

Measuring conflation success

Marta Padilla-Ruiz^{*}
Carlos López-Vázquez^{**}

Recibido el 15 de agosto de 2016; aceptado el 24 de enero de 2017

Resumen

Nos encontramos inmersos en la era del *Big Data*, en la cual existe una enorme cantidad de datos heterogéneos en escala de tiempo y espacio. Estos datos comienzan a ser transmitidos en tiempo real desde diferentes dispositivos y sensores desplegados en el nuevo concepto de ciudades inteligentes. Los procesos de confluación juegan un rol importante en este escenario, definidos como el procedimiento para la combinación e integración de diferentes fuentes de datos, consiguiendo así mayor riqueza de información en el resultado. Estos procesos permiten la actualización de bases de datos geográficas (BDG) a partir de diferentes fuentes de datos, en la que una de ellas es más precisa, exacta o vigente que la otra. El proceso seguido para la confluación geométrica requiere transformar elementos geográficos de un conjunto de datos a otro, de forma que se minimicen las discrepancias entre ellos. La exactitud debe ser tomada en cuenta en este proceso y los resultados deben ser medidos y evaluados con el objetivo de estimar la calidad del producto. En este artículo se describe el proceso de evaluación de la confluación, junto con las diferentes métricas y estrategias seguidas para evaluar lo adecuado del proceso.

Palabras clave: *confluación, fusión de datos, integración de datos, exactitud espacial, éxito de la confluación.*

Resumo

Nos encontramos imersos na era do *Big Data*, na qual existe uma enorme quantidade de dados heterogêneos na escala de tempo e espaço. Estes dados começam a ser transmitidos em tempo real desde diferentes dispositivos e sensores empregados no novo conceito de cidades inteligentes. Os processos de confluência jogam um rol

^{*} Campus SUR de la Universidad Politécnica de Madrid (UPM), 28031 Madrid, España, correo electrónico: marta.padilla.ruiz@alumnos.upm.es

^{**} Laboratorio LatinGEO, SGM+ORT, Facultad de Ingeniería, Universidad ORT, Uruguay, correos electrónicos: carloslopez@uni.ort.edu.uy; carlos.lopez@ieee.org

importante neste cenário, definidos como o procedimento para a combinação e integração de diferentes fontes de dados, conseguindo assim maior riqueza de informação no resultado. Estes processos permitem a atualização de bases de dados geográficas (BDG) a partir de diferentes fontes de dados, onde uma delas é mais precisa, exata ou atual que a outra. O processo seguido para a conflagação geométrica requer transformar elementos geográficos de um conjunto de dados a outro, de forma que se minimizem as discrepâncias entre eles. A exatidão deve ser tomada em conta neste processo e os resultados devem ser medidos e avaliados com o objetivo de estimar a qualidade do produto. Neste artigo se descreve o processo de avaliação da conflagação, junto com as diferentes métricas e estratégias seguidas para avaliar o adequado do processo.

Palavras chave: *Conflagação, Fusão de Dados, Integração de Dados, Exatidão Espacial, Êxito da Conflagação.*

Abstract

We are immersed in the Big Data era, where there is a large amount of heterogeneous data, both in time and spatial scales. This data starts to be streamed in real time from different devices and sensors, well illustrated by the new concept of Smart Cities. Conflation processes play an important role in this scenario, defined as the procedure for the combination and integration of different data sources, improving the level of information of the result. It also allows to update geographical databases (GDB), conflating different kind of sources where one of them is more accurate or updated than the other. Regarding geometric conflation, the procedure involves transforming features from one data source to another, minimizing the geometric discrepancies between them. Accuracy has to be taken into account in these processes, and the results need to be measured and evaluated in order to have a better understanding of product quality. In this paper, conflation evaluation process is described along with the different metrics and approaches to assess its accuracy.

Key words: *Conflation, Data fusion, Data integration, Spatial Accuracy, Conflation Success.*

Introduction

The term Smart Cities pictures the city as a constellation of instruments, such as devices or sensors, that are inter-connected through multiple networks. In this new concept, huge amounts of data start to be streamed in real time and a new urban environment is being generated, quite different as anything we have experienced before (Batty *et al.*, 2012). This scenario is important to geographic information as it increases the data available to make maps and could allow to update them in real time.

Therefore, a Web of Sensors is what the Smart Cities are developing, and it could represent an architecture for pervasive spatial data. The desirable scenario would be an interoperable system (for different data, sources, sensors and devices) with all of this data accessible to everyone and ready to use.

This new challenge makes the combination of diverse information necessary. There is a large amount of heterogeneous data, sometimes duplicated from different sources, with different time and spatial scales and different accuracy. This heterogeneous lineage results in geometric, topological and semantic discrepancies among them (Casado, 2006).

Conflation processes deals with these problems, defined by Cobb *et al.* (1998, 2000) and Edwards and Simpson (2002) as the procedure for the combination and integration of different data sources, improving the level of information of the result. Following the definition, it is clear that conflation processes could play an important role in the situation described above. They allow to obtain conflated datasets from heterogeneous sources, covering the same geographical zone and describing the same information, even if they do not share the same density and accuracy.

Regarding conflation classification, Ariza *et al.* (2011), proposes a classification distinguishing between four different criteria:

- According to the criteria used to match the objects (Casado, 2006): geometric conflation deals with the problem of transforming features from one map to another (target map), minimizing the geometric differences between them. Semantic conflation deals with differences in meaning of terms and aims to the homogenization of existing feature classes in a map. Topological conflation deals with topological relationships, which need to be preserved and have to be regenerated in case there is joining, merging or disappearance of features.
- According to the representation model used: between vectors, between vector and raster, and between rasters.
- According to the categorization problem: vertical conflation and horizontal conflation (Yuan and Tao, 1999), and temporal conflation (Ariza *et al.*, 2011). Vertical conflation refers to datasets that occupy the same geographical region. Horizontal conflation is related to detect the differences between the common boundaries of adjacent datasets. Temporal conflation refers to datasets that occupy the same geographical zone at two different points in time.
- According to the level of automatization (Lemarie and Raynal, 1996): automatic, semiautomatic or completely manual.

Conflation processes could provide new information that was not there when the datasets were separated. However, a third dataset is not generated without problems

or errors (Saalfeld, 1983). As in any other process, the results need to be measured and evaluated, in order to have a better understanding of product quality.

This paper deals with the evaluation of conflation processes and its metrics. Traditionally, conflation has been performed and evaluated subjectively by cartographic experts. In particular, there exist a number of different metrics found in the literature devoted measuring conflation success. Those metrics are objective, independent of the user context or the conflation purpose, and global, taking the value obtained as representative for the entire region. Most of the metrics do not consider the intended end user application, thus they are solely related to the dataset. A new term, Ephemeral Conflation, is arising, where user context and purpose are being taken into account. In an ephemeral use of data, the user is seeking an utilitarian end, solving a particular query or quickly visualizing without having the intention of storing the information (González *et al.*, 2013). This scenario is the most likely when an user is in a Smart City, engaging with all the connected devices around him/her, wanting the information they need on the move and at the right time and place. In this case, the evaluation of the conflation processes should be assessed with a different criterion, and the success metrics needs to reflect the non-permanent use aspect of the product, the user context and subjectivity.

This paper is organized as follows. The workflow for a conflation evaluation is described in section 2. Section 3 presents the new concept of ephemeral conflation and discuss about weak and strong metrics. Section 4 deals with the strong metrics to assess conflation success and at the end of the paper, the conclusions and references are presented.

Conflation evaluation

Ariza *et al.* (2011) proposed a conceptual model for a general conflation process. First step involves preprocessing, testing if two datasets have the same format, scale, cartographic projection and reference system. Second step determine the homologous elements of both datasets, using semantic filters and ontologies as debug operators. Once homologous elements are obtained, it is possible to evaluate the differences. Then, the most appropriate adjustments for the transformation are established and after erasing the possible differences, both datasets can be matched.

Saalfeld (1983), distinguishes two steps for the geometric conflation procedure: (a) to identify homologue objects in both cartographies, and record its coordinates; and (b) to find and apply an adequate numerical transformation to objects in A trying to fit the corresponding ones in B.

These general workflows do not include the conflation evaluation stage. However, it is needed to measure the conflation success in order to evaluate the results of the process, compare different transformation functions, algorithms, methods, or as a quality metric of the resulting product. This could help to select the best meth-

od to perform a conflation process given a defined scenario. The ordering criterion among feasible transformation methods is a byproduct related to the metric of success (López-Vázquez *et al.*, 2009).

A simplified workflow for a conflation process based on the reviewed literature but adding the evaluation stage, is described in Figure 1. It requires: (a) to have two different datasets (A and B), (b) to extract the features in both sources, (c) to perform a matching between them, identifying the homologous objects in both datasets and recording its properties and/or coordinates, (d) to select a transformation function, (e) to apply the transformation to fit objects in A to the corresponding ones in B and (f) to apply an evaluation metric to assess success.

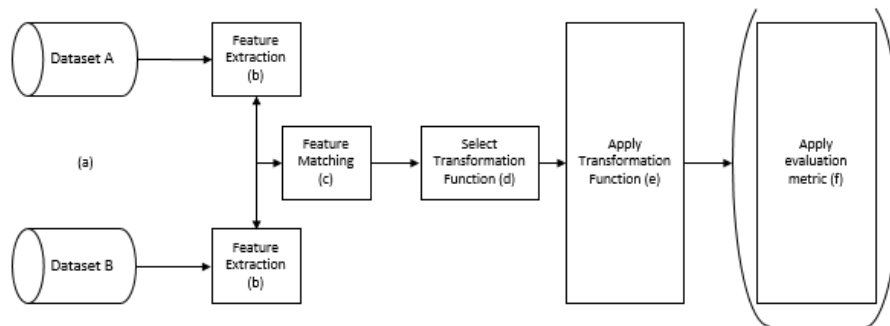


Figure 1. Workflow in a Geometric conflation process: (a) two different datasets, (b) feature extraction, (c) feature matching, (d) Transformation Function selection, (e) transformation function application and (f) Apply evaluation metric. Source: compiled by the authors based on the literature.

The conflation evaluation takes part in step (f) and it is optional in a conflation process. However, it is always highly recommendable to evaluate the conflation product.

Ephemeral conflation: introducing weak metrics

Metrics to measure conflation success are usually objective and quantitative, obtained as a result of field work and computation and performed by cartographic experts. However, this is not the only possibility to measure conflation success. González *et al.* (2013) describe what would be an “ephemeral use of data”: the user could only seek an utilitarian or immediate end, solving a particular query or quickly visualizing or displayed on a screen. Despite they illustrate the concept with geometric conflation, its implications are more general, being the most likely scenario in a smart city when the user is engaging with all the connected devices

around him/her, wanting the information they need on the move and at the right time and place.

This new perspective changes what quality and accuracy means and, consequently, their evaluations and metrics. In an ephemeral conflation context, the positional accuracy is not that important provided the utilitarian or immediate end of the user have been accomplished. That means a road enriched with directions information with 10 meters of error in a conflation process could help an user to find his way to an airport, although the positional accuracy is not high. The information speed is more important than the result “quality”.

In this context, they define a new categorization for metrics:

- Strong metric. objective statistics based on traditional accuracy: Root Mean Square Error (RMSE), the Fréchet distance (Aronov *et al.*, 2006), the Hausdorff distance (Knauer *et al.*, 2009) as well as others.
- Weak metric. Any metric characterizing ephemeral conflation, considering the context and the user.

They remark that the visual impact of conflation processes is not always directly related to the value of a strong metric. Depending on the conflation goal, the author argues that alternative metrics deserve further consideration. They do not offer a computable procedure to evaluate such weak metric so we will concentrate on the strong metrics.

Strong metrics to assess conflation success

The most common way to measure geometric conflation success in geographical data (coordinates) is the RMSE, as prescribed for example in the NSSDA Standard (hereinafter FGDC, 1998). However, there are other dimensions to consider as well as ways to assess conflation success. In this section, a review of the metrics to assess conflation success are presented.

Metrics as a percentage

Completeness and correctness

Completeness and correctness are two quality dimensions described by Wiedemann *et al.* (1998) barely considered in practical quality reports. For them, percentage metrics were used to report conflation success. They can be applied to point features (Chen *et al.*, 2004), line features (Zhang *et al.*, 2016, Chen *et al.*, 2006, Chen *et al.*, 2008, Yang *et al.*, 2012) and polygon features (Sledge *et al.*, 2011). As usual, in order to apply these metrics, it is needed to have a ground truth set of features (reference features) to which the conflated dataset will be compared. They might be

points with known coordinates, manually extracted features from an image (if lines or polygons are considered), or even place names.

Chen *et al.* (2004) described an information integration approach that uses common vector datasets to automatically conflate imagery with street maps. They conducted an experiment using two test dataset of freely available street maps to evaluate their approach, applying correctness metric as the percentage of the tested maps whose point pattern aligns to the corresponding point pattern on the imagery (taken as a reference). Geometry is not considered and the metric is applied just to point features. As a result for the experiment, 87.1% of the tested maps accurately aligned while 12.9% of the maps misaligned with the image.

Sledge *et al.* (2011) presented a vector-to-imagery building conflation system intended to align building outlines to buildings automatically extracted from imagery, in order to improve their geometric accuracy. To evaluate the performance of the system, they select 20 imagery-tiles of size 350 m² with a resolution of 0.6-m per pixel and vector GIS outlines. As a reference features, ideal representations of buildings were manually extracted. Instead of using a traditional geometric metric, they use the conflation correctness score for each scene as a percentage of the number of correct building correspondences found versus the total number of correct correspondences. The total numbers of correspondences are the total aligned buildings found by the system, while the correct building correspondences are the ones where the buildings align with the reference buildings extracted by hand. In this case, the metric is applied to polygon features. The results were presented as a summary for the 20 test areas showing the displacement between the buildings outlines and the reference, whose high values evidence that the original buildings have poor positional accuracy. The results for correctness are high, around 90-100%. However, some scores are lower (around 80-85%), due to a segmentation fail to find buildings, hidden by trees or shades, which is another issue to take into account when working with automatically vector extractions.

Zhang *et al.* (2016) performs a road network conflation process to allow multi-modal navigation. Their process is shown in Figure 2 and involves five steps: (1) road-network matching between datasets; (2) identification of the pedestrian ways; (3) geometric transformation of the pedestrian ways to eliminate geometric inconsistency; (4) topologic and semantic remodeling of the conflated road network, creating new intersections, transferring semantic information and generating unique IDs for each geographic entity; and (5) error checking and correction.

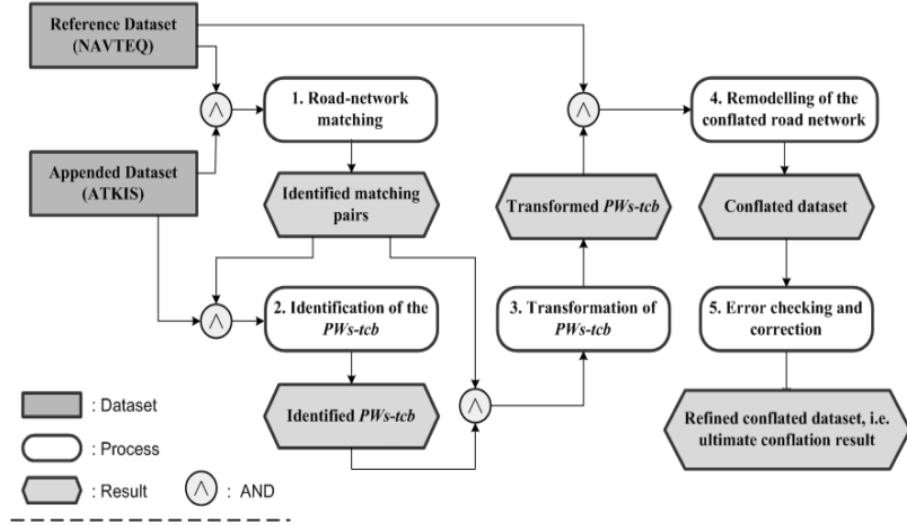


Figure 2. Steps taken to perform a road network conflation process to allow multi-modal navigation. From Zhang *et al.* (2016).

In step 5, the conflated pedestrian ways are classified into four categories: duplicated conflated pedestrian ways, partial duplications, conflated pedestrian ways that are possibly wrong and reliable conflated pedestrian ways (which are the ones not belonging to any of the other categories).

After the process, an evaluation for the automatic conflation approach is performed, selecting three examples of NAVTEQ roads with additional ATKIS pedestrian ways in Germany. Automatic results are compared with manually produced ones, computing two different metrics: overall correctness and conflation correctness, as described below:

$$\text{Overall Correctness} = \frac{AF - UCF}{AF} \times 100 \quad (1)$$

$$\text{Conflated Correctness} = \frac{CF - UCF}{CF} \times 100 \quad (2)$$

Where:

- AF is the number of ATKIS features.
- CF is the number of the conflated features.
- UCF is the number of mismatched conflated features.

In this case they use the number of road features to compute the metric of success. There are a total of 20,285 NAVTEQ features (reference) and 31,112 ATKIS features. After the conflation process, 10,022 ATKIS features were successfully conflated, while only 65 features were unfavorable, which means they were conflated either inaccurately or unnecessarily. Results for both metrics are high, achieving 99.79% for overall correctness and 99.35% for conflation correctness.

Chen *et al.* (2006) performs an automatic map-imagery geometric conflation using a road network. They developed a new metric called Buffer Method (Figure 3) to measure how well the features (roads) on the map align to the corresponding features (roads) on the imagery. The metric is based on a methodology to evaluate automatically road extraction from imagery (Wiedemann *et al.*, 1998). In the evaluation process, they compare this network with a reference one, extracted from manually plotted road axes (segments) and road sides. To understand the metrics applied within the Buffer Method context, it is needed to explain what they consider as matched features:

- Both networks are split into short pieces of equal length.
- A constant predefined road-width is constructed around the reference road network.
- Every portion of the conflated road network within the buffer width from the reference road network is considered as matched.
- The direction difference between matched road axis and reference road axis must be less than a pre-defined threshold d .

The pieces of the conflated roads within the buffer to the reference roads with consistent direction are considered as matched. The drawback of this procedure is that the performance is highly affected by the predefined constant buffer width.

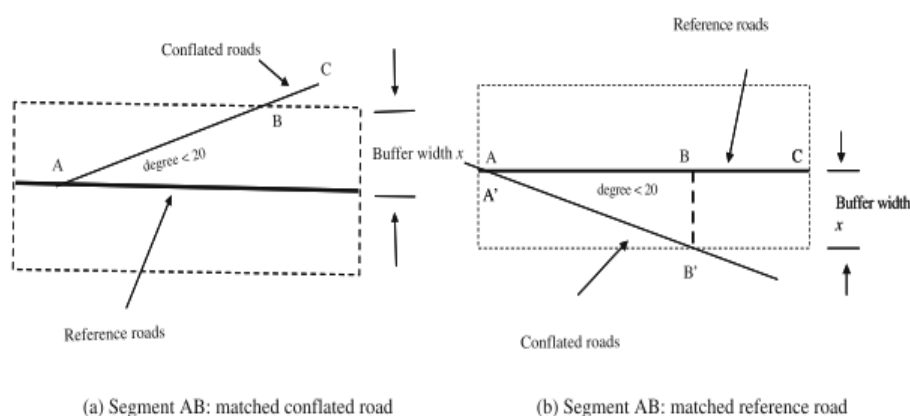


Figure 3. Buffer method for evaluating completeness and correctness. From Chen *et al.* (2006).

Taking this into account, they apply completeness and correctness as a metric of success, defined as follows:

$$\text{Correctness} = \frac{\text{Length(MCR)}}{\text{Length(CR)}} \quad (3)$$

Where:

- CR is the set of road segments that compose the conflated map road network.
- MCR is the set of conflated map road segments that can be matched to the corresponding reference road segments.

$$\text{Completeness} = \frac{\text{Length(MRR)}}{\text{Length(RR)}} \quad (4)$$

Where:

- RR is the set of road segments that compose the reference road network.
- MRR is the set of reference road segments that can be matched to the corresponding conflated map road segments.

Basically, correctness is the percentage of correctly conflated roads with respect to the total conflated roads and completeness corresponds to the percentage of the reference features for which the conflation process is successful, meaning what is missing if the reference features are replaced with the conflated ones.

In this case, one of the differences with Zhang *et al.* (2016) is that they compute these metrics using the length of the roads instead of counting the features. They conduct several experiments, using two different datasets: orthoimagery (0.3 m/pixel and 1 m/pixel resolution) and vector data. Four test areas were considered. To avoid the drawback of the method, instead of using a constant buffer width for each road segment, the real road widths in the imagery are used as the buffer width. Hence, the roads with different widths have different buffer widths. They calculated completeness and correctness for both original and conflated features for each vector dataset respectively, using a 3D graph to report the results, as shown in Figure 4. The values are grouped by tested areas in the X-axis, percentage values are shown in the Y-axis and the third variable (color) shows if the lines are the originals or the conflated ones.

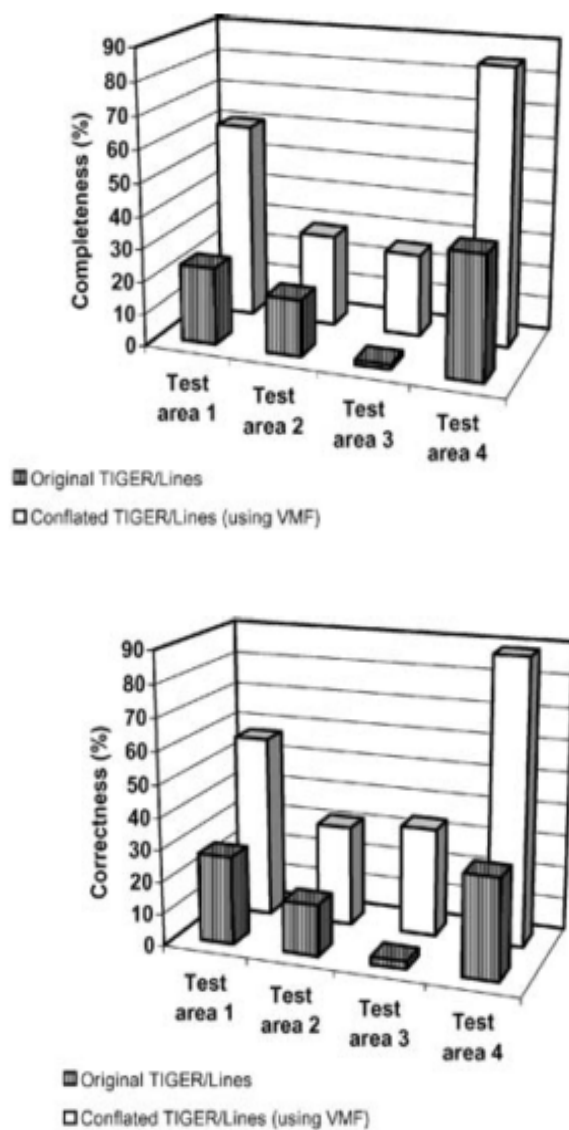


Figure 4. Completeness and correctness assessment as reported by Chen *et al.* (2006).

Therefore, they report a percentage of change from original to conflated features. All percentages result as an improvement, varying from 10% to 80% for both completeness and correctness.

The methodology and metrics described by Chen *et al.* (2006) are found also in Chen *et al.* (2008) and Yang *et al.* (2012). Chen *et al.* (2008) performed a raster map-orthoimagery conflation and the metrics are applied to vectorized map road pixels. Yang *et al.* (2012) develop an algorithm intended to improve the quality of hybrid maps by removing shape disagreements and spatial inconsistencies that arise during conflation of road maps with aerial images. They use road intersections as the corresponding features in the two datasets, applying correctness metric.

Wiedemann *et al.*, (1998) described another metric (that they called quality) expressed as a combination of completeness and correctness into a single measure.

$$\text{quality} = \frac{\text{Completeness} * \text{Correctness}}{\text{Completeness} - \text{Completeness} * \text{Correctness} + \text{correctness}} \quad (5)$$

Quality is a more general measure of the final result where completeness and correctness are combined into a single measure.

Yang and Zhang (2015) proposes an approach to conflate crowdsourced road networks with points of interest (POIs), based on pattern mining. Geometric patterns represent the geometric characteristics (e.g., shape, connectedness, density, or distribution) repeated with sufficient regularity within an object or between objects (Mackaness and Edwards, 2002). They conduct an experiment with two data sets. The road networks were retrieved from OSM and the POIs from VGI data and professional agencies. To evaluate the geometric adjustment, they compare the spatial relations of the POIs and road networks with map services websites. They select 10% of the POIs near the road segments and manually check if the spatial relations are consistent between them, as shown in Figure 5. If they are consistent, they are considered as correct. If an inconsistent spatial relation or large discrepancy occurs to one POI, it is marked as incorrect. Finally, if no corresponding records with the name of POI are found in the map services sites, the POI is marked as uncertain. They count the total number of correct, incorrect and uncertain points.



Figure 5 . Manual inspection of the spatial relations between POIs and road networks as performed by Yang and Zhang (2015).

Only for the correct number of points, they give a percentage related to the total points selected. Therefore, overall, this metric only reports a percentage of correct points over only 10% of the POIs near the road segment.

Improvement percentage

Doytsher *et al.* (2001) developed a linear-based map conflation algorithm. They applied the algorithm using target datasets (i.e., more accurate), composed of roads and hydrologic features extracted from photogrammetry as well as source datasets obtained from 1:50,000 topographic maps. Additional objects were extracted from the photogrammetric data as test objects. They calculated global statistical measures as the mean distances between the positions of the source and target objects. Comparing the mean distance differences between the original dataset and the conflated dataset, a percentage is calculated as an improvement in positional accuracy, as described below.

$$\text{Improvement \%} = \frac{(\bar{d}_o - \bar{d}_c) * 100}{\bar{d}_o} \quad (6)$$

Where:

- \bar{d}_o = mean difference between target and original (meters)
- \bar{d}_c = mean difference between target and conflated (meters)

Area increase ratio

Touya *et al.*, (2013) proposes a framework based on least squares adjustment that provides a geometrical conflation which preserves the characteristic shapes of geographic data. Their process involved different steps: (a) Defining constraints to preserve shape, (b) conflating data and maintaining data consistency, (c) computing the constraints on each point and (d) transforming the constraints into a linear equation system which is solved using least square adjustment. They evaluate the framework conducting an experiment to conflate two datasets: a very accurate one, containing city limits and network information like roads, paths and rivers; and a less accurate dataset, containing city limits, cadastral land use parcels and a building layer. They compare the results obtained, to the results applying rubber sheeting method on the same test case, measuring the shape preservation, which is defined as the shape of the conflated features compared to the shape of the corresponding initial features. They use five different metrics, computing RMSE errors for 200 features and comparing the Least Squares conflation and the Rubber sheeting conflation with the initial data. One of the metrics used is the Area Increase Ratio, which measures as a percentage if the conflation increased or decreased the feature area. The results show that the geometrical conflation adjusted by least squares preserves shapes better than using rubber sheeting method, having a less area increase ratio.

Metrics as a length

Conflation success usually involves datasets with different geometric accuracy. It is fit to use metrics related with geometry, which can be expressed as a length.

RMSE

Geometric conflation is usually characterized by using a traditional accuracy metric, the RSME described by FGDC (1988) in their National Standard for Spatial Accuracy (NSSDA). RMSE is the square root of the average of the set of squared differences between dataset coordinate values and coordinate values from an independent source of higher accuracy for homologous points. This means that this standard assumes that there exist reference data more accurate than the dataset.

Accuracy is reported in ground distances at the 95% confidence level, which means that 95% of the positions in the dataset will have an error with respect to true ground position that is equal to or smaller than the reported accuracy value. This value reflects all uncertainties, including those introduced by geodetic control coordinates, compilation, and final computation of ground coordinate values in the product (FGDC, 1998). Horizontal accuracy is defined as follow:

$$RSME_r = \sqrt{RSME_x^2 + RSME_y^2} \quad (7)$$

Where:

$$RSME_x = \sqrt{\sum \frac{(x_{data} - x_{check})^2}{n}} \quad (8)$$

$$RSME_y = \sqrt{\sum \frac{(y_{data} - y_{check})^2}{n}} \quad (9)$$

- x, y_{data} = dataset coordinates values
- x, y_{check} = independent source of higher accuracy coordinates values
- n = number of control points

There are two different cases for computing accuracy according to the NSSDA:

- Case 1:

$$\begin{aligned} RSME_x &= RSME_y \\ Accuracy_r &= 1.7308 * RSME_r \end{aligned} \quad (10)$$

- Case 2:

$$\begin{aligned} RSME_x &\neq RSME_y \\ Accuracy_r &\sim 2.4477 * (RSME_x + RSME_y) \end{aligned} \quad (11)$$

Using RMSE as defined above, implies to use an accuracy standard to measure conflation success. The issue with this application is that in the conflation case, there is not always reference data available. In our setting, the goal is to measure the discrepancies between two datasets after the conflation process but one is not necessarily highly more accurate than the other. However, the standard is used to measure conflation success since it fits to the problem.

As an example for applying NSSDA statistics to a conflation process, López-Vázquez *et al.* (2009) used RMSE as a metric of success, applying it to different geometric conflation algorithms to data from an urban area. Their goal was to compare different conflation methods. They used orthorectified images of higher accuracy as a ground reference. To simulate a realistic situation, 20 control points

out of 100 available are selected in each evaluation: 8 being fixed (defining the convex hull of the data) and 12 being random. The remaining 80 points are used as witnesses of the accuracy of the transformation, since their displacements are also known. Four different methods were considered: a) Ordinary kriging (Samper and Carrera, 1990) b) Inverse Distance Weighting c) GRIDDATA (Matlab, 2009) and d) GRIDFIT (D'Errico, 2006). Five runs of the calculations were performed. The results show that some methods perform better than others in some cases but not in all of them, while some methods were systematically worse. The author stated that is needed more research to improve confidence.

Yang *et al.* (2012) also applies RMSE as a metric of success for a conflation process. They compare the results of two different approaches, computing RMSE in four different samples. The original dataset coordinate values are compared to the original ones and the results of using both approaches, being RMSE the square root of the distance from the location of the references to the location of the originals and aligned using both approaches. In terms of the overall results for the four sample sites, the RMSE improved from 14.2 to 9.3 m.

Displacement

This metric is applied to linear (Chen *et al.*, 2006) or point features (Chen *et al.*, 2004) and they also need a reference to compare with. In these cases, conflation success is expressed as a length representing the positional accuracy and meaning how far are the conflated features on the map from the real (reference) features.

In Chen *et al.* (2004) a quantitative analysis is conducted, using the road intersections (points) instead of the whole lines (roads). The displacement of the road intersections to the corresponding road intersections in the imagery is measured in meters, and the mean of these point displacements are used to evaluate the accuracy of the considered algorithms. For test dataset 1, 281 intersections were used and the mean point displacement improves from 27 to 8.35 meters; while in test dataset 2, with 240 intersections, the mean point displacement improves from 24 to 10.9 meters.

Linear features can also be considered. In Chen *et al.* (2006) they calculate these displacements considering a buffer around the reference line axes and computing the proportion of the conflated map line pixels that lies within the buffer (Figure 6). They state that RMS difference quality measure as described in Wiedemann *et al.* (1998) does not meet the requirements needed to compute how far the conflated road network is from the reference road network, because it only allows to measure how far the matched conflated road network is from the reference road network. Displacement is described as the percentage of the total length of the conflated map lines that is within a specific distance of the reference lines. (e.g., 95% of the conflated lines are within 5 meters of the reference lines).

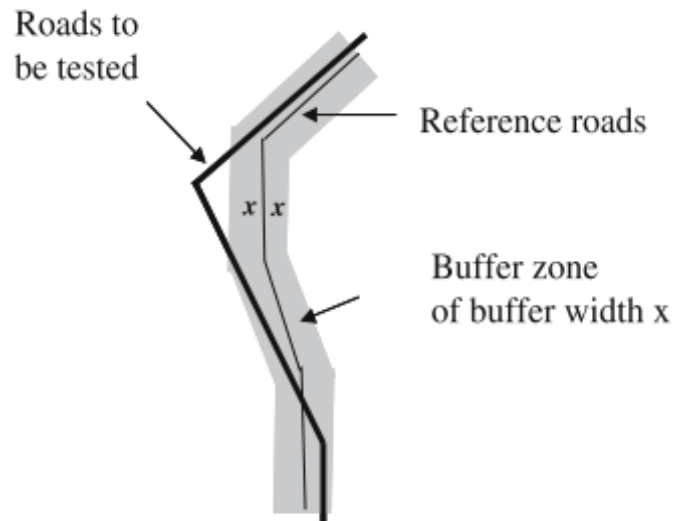


Figure 6. Positional accuracy evaluation using buffers. From Chen *et al.* (2006).

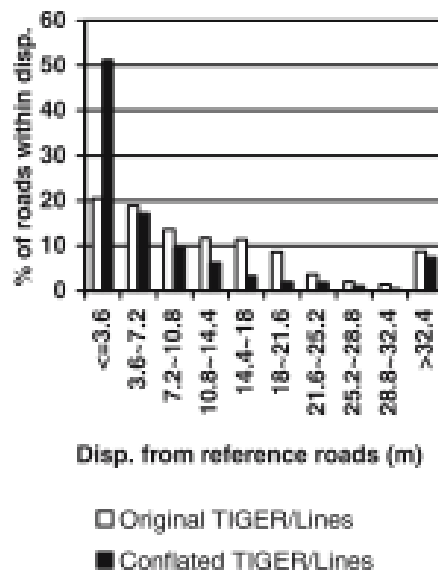


Figure 7. Positional accuracy assessment. Displacement values are grouped every 3.6 meters in the x-axis and the y-axis shows the percentage of conflated roads which have that displacement tolerance. From Chen *et al.* (2006).

They gradually increase the buffer-width, varying from 3.6 m (i.e., the U.S. standard lane width) to 32.4 m (i.e., nine times the U.S. standard lane width). Therefore, instead of computing a number, the positional accuracy is reported as a function. The x-axis shows the displacement values grouped every 3.6 meters and the y axis shows the percentage of conflated roads lying within the displacement range presented by the x axis. Color is used to distinguish between original and conflated lines. An example is given in Figure 7, where 50% of the conflated roads have a displacement value less than 3.6 meters, 18% between 3.6 and 7.2 meters, 10% between 7.2 and 10.8 meters and so on.

Hausdorff distance

Another metric used by Touya *et al.*, (2013) in their geometric conflation adjusted by least squares was Hausdorff distance. The Hausdorff distance measures how far two subsets of a metric space are from each other, represented as the greatest of all the distances from a point in one set to the closest point in the other set. It can be used, among others, to measure the similarity of contours (Abbas, 1994). As the same case with area increase ratio, they computed RMSE errors for 200 features, comparing The Least Squares conflation and the Rubber sheeting conflation with the initial data. The results show that the geometrical conflation adjusted by least squares preserves shapes better than using rubber sheeting method, being the Hausdorff distance shorter.

Other length metrics

In Doytsher *et al.* (2001), as explained before, mean distance differences between the original dataset and the conflated dataset are calculated and global statistical measures are computed from these differences:

- Mean: average of the set of distance differences. Particularly, the number obtained as the sum of the differences divided by the total number of differences calculated.
- Maximum and minimum: highest and lowest value of the differences in absolute value.
- Standard deviation. The standard deviation is defined as the positive square root of the differences between the distances and the mean distance, expressing the deviations of the differences from the mean.

After calculating global statistical measures, they apply the same procedure to specific examples for each kind of feature (roads and hydrology), for the original and the conflated dataset.

Metric as time

Execution time always need to be evaluated in the case of automatic conflation, since it could represent a way to measure the algorithm quality or performance. Moreover, in an ephemeral context, execution time evaluations are decisive since it represents the time the user will wait till the conflation process is finalized. However, in this case, the metric does not measure the product quality but the process.

In some cases, the running time of the conflation algorithms is mainly dominated by the feature matching routine (third step in a conflation process, Figure 1), and the transformation function applied (fifth step in a conflation process, Figure 1) is not relevant in this computation time. This running time mainly depends on the number of features, not on the maps sizes or map scales. Therefore, the running time of the point matching routine can be used as the overall execution time, gradually increasing the number of points on the imagery (Chen *et al.*, 2004, Chen *et al.*, 2006, Chen *et al.*, 2008).

On the other hand, conflation process for which the transformation function is computationally intense, including the required matching process in the execution time does not drastically change the result. The processing time is negligible compared to conflation computation time (Touya *et al.*, 2013).

There are also examples that compute the execution time including the whole process: matching, transformation and data reading and writing (Zhang *et al.*, 2016).

Other possibilities for metrics

Other metrics to measure shape preservation used by Touya *et al.* (2013), in their geometric conflation adjusted by least squares, were not expressed as a percentage, length or time:

- Surface distance (Figure 8). It measures the similarity of shapes, described as the ratio between the areas of the intersection of two polygons and their union.
- Turning function. It measures the preservation of angles independently to translations, as described in Arkin *et al.* (1991).
- Polygon signature. It measures the similarity of shapes independently to translations, as described in Vauglin *et al.* (1998).

They computed RMSE errors also for these metrics, comparing the Least Squares conflation and the Rubber sheeting conflation with the initial data. The results show that the geometrical conflation adjusted by least squares preserves shapes better than using rubber sheeting method in all the cases: turning function distance is twice smaller in RMSE error in the case of Least Squares adjustment, which means that angles are better preserved; polygon signature is smaller too,

which means it increases less the area of the features; however, surface distance is higher, due to the translation implied by the shape preservation constraints.

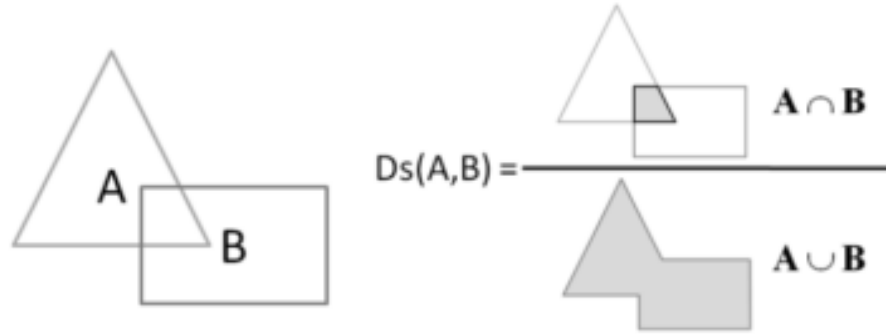


Figure 8. Definition of surface distance. From Touya *et al.* (2013).

Conclusion

Correctness and completeness are two common metrics for measuring conflation success. They are used in the literature for all kind of features (polygons, lines and points), with similar definitions and for evaluating geometric conflation success. Most of the quantitative numerical results are high, always above 80%. However, this is only a global percentage for the whole map and its representativity could be discussed. Is it a conflation process better than other if they have 10% more of correct features?

In the case of line features, Chen *et. al* (2006, 2008) improve the definition not using the whole features but lengths of segments, applying a matching buffer method (Wiedemann *et al.*, 1998) to a conflation evaluation. Analyzing the results using partial length instead of the whole features drives to worse correctness percentages, which makes sense since the “zoom” in the evaluation is being increased.

Metrics as a length give a distance in meters instead of a percentage, usually explaining through a measurement of how far the conflated features are from the original ones. RMSE as defined in the NSSDA standard (FDGC, 1998), only take into account the results in the control points, without considering the geographical position of them. Control points could be badly distributed and thus there would be a bias in the results. In the case of line features, Chen *et. al* (2006, 2008) make the measurement more local by measuring the percentage of the total length of the (conflated) roads that is within a specified distance x to the reference roads, instead of computing only an average (Chen *et al.*, 2004). This allows to group the results by distance threshold, obtaining a percentage for each of them.

General statistics measures calculated by Doytsher *et al.* (2001) could be useful to study the distribution of the distance differences between the original dataset and the conflated dataset. However, the result obtained is only an improvement percentage from original to conflated features.

Considering the data producer point of view, conflation execution time is worth considered. Regarding execution time, the results are always variable, as they depend on a lot of parameters: map scale, number of features, information available for the map (e.g. road directions), matching algorithm used and conflation algorithm used. It is hard to determine a general time threshold and it will be depending of each case of study. Also, none of the literature reviewed provide a relative value (e.g. seconds per object feature) which means the results are not comparable.

Other possibilities for metrics are found, besides from lengths, percentages or time units. They are distances used to measure shape preservation, which could evaluate conflation success analyzing how the shape of the conflation features change. Shapes could also be compared using other conflation methods.

Open problems

Most of the literature reviewed evaluate the geometric conflation and only a few of them consider other aspects which can be taken into account in these processes (object shapes, presence and absence of objects...). Moreover, the metrics used are basic, being mostly percentages and displacements. Only metrics as the RMSE, Hausdorf distance and other possible distances (e.g. Fréchet distance) are defined mathematically, which brings more consistency to the conflation evaluation reports.

There is a lack of a standard metric to measure conflation success. The RMSE as defined by NSSDA (FDGC, 1998) is one of the exceptions found, but there could be another possibilities as ASPRS standards for Positional Accuracy for Digital Geospatial Data (ASPRS, 2015) and UNE 148002 standard (2016), recently published.

All metrics are strong metrics in the sense of González *et al.* (2013), as they are objective and do not take user context and purpose into account. Moreover, they evaluate the map as a whole. The numbers and percentages obtained are taken as representative of the entire region, even though there could be regions with good accuracy and others with less accuracy.

Execution time metrics found in traditional conflation processes are not sophisticated. This could be due to the fact that although the execution time is always important, it is only a data production issue. Nevertheless, in an ephemeral context, execution time evaluations are decisive since it represents the time the user will wait till the conflation process is finalized and it should be strongly considered.

Furthermore, all of these metrics are intrinsic to the data and do not consider neither the user-context nor the purpose. There is a need for developing weak met-

rics if the ephemeral conflation wants to be used and applied, needing further considerations and development since it could represent a new scenario in cartography, linked to the concept of smart cities and pervasive spatial data.

References

- Abbas, I. (1994). “Base de Données Vectorielles et Erreur Cartographique: Problèmes Posés par le Contrôle Ponctuel; une Méthode Alternative Fondée sur la Distance de Hausdorff”, Ph.D. Thesis, Université de Paris, Paris, France, 197 pp.
- 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 Trans. Patt. Anal. Mach. Int.*, 13:209-216.
- Aronov, B.; Har-Peled, S.; Knauer, C.; Wang, Y. and Wenk, C. (2006). “Frechet distance for curves, revisited”, *Lecture Notes in Computer Science*, 4168:52-63.
- ASPRS (2015). “Positional Accuracy Standards for Digital Geospatial Data. American Society for Photogrammetry and Remote Sensing”. Available at <<http://cort.as/oc0N>> accessed November 19th, 2016.
- Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; G. Ouzounis and Portugali, Y. (2012). “Smart cities of the future”, *The European Physical Journal: Special Topics*, 214(1):481-518.
- Casado, M.L. (2006). “Some basic mathematical constraints for the geometric conflation problem”, in *Proceedings of the 7th international symposium on spatial accuracy assessment in natural resources and environmental sciences*, Lisbon, Portugal, 264-274.
- Chen, C.C.; Knoblock, C.A. and Shahabi, C. (2006). *Automatically conflating road vector data with orthoimagery*. *GeoInformatica*, 10(4), 495-530.
- (2008). “Automatically and accurately conflating raster maps with orthoimagery”, *GeoInformatica*, 12(3):377-410.
- Chen, C. and Knoblock, C. (2004). “Automatically and accurately conflating orthoimagery and street maps”, in *Proceedings of the 12th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS’04)*, ACM Press, Washington, D.C., pp. 47-56.
- Cobb, M. *et al.* (1998). “A rule-based approach for the conflation of attributed vector data”, *Geoinformatica*, 2(1):7-35.
- Cobb, M.; Petry, F. and Shaw, K., (2000). “Fuzzy spatial relationship refinements based on minimum bounding rectangle variations”, *Fuzzy Sets and Systems*, 113(1):111-120.
- D’Errico, J. (2006). “Understanding GRIDFIT, 2006”, available at <<http://cort.as/oc0X>>, accessed July 22, 2016.

- Doytsher, Y.; Filin, S. and Ezra, E. (2001). "Transformation of Datasets in a Linear-based Map Conflation Framework", *Surveying and Land Information Systems*, 61(3):159-169.
- Edwards, D. and Simpson, J. (2002). "Integration and access of multi-source vector data", Symposium on geospatial theory, processing and applications, 9-12 July, Ottawa, Canada.
- FGDC (1998). *Geospatial Positioning Accuracy Standards; Part 3: National Standard for Spatial Data Accuracy*, Federal Geographic Data Committee, FGDC-STD-007.3, Washington, D.C., 28 pp. Available at <<http://cort.as/oc0f>>, accessed July 22, 2016.
- González, C.-H., López-Vázquez, C. and Bernabé, M.-Á. (2013). "Ephemeral Conflation", *The Cartographic Journal*, 50(1):43-48.
- Knauer, C.; Löffler, M.; Scherfenberg, M. and Wolle, T. (2009). "The directed Hausdorff distance between imprecise point sets", *Lecture Notes in Computer Science*, 5878:720-729.
- Lemarie, C. and Raynal, L. (1996). "Geographic data matching: first investigations for generic tool", *Proceedings of the GIS/LIS096 annual convention*, 19-21 November, Denver, ACSM, pp. 405-420.
- López-Vázquez, C., and González, C.H. (2009). "The need of a framework to compare Geometric Conflation Algorithms", in *12th AGILE International Conference on Geographic Information Science*, Hannover, pp. 1-5.
- Mackaness, W. and Edwards, G. (2002). "The importance of modelling pattern and structure in automated map generalisation", in *Joint workshop on multi-scale representations of spatial data*, 7-8 July, Ottawa, ON, pp. 1-11.
- MATLAB (2009). *Data gridding*, available at <<http://cort.as/oc0p>>, accessed July 22, 2016.
- OGC (2016), *Sensor Web Enablement (SWE)*, available at <<http://cort.as/oc0t>>, accessed July 20, 2016.
- UNE 148002 (2016), *Methodology for positional accuracy assessment of geographic information*, Asociación Española de Normalización y Certificación (AENOR). Normas UNE (Una Norma Española), available at <<http://cort.as/oc1->>, accessed November 19, 2016.
- Ruiz, J.J.; Ariza, F.J.; Ureña, M.A. and Blázquez, E.B. (2011). "Digital map conflation: a review of the process and a proposal for classification", *International Journal of Geographical Information Science*, 25(9):1439-1466.
- Saalfeld, A. (1983). "Conflation: automated map compilation", PhD thesis. CS-TR-3066, University of Maryland College Park, College Park, MD, USA, 133 pp.
- Samper, J. and J. Carrera (1990). *Geostadística: aplicaciones a la Hidrología Subterránea*, Ed. CIMNE, Barcelona, Spain, 484 pp.

- Sledge, I.; Keller, J.; Song, W. and Davis, C. (2011). "Conflation of vector buildings with imagery", *IEEE Geoscience and Remote Sensing Letters*, 8(1):83-87.
- Touya, G.; Coupé, A.; Jollec, J.; Dorie, O. and Fuchs, F. (2013). "Conflation Optimized by Least Squares to Maintain Geographic Shapes", *ISPRS International Journal of Geo-Information*, 2(3):621-644.
- Vauglin, F.; Bel Hadj Ali, A. (1998). "Geometric Matching of Polygonal Surfaces in GISs", *Proceedings of ASPRS Annual Meeting*, Tampa, FL, USA, 30:1511-1516.
- Wiedemann, C.; Heipke, C.; Mayer, H. and Jamet, O. (1998). "Empirical evaluation of automatically extracted road axes", *Empirical Evaluation Techniques in Computer Vision*, 172-187.
- Yang, B., and Zhang, Y. (2015). "Pattern-mining approach for conflating crowdsourcing road networks with POIs", *International Journal of Geographical Information Science*, 29(5):786-805.
- Yang, S.; Kim, C.; Huh, Y. and Yu, K. (2012). "Removal of spatial inconsistency and shape disagreement for conflation of road maps with aerial images", *Canadian Journal of Remote Sensing*, 38(6):723-737.
- Yuan, S. and Tao, C. (1999). "Development of conflation components", *Proceedings of geoinformatics'99 conference*, Ann Arbor, USA, pp. 1-13.
- Zhang, M.; Yao, W.; and Meng, L. (2016). "Automatic and Accurate Conflation of Different Road-Network Vector Data towards Multi-Modal Navigation", *ISPRS International Journal of Geo-Information*, 5(5):68-84.