M. and Leibovici, D. G. (2016) ‘A BPMN solution for chaining OGC services to quality assure location-based crowdsourced data’, *Computers and Geosciences*, 87, pp. 76–83. doi: 10.1016/j.cageo.2015.12.003.

Mobasheri, A. (2013) ‘Exploring the possibility of semi-automated quality evaluation of spatial datasets in Spatial Data Infrastructure’, *Journal of ICT Research and Applications*, 7(1), pp. 1–14. doi: 10.5614/itbj.ict.res.appl.2013.7.1.1.

Montgomery, D. C. and Runger, G. C. (2003) *Applied statistics and probability for engineers*. 3rd ed. New York: John Wiley & Sons.

Mueller, M. and Pross, B. (2015) *OGC WPS 2.0 interface standard*.

Mustière, S. and Devogele, T. (2008) ‘Matching networks with different levels of detail’, *GeoInformatica*, 12(4), pp. 435–453. doi: 10.1007/s10707-007-0040-1.

Norris, J. (2015) *Future trends in geospatial information management: the five to ten year vision - 2nd edition*. Available at: ggim.un.org/docs/meetings/GGIM5/Future Trends in Geospatial Information Management the five to ten year vision.pdf.

OGC (2014) ‘About OGC’. Available at: http://www.opengeospatial.org/ogc.

Olteanu, A.-M., Mustière, S. and Ruas, A. (2006) ‘Matching imperfect spatial data’, in Caetano, M. and Painho, M. (eds) *Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Lisbon, Portugal, pp. 694–704.

Olteanu-Raimond, A.-M., Mustière, S. and Ruas, A. (2015) ‘Knowledge formalization for vector data matching using belief theory’, *Journal of Spatial Information Science*, 10(2015), pp. 21–46. doi: 10.5311/JOSIS.2015.10.194.

Pendyala, R. M. (2002) *Development of GIS-based conflation tools for data integration and matching*. Available at: http://www.fsutmsonline.net/images/uploads/reports/fdot_bc353_21_rpt.pdf.

Plazanet, C. (1997) ‘Modeling geometry for linear feature generalization’, in Craglia, M. and Coucleis, H. (eds) *Geographic Information Research: Bridging the Atlantic,*. London: Taylor & Francis, pp. 264–279.

Portele, C. (2007) *OpenGIS® Geography Markup Language (GML) Encoding Standard*. Wayland. Available at: http://portal.opengeospatial.org/files/?artifact_id=20509.

Portele, C. (2011) *Final test cases*.

Ramadhani, S., Bennett, R. and Nex, F. (2016) ‘UASs for cadastral applications’, *GIM International*, 30(12), pp. 30–31.

Rodríguez-Avi, J. (2011) 'Estadística', in Experto Universitario en Evaluación de la Información Geográfica (4a Edición). Máster Universitario en Evaluación y Gestión de la Calidad de la Información Geográfica (1a Edición). Jaén, España: Universidad de Jaén.

Ruiz-Lendínez, J. J. (2012) *Automatización del control de calidad posicional de la cartografía*. Universidad de Jaén.

Ruiz-Lendínez, J. J., Ariza-López, F. J. and Ureña-Cámara, M. A. (2016) 'A point-based methodology for the automatic positional accuracy assessment of geospatial databases', *Survey Review*, 48(349), pp. 269–277. doi: 10.1179/1752270615Y.0000000030.

Saalfeld, A. J. (1988) 'Conflation automated map compilation', International Journal of Geographical Information Systems, 2(3), pp. 217–228.

Safra, E., Kanza, Y., Sagiv, Y. and Beeri, C. (2010) 'Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets', International Journal of Geographical Information Science, 24(1), pp. 69–106. doi: 10.1080/13658810802275560.

Samal, A., Seth, S. and Cueto, K. (2004) 'A feature-based approach to conflation of geospatial sources', *International Journal of Geographical Information Science*, 18(5), pp. 459–489. doi: 10.1080/13658810410001658076.

Schut, P. (2007) *OpenGIS® Web Processing Service*. Wayland. Available at: http://portal.opengeospatial.org/files/?artifact_id=24151.

Tagg, C. (2015) 'Marching the step', *Geoconnexion*, 14(8), pp. 51–52. Available at: http://www.geoconnexion.com/articles/marching-the-step/.

Tong, X., Shi, W. and Deng, S. (2009) 'A probability-based multi-measure feature matching method in map conflation', *International Journal of Remote Sensing*, 30(20), pp. 5453–5472. doi: 10.1080/01431160903130986.

Tong, X., Liang, D. and Jin, Y. (2014) 'A linear road object matching method for conflation based on optimization and logistic regression', *International Journal of Geographical Information Science*. Taylor & Francis, 28(4), pp. 824–846. doi: 10.1080/13658816.2013.876501.

Traxler, C. and Hesina, G. (2017) *Interacting with Big Geospatial Data*, *GIM magazine*. Available at: https://www.gim-international.com/content/article/interacting-with-big-geospatial-data.

Tversky, A. (1977) 'Features of similarity', *Psychological Review*, 84(4), pp. 327–352.

USACE (2015) *Engineering and design - photogrammetric and LiDAR mapping*. Washington, USA.

Van Rijsbergen, C. J. (1979) Information retrieval. 2nd ed. Newton: Butterworth-Heinemann. Available at: http://www.dcs.gla.ac.uk/Keith/Preface.html.

Van Wijngaarden, F., van Putten, J., van Oosterom, P. and Uitermark, H. (1997) 'Map integration - update propagation in a multi-source environment', in Proceedings of the

5th ACM international workshop on Advances in geographic information systems. Las Vegas, USA: ACM, pp. 71–76. doi: 10.1145/267825.267844.

Vretanos, P. A. (2010) *OpenGIS Web Feature Service 2.0 Interface Standard*. Available at: http://portal.opengeospatial.org/files/?artifact_id=39967.

Walter, V. (1997) *Zuordnung von raumbezogenen daten - am beispiel der datenmodelle ATKIS und GDF*. Universität Stuttgart.

Walter, V. and Fritsch, D. (1999) 'Matching spatial data sets: a statistical approach', *International Journal of Geographical Information Science*, 13(5), pp. 445–473. doi: 10.1080/136588199241157.

Whiteside, A. and Evans, J. D. (2008) *Web Coverage Service (WCS) Implementation Standard*. Wayland. Available at: http://portal.opengeospatial.org/files/?artifact_id=27297.

Xavier, E. M. A., Meyer, W. and Lunardi, O. A. (2014) 'Banco de dados geográficos do Exército Brasileiro: arquitetura e resultados', in *V Jornadas Ibéricas de Infra-estruturas de Dados Espaciais*. Lisbon, Portugal. Available at: http://www.idee.es/resources/presentaciones/JIIDE14/20141106/BancoDadosGeograficos_presentacion.pdf.

Xavier, E. M. A., Meyer, W. S. and Lunardi, O. A. (2015a) 'Aplicações, tendências e desafios em infraestruturas de dados espaciais', *Bahia Análise & Dados*, 25(4), pp. 699–714.

Xavier, E. M. A., Ariza-López, F. J. and Ureña-Cámara, M. A. (2015b) 'Web service for positional quality assessment: the WPS tier', in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W5*. La Grande Motte, France, pp. 257–262. doi: 10.5194/isprsannals-II-3-W5-257-2015.

Xavier, E. M. A., Ariza-López, F. J. and Ureña-Cámara, M. A. (2015c) 'WPS for positional quality control applying the method proposed in UNE 148002', in *VI Jornadas Ibéricas de Infraestructuras de Datos Espaciales*. Sevilla, Spain.

Xavier, E. M. A., Ariza-López, F. J. and Ureña-Cámara, M. A. (2016a) 'A survey of measures and methods for matching geospatial vector datasets', *ACM Computing Surveys*, 49(2), p. Article 39. doi: 10.1145/2963147.

Xavier, E. M. A., Ariza-López, F. J. and Ureña-Cámara, M. A. (2016b) 'Using third party data to update a reference dataset in a quality evaluation service', in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B4*. Prague, Czech Republic, pp. 55–61. doi: 10.5194/isprsarchives-XLI-B4-55-2016.

Xavier, E. M. A., Ariza-López, F. J. and Ureña-Cámara, M. A. (2017) 'MatchingLand, geospatial data testbed for the assessment of matching methods', in press.

Yang, C., Raskin, R., Goodchild, M. F. and Gahegan, M. (2010) 'Geospatial cyberinfrastructure: past, present and future', *Computers, Environment and Urban Systems*. Elsevier Ltd, 34(4), pp. 264–277. doi: 10.1016/j.compenvurbsys.2010.04.001.

Yang, B., Zhang, Y. and Luan, X. (2013) ‘A probabilistic relaxation approach for matching road networks’, *International Journal of Geographical Information Science*, 27(2), pp. 319–338. doi: 10.1080/13658816.2012.683486.

Yuan, S. and Tao, C. (1999) ‘Development of conflation components’, in *Proceedings of Geoinformatics*. Ann Arbor, USA, pp. 1–13.

Zhang, X., Ai, T., Stoter, J. and Zhao, X. (2014) ‘Data matching of building polygons at multiple map scales improved by contextual information and relaxation’, *ISPRS Journal of Photogrammetry and Remote Sensing*. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS), 92, pp. 147–163. doi: 10.1016/j.isprsjprs.2014.03.010.

Zhao, P., Foerster, T. and Yue, P. (2012) ‘The Geoprocessing Web’, *Computers & Geosciences*, 47, pp. 3–12. doi: 10.1016/j.cageo.2012.04.021.