Research article

Harmonization of categorical maps by alignment processes and thematic consistency analysis

Rita Nicolau[1],*, Nadiia Basos[1], Filipe Marcelino[1], Mário Caetano[2] and José M. C. Pereira[3]

[1] DGT—Direção-Geral do Território, Lisboa, Portugal
[2] DGT—Direção-Geral do Território; NOVA IMS—NOVA Information Management School, Universidade Nova de Lisboa, Lisboa, Portugal
[3] CEF/ISA—Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade de Lisboa, Lisboa, Portugal

* **Correspondence:** Email: rnicolau@dgterritorio.pt.

Abstract: This paper describes an approach for harmonizing historical vector categorical maps with related modern maps. The approach aims at the correction of geometric distortions and semantic disagreements using alignment processes and analysis of thematic coherence. The harmonized version of the map produced by this approach can already be overlaid with other maps, what was unfeasible with the original map.

The positional errors of the old map are reduced by two consecutive geometric adjustments, which use transformations usually available in most GIS software. The thematic consistency between the old and the modern map is achieved by harmonizing their classification systems and by the inclusion of specific contents missing in the early map, but represented in the modern map (e.g. small rivers).

This approach was tested in the geometric and thematic harmonization of the Portuguese Land Cover/Land Use (LCLU) map for 1990 (COS90). In this test, the 1995 orthorectified aerial images and the 1995 LCLU map (COS95) were used as reference sources of higher positional accuracy, to align the COS90 map. COS90 was firstly adjusted with the 1995 aerial images by an NTV2 grid transformation, developed by the authors. Then, for reduction of the local distortions, the map resulting from the first transformation was aligned with the COS95 by a rubber-sheeting linear interpolation transformation. This geometric harmonization enabled a decrease of the Root Mean Square Error of COS90 from 204 meters to 13 meters. The thematic harmonization of COS90 enabled its comparison

with modern related maps, and the integration of 201 river sections, that were missing because the specifications used in the production of the original map did not allow their representation.

Keywords: categorical maps; geometric and thematic harmonization; positional errors

1.   Introduction

This research was motivated by the need to compare an old categorical map, in vector format, with related modern maps to study landscape evolution. Historical maps may contain geometric distortions caused by outdated production processes, and/or make use of outdated classification systems. In these situations, their overlay with a related contemporary map has to be preceded by their geometric and semantic harmonization. The geometric harmonization aims at correcting the positional errors of the old map, and the semantic (or thematic) harmonization intends to make its classification system compatible with that of the related modern map, as well as the correction of classification inconsistencies.

This paper proposes an approach for harmonizing an old, vector categorical map with related modern maps. The approach aims at the correction of geometric distortions and semantic disagreements using alignment processes and analysis of thematic coherence. The harmonized version of the early map produced by this approach can already be overlaid with other maps, what was unfeasible with the original map. While the thematic or semantic harmonization of maps is essential to perform spatial analysis, such as change assessments, the geometric harmonization of a map improves the change estimates resulting from its comparison with similar maps because it contributes to the reduction of the erroneous change areas caused by positional errors [1].

Our approach was tested in the geometric and thematic harmonization of the Portuguese Land Cover/Land Use (LCLU) map for 1990 (COS90). Despite the lack of information on the thematic and the geometric accuracy of COS90, it is well known that this map, produced in the late nineties, presents geometric distortions and thematic inconsistencies that require corrections [2]. While the geometric errors of COS90 resulted mostly from its outdated production process, the thematic errors were mainly caused by discrepancies in the classification of land cover units distributed by several sheets of the original map. The above-mentioned study [2] tested a methodology for the geometric correction of COS90 that involved its comparison with an equivalent LCLU map with higher positional accuracy. For this purpose, COS90 was overlaid with a 2004 LCLU map and harmonized by reclassification and generalization operations. The authors concluded that the improved COS90 produced by the adopted methodology still contained geometric errors, which could have been partially corrected if a map adjustment had been applied before the tested methodology. Therefore, in our study, we opted first to produce an adjusted COS90 dataset, which can be later compared (overlaid) with a similar dataset for the correction of remaining geometric errors.

1.1. Geometric harmonization of categorical maps

The identification of geometric errors and thematic inconsistencies on a categorical map may be accomplished by its comparison (overlay) with a related independent map, of higher positional accuracy (preferably of a close date) [1,3,4]. By ensuring that the time lag between maps is as short as possible, we reduce the areas of change in the overlay map, which is used to identify polygons that may correspond to false changes (slivers). Slivers are polygons (usually small and elongated) that may result from the overlay of two vector maps, which portray the same reality. They are caused by the misalignment of the boundaries of the same object (polygon) represented in the two maps and reflect positional errors of the maps [3,4,5]. The problem of positional error when overlaying spatial data from different sources has been widely addressed in the literature, and the methodologies proposed to solve it depend on the type of the maps to be compared, and on their technical characteristics [1,3,4,6,7].

The most common procedure to reduce the creation of slivers resulting from the overlay of two vector maps consists in specifying a tolerance (also known as fuzzy tolerance) for carrying out the overlay. This tolerance defines the maximum distance allowed to move any point (node or vertex) of a polygon boundary [6]. Therefore, in the overlay process the points within the fuzzy tolerance distance of each other, are moved to a common position. The fuzzy tolerance is set based on the characteristics of the maps being compared, namely the scale, geometric accuracy, minimum mapping unit and minimum distance between lines. Several alternative methods enable the identification and elimination of slivers, but they are not available in standard GIS as an automated process [8,9].

When a map has large geometric distortions, it should be aligned or adjusted to a reference dataset by spatial transformations [4,10–13], before comparison with a related map. The alignment or adjustment transformations are also used for georeferencing historical maps and registering images/maps [14,15]. Their application requires the identification of matching objects on the two images/maps representing similar contents. These objects may be polygons, lines or points, the latter being the most common. The association between objects is usually established by pairs of points that link a location in the map/image (to be adjusted) to the corresponding location in the reference map/image. These pairs are also known as ground control points. Based on the collected pairs of points, the first map is aligned with the second using a global or a local transformation model [16–19]. The global methods make use of all the available control points to identify a single transformation, to be applied to the whole spatial domain. The local methods perform by sections of the spatial domain, applying at each a distinct transformation, which is identified based on neighboring control points. Conversely to global methods, local methods preserve the positions of the control points used in each transformation, producing estimates that match their original positions. On the other hand, local transformations do not allow control over distortions throughout the spatial domain, as global transformations do. Thus, the local methods allow for more efficient control of local distortions, while global transformations are most appropriate when the map distortions are homogeneous [12,16–19]. Depending on the type of estimates produced for the point positions (exact or approximate), the above-mentioned techniques are also classified into exact or approximate (non-exact) transformation models. Examples of global non-exact techniques are the similarity, the affine, the projective and the polynomial transformations. Examples of local transformations that produce exact estimates are finite elements, thin-plate splines, and rubber-sheeting [14,16–18].

The performance of a geometric transformation may be evaluated by statistical analysis of the distances between the point positions before and after the adjustment. Since the exact transformation models produce estimates that match the original points positions, the evaluation of their performance should be based on a set of control points that is independent of the set used for the adjustment [14].

2. Data and methods

2.1. Data

The base information of this study includes the Portuguese Land Cover/Land Use (LCLU) reference maps: COS90 and COS95. Both are vector categorical maps that cover Mainland Portugal for 1990 and 1995, respectively. In addition to these maps, we use the 1995 aerial orthorectified images. These images and the COS95 map were chosen to harmonize COS90 because they were the available datasets closest to 1990 (with known positional accuracy and Mainland coverage). In the first stage of the study, the aerial images and COS95 are used, as reference sources of higher positional accuracy, to adjust the COS90 map. Next, the classification system of COS90 is harmonized with that of COS95.

The 1995 images were obtained from a 1:40,000 scale flight and orthorectified with a Digital Terrain Model, generated from the altimetric data at scale 1:25,000. For improving its positional accuracy, some images were adjusted using the 2010 orthophotos as reference. Despite the lack of knowledge of a formal evaluation of the accuracy of these orthorectified images, it is reasonable to expect that their positional accuracy should be about three meters, based on the acquisition process above described.

COS90 production was initiated in the late nineties. It was based on the manual delineation of visually interpreted LCLU units, over non-orthorectified aerial photographs. Initially delineated on paper, these units were subsequently digitized [20,21]. The resulting vector map was structured into sheets according to the Portuguese map series 1:25,000, using the Hayford Gauss—Lisbon Datum coordinate system.

The classification system (nomenclature) of COS90 is non-hierarchical [20,21]. It is composed of 942 classes, establishing a wide variety of combinations between the main LCLU types: Agricultural, Forestry, Uncultivated, Unproductive, Social, and Water. This nomenclature is difficult to integrate with other nomenclatures, in particular with those resulting from hierarchical classification systems, such as the nomenclature of the CORINE Land Cover (CLC) map, which constitutes an LCLU reference product at the European level [20].

In 2007, significant changes were introduced in the production of most recent COS maps (namely COS95, COS2007 and COS2010) [22]. An effort to harmonize with international standards and practices led to the use of the European Terrestrial Reference System 1989 (ETRS89) and the adoption of a nomenclature compatible with the CLC nomenclature, up to the third level. Thus, a hierarchical classification system, encompassing five levels of thematic detail, was adopted. The first level of the nomenclature describes the following LCLU classes: Artificial areas (1), Agricultural, and agro-forestry areas (2), Forests and natural and semi-natural areas (3), Wetlands (4) and Water bodies (5).

The COS maps after 1990, were produced through visual interpretation of high spatial resolution orthorectified aerial images [22]. As the coordinate system ETRS89/PT-TM06 was adopted in their production, COS90 was also later converted into this system.

Although the COS maps most recently produced share the same nomenclature, the 1995 map (COS95) has less thematic detail than its successors (e.g. at the most detailed level of the nomenclature COS2007 has 225 classes, while COS95 has only 89 classes). Table 1 compares the technical specifications of COS90 with those of COS95. It reveals that the geometric accuracy of COS90 is unknown and that COS90 has greater thematic detail than COS95. Thus, the thematic harmonization of COS90 with COS95 will cause a reduction of the thematic detail of COS90. Table 1 also shows that the minimum distance between lines of objects represented in COS90 and COS95 is different (40 and 20 meters respectively). As COS95 has a smaller minimum distance between lines than COS90, some of the contents represented in COS95, although present in 1990 are not depicted in COS90. Another difference between the two maps concerns the Minimum Mapping Unit (MMU). The MMU of COS90 is lower than that of COS95. This means that COS90 contains areas (smaller than one hectare) that are not represented on COS95, even if they existed in 1995.

Table 1. Technical specifications of the Portuguese Land Cover/Land Use maps: COS90 and COS95. Source: Adapted from [20,23].

	COS90 (improved after 2007)	COS95
Data model	Vector (polygons)	Vector (polygons)
Scale	1/25,000	1/25,000
Minimum Mapping Unit (MMU)	<1 ha (about 0.5 ha)	1 ha
Minimum distance between lines	40 m	20 m
Coordinate reference system	ETRS89/PT-TM06	ETRS89/PT-TM06
Nomenclature	Non-hierarchical classification system (942 classes)	Hierarchical classification system with one level of detail for classes 1, 4 and 5 and five levels of detail for classes 2 and 3 (89 classes at the most detailed level)
Geometric Accuracy	Unknown	≤ 5.5 m
Thematic Accuracy	Unknown	$\geq 85\%$
Base data	1990 non-orthorectified aerial photographs; 1995 orthorectified aerial images (spatial resolution: 1 m)	1995 orthorectified aerial images (spatial resolution: 1 m)

Figure 1 shows the COS90 map reclassified into five LCLU classes, according to the first level of the CLC nomenclature.
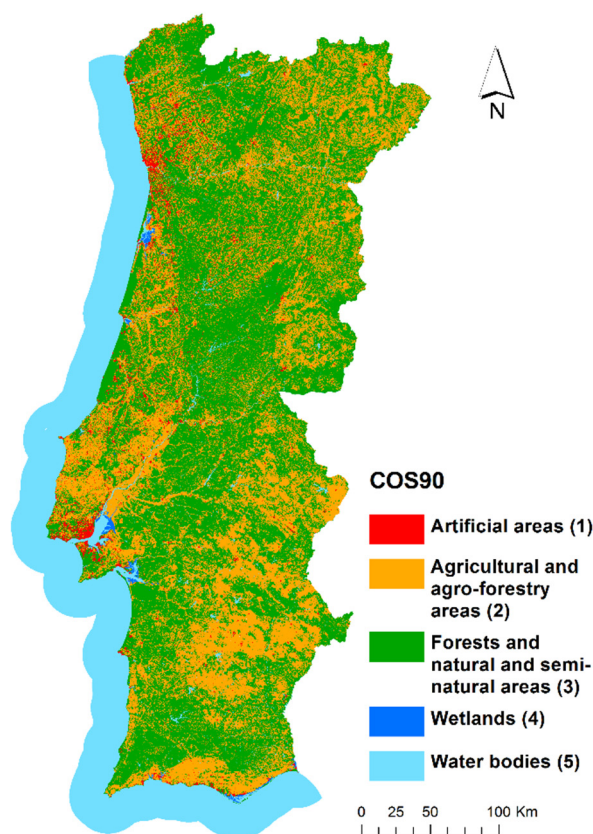
Figure 1. COS90 map reclassified into the five classes integrating the first level of the CORINE Land Cover nomenclature.

2.2. Methods

In the geometrical harmonization of COS90, we applied spatial transformations available in Geographic Information System (GIS) software or libraries, except one which was developed in the Python language. For the thematic harmonization of COS90, we applied reclassification procedures, followed by an analysis of the contents that were missing in COS90 but could be imported from COS95 map (e.g. small rivers). The harmonization of COS90 included the following steps:

I. Horizontal positional accuracy assessment of the COS90 map;

II. Spatial adjustment of COS90 map using the 1995 orthorectified aerial images as a reference;

III. Horizontal positional accuracy assessment of the adjusted COS90 map resulting from II;

IV. Spatial adjustment of COS90 map produced by step II using the COS95 map as a reference;

V. Horizontal positional accuracy assessment of the adjusted COS90 map resulting from IV;

VI. Reclassification of the COS90 map produced by step IV using a COS90-COS95 nomenclature correspondence table;

VII. Inclusion of some thematic contents missing in COS90 but represented in COS95.

2.2.1. Horizontal positional accuracy assessment of the COS90 map

As the positional accuracy of the COS90 map was unknown, the first stage of the study (I) comprised its quantification. Geometric harmonization intends to increase the positional accuracy of a map. Therefore, the knowledge of the positional accuracy of the map is useful before the harmonization process [19]. The assessment of the planimetric positional accuracy of COS90, required the identification of pairs of points connecting selected locations in COS90 to corresponding locations in a reference map/image. Due to the lack of adequate contemporary data to correct COS90, the orthorectified aerial images of 1995 were selected as a reference dataset of higher positional accuracy. For this purpose, 2653 pairs of control points were manually collected in positions that did not change between 1990 and 1995. Their placement ensured the existence of at least three control points per 160 km^2 (the area of a map sheet of the original COS90) and a Mainland density of 0.03 control points per km^2.

As the distance between a pair of control points depicts the magnitude of a positional error in our map, their average gives an overall error index. According to the US National Standard for Spatial Data Accuracy [24] the horizontal positional accuracy can be quantified through the Root Mean Square Error (RMSE), which derives from the RMSE in X and Y components (RMSE$_x$ and RMSE$_y$). The US National Standard for Spatial Data Accuracy (NSSDA) states, however, that spatial accuracy should be reported in ground distances at a 95% confidence level. Since the computation of the NSSDA accuracy measure differs if RMSE$_x$ ≠ RMSE$_y$ or RMSE$_x$ = RMSE$_y$, the value to be reported will correspond to (2.4477 × 0.5 × (RMSE$_x$ + RMSE$_y$)) or to (1.7308 × RMSE), respectively [24]. This study presents both the RMSE and the horizontal accuracy at a 95% confidence level.

The accuracy assessment was repeated following each spatial transformation applied to COS90 (steps III and V). For this we collect, at each time, a set of control points independent from the set used to transform the map.

2.2.2. Geometric harmonization of the COS90 map

Previous studies [19,25] show that two consecutive geometric adjustments enable a more effective correction of map distortions, therefore in the geometric harmonization of COS90, we also applied two consecutive geometrical adjustments. According to referred studies, while the first adjustment intends to produce a globally adjusted map, the second adjustment intends to correct local distortions remaining from the first adjustment.

Figure 2 explains the workflow used for the geometric harmonization of COS90. In the first adjustment (step II), the COS90 map was aligned with the 1995 orthorectified aerial images, using 2013 pairs of control points. Initially, we did not make use of all the control points collected for the first COS90 positional accuracy assessment, due to the need to evaluate the performance of the transformations we intended to test with part of the collected control points. Therefore, we reserved a set of 640 pairs of control points for the evaluation of the performance of the transformations, a procedure also known as model validation. After testing several transformations, their estimates (obtained for the validation control points) were compared using the Root Mean Square Error (RMSE) parameter. The transformation with the lowest RMSE was chosen to align the COS90 dataset with the

1995 orthorectified aerial images using all the collected control points (2653). As the adjustments depend on the distribution and density of the control points [12,16], it was ensured that both sets of control points (testing set and validation set) had at least two control points per 160 km^2.

In the second adjustment (step IV), the COS90 map resulting from the first adjustment was aligned with the COS95 map. Although the two maps are not independent, COS95 was selected to perform the second adjustment not only due to its higher positional accuracy than the adjusted COS90 map, but also due to our interest in assessing the LCLU changes between 1990 and 1995. To link the locations of the adjusted COS90 map with the corresponding locations of the COS95 map, we used a new set of control points, composed of 32035 pairs. These control points were obtained by a semi-automated procedure. Firstly, we generated a network of dummy pairs of control points over Mainland Portugal using a Python script that placed a dummy control point in each cell of a 2000 m fishnet (a net of rectangular cells), ensuring the presence of a dummy control point per 4 km^2. The distance between the origin and destination of the dummy pair of control points had to be very small to avoid displacing or distorting correctly adjusted areas, so it was set at a shift of 0.5 meters in X and Y directions. Such an approach made the dummy distances to be 0.707 m length in the NE direction. Then, the COS90 and COS95 maps were visually inspected at the scale 1:20,000 and in the areas of local distortions, the dummy pairs of control points were manually replaced by real pairs of control points. This procedure allows that the pairs of control points placed manually, adjust the map locations presenting local distortions, while the pairs of control points placed in an automated way, hold map locations considered already aligned. The pairs of dummy control points that were manually replaced by real pairs (in locations of local distortions identified by visual inspection) represent 43% of the total used to transform the map. For the validation of the second adjustment, we used 640 new, manually collected, pairs of control points.

Several global and local methods were tested in the first adjustment. The global methods comprise the following transformations: 1st order polynomial—Affine; 2nd order polynomial; 3rd order polynomial; Projective; and Similarity. The local methods include transformations using: Spline; Rubber-sheeting—Linear interpolation; Rubber-sheeting—Natural neighbor; and NTv2 grid (with different cell sizes). The latter is a local transformation method in which the spatial domain is subdivided into a quadrangular mesh (an NTv2 format grid), such as that proposed by Junkins and Farley [26]. In our study, the work of those authors was adapted to transform coordinates between reference systems that do not involve a change of datum. The NTv2 grid is created using a chosen interpolation method. The vertices of the NTv2 rectangular cells are generated homogeneously/regularly according to the chosen cell size, and their values are interpolated from the surrounding control points. Later in the adjustment, the dataset vertices falling within each grid cell were transformed by bilinear interpolation using the four vertices of the cell. To test this method, several NTv2 grids with different cell sizes (0.005º; 0.01º; 0.02º; 0.04º; where 0.01º = 36 seconds ≈ 1 km) were created using the set of validation control points and different interpolation methods (Minimum curvature; Kriging; Inverse Distance Weighting; Natural neighbor).

The transformation chosen for the second adjustment (step IV) was a Rubber-sheeting, using linear interpolation [27,28]. This transformation uses two temporary triangulated irregular networks (TIN) to interpolate changes in X and Y coordinates of map objects that are nearby given control points. These constitute the nodes of the TIN, the surface of which was created by linear interpolation. Its

adoption is advised to correct local distortions, following a dataset global transformation [29]. A comparative study of several geometric transformations used to register an extract of the historical "Map of France" [12], concluded that in the presence of a large number of ground control points the linear transformation based on Delaunay triangulation (i.e. a Rubber-sheeting using linear interpolation) was more accurate than the other models tested.
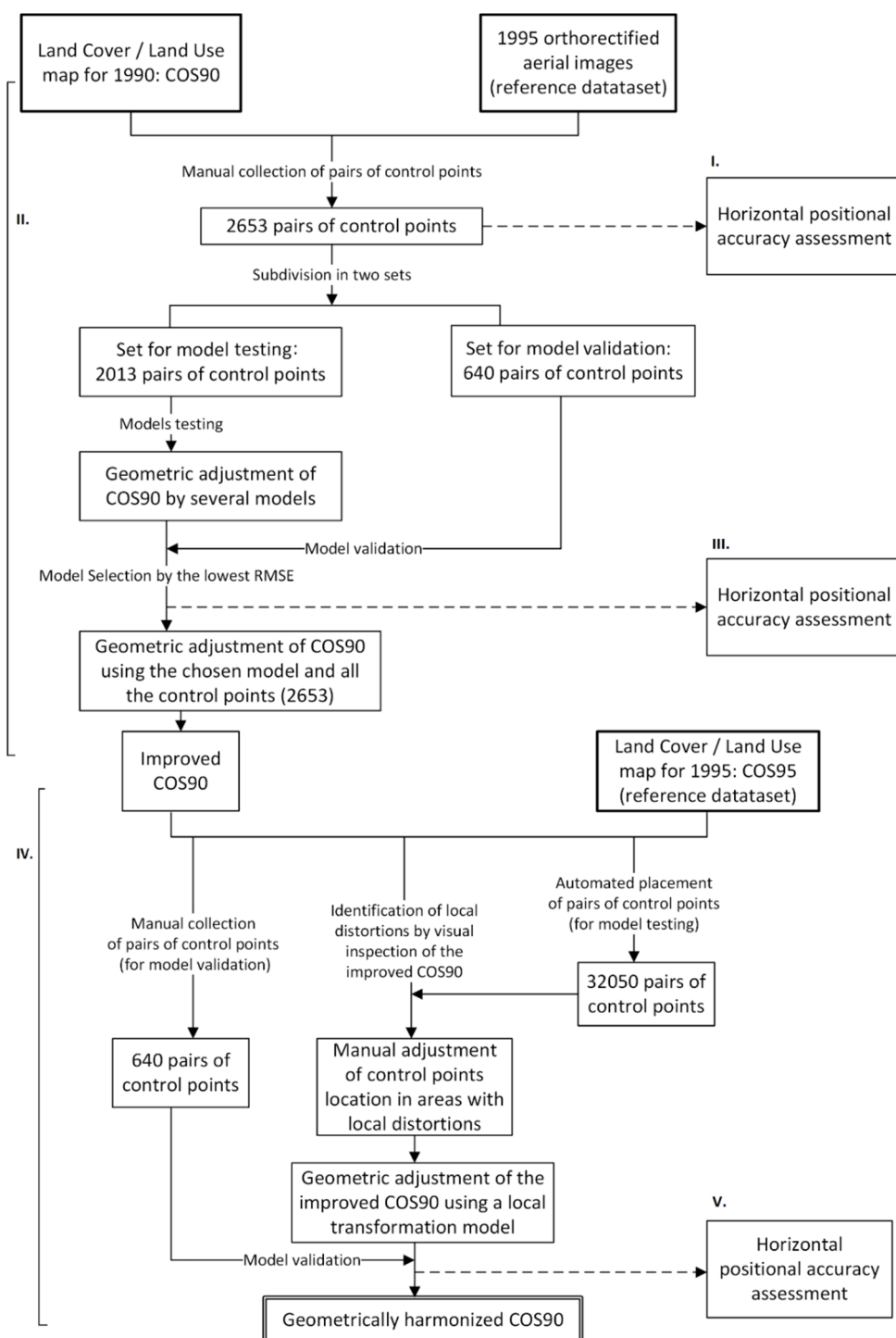

Figure 2. Workflow used for the geometric harmonization of COS90.

2.2.3. Thematic harmonization of the COS90 map

To harmonize the semantic content of the two LCLU maps, the COS90 map, geometrically harmonized, was reclassified into the LCLU classes of the COS95 map (step VI). For this, we used a COS90-COS95 nomenclature correspondence table, which ensured that the COS90 map had a classification system equivalent to those of the COS95 map, up to the third level. As COS95 has less thematic detail than COS90, this process led to a generalization of the COS90 map.

The minimum distance between lines of objects represented in COS90 and COS95 is different (40 m and 20 m respectively), therefore, some of the thematic contents portrayed by COS95 are not represented in COS90, although they were present in 1990. An analysis of the contents represented in COS95 that were missing in COS90 revealed that such objects are typically roads and rivers. The small sections of rivers that were missing in COS90 were identified by comparing the polygons of the corresponding class in both maps. Such polygons were imported from COS95 into COS90. This approach was not applied to the roads because the time lag between the two maps does not ensure that the entities represented in COS95 do not correspond to real landscape changes. In fact, in the period 1990–1995 several new roads were built with European Community funds.

3. Results

The positional accuracy assessment of the COS90 map, before its geometric adjustment, revealed horizontal errors ranging between 142.4 and 246 meters, with a mean of 203.3 meters and a standard deviation of 14.7 meters. The RMSE estimate obtained for this early COS90 map is 203.8 meters. As the error in X component was lower than the error in the Y component ($RMSE_x = 103.8$ m; $RMSE_y = 175.4$ m), the dataset tested 341.7 meters horizontal accuracy at a 95% confidence level. This RMSE estimate (203.8 m) is extremely high when compared to the equivalent estimate for COS95 (5.5 m), which justifies the need to improve the COS90 geometry. The validation results of the methods tested in the first geometric adjustment of the COS90 map are presented in Table 2. These results show that most global and local transformations had similar performances, leading to RMSE values close to 22 m. Due to its lowest RMSE (21.85 m), the adjustment method chosen was the NTv2 grid with 0.02 degrees (cell size of 3667.4 km in the north and 3951.6 km in the south), created by Inverse Distance Weighting (IDW) interpolation method with a power value equal to one. This local transformation is hereinafter referred to as NTv2 0.02° IDW (power = 1).

Table 2. Validation results for the transformations tested in the first geometrical adjustment of the COS90 map.

Transformation methods	Implemented using	RMSE (m)
Non-exact		
1st order polynomial (Affine)	GDAL library	22.19
2nd order polynomial	GDAL library	22.11
3rd order polynomial	GDAL library	22.09
Projective	ArcGIS	22.17
Similarity	ArcGIS	22.21
Exact		
Spline	GDAL library	25.29
Rubber-sheeting—Linear interpolation	ArcGIS	23.37
Rubber-sheeting—Natural neighbor	ArcGIS	22.73
NTv2 grid with 0.01°—Minimum curvature	*	24.42
NTv2 grid with 0.01°—Kriging (slope = 1)	*	22.94
NTv2 grid with 0.01°—Minimum curvature (tension = 1)	*	21.98
NTv2 grid with 0.01°—Inverse Distance Weighting (power = 1)	*	21.87
NTv2 grid with 0.01°—Inverse Distance Weighting (power = 1.5)	*	22.06
NTv2 grid with 0.01°—Inverse Distance Weighting (power = 2)	*	22.42
NTv2 grid with 0.005°—Inverse Distance Weighting (power = 1)	*	21.95
NTv2 grid with 0.02°—Inverse Distance Weighting (power = 1)	*	21.85
NTv2 grid with 0.04°—Inverse Distance Weighting (power = 1)	*	21.96
NTv2 grid with 0.01°—Natural neighbor	*	22.90

* programmed in Python with interpolations performed in ArcGIS or Surfer.

The COS90 map adjusted by the NTv2 0.02° IDW (power = 1) transformation, presented horizontal errors ranging between 0.7 and 81.7 meters, with a mean of 19 meters and a standard deviation of 10.7 meters. The RMSE obtained for this map (21.9 meters) is substantially lower than that of the early COS90 (203.8 m). The first adjustment did also decrease the COS90 horizontal mean error (from 203.3 to 19.0 meters) and the maximum horizontal error (from 246 to 81.7 meters). Despite these results, we decided to apply a second geometric adjustment to the already transformed COS90, grounded on the following:

- The RMSE estimate achieved after the first adjustment (21.9 m) was considered very high compared to the positional accuracy estimate of COS95 (5.5 m);
- The low quality of some of the orthophotos used in the first adjustment hampered the identification of reliable control points;
- The visual inspection of the COS90 map resulting from the first adjustment, in areas with highest adjustment errors revealed the existence of significant local distortions.

As explained in the methods section, the transformed COS90 map was aligned with the COS95 map by a Rubber-sheeting using linear interpolation. The map resulting from this transformation presented horizontal errors ranging between 0.9 and 42.3 meters, with a mean of 11.9 meters and a standard deviation of 5.1 meters. The RMSE estimate of the COS90 map after this second adjustment

is 13 meters, which reflects a considerable reduction of the corresponding value after the first adjustment (21.9 m). Although this value exceeds twice the RMSE of COS95, it is considered acceptable given the COS90 production process. As the additional corrections applied to COS90 did not involve further adjustments, this is the RMSE of the final map produced by our study. Since its error in X component is lower than the error in the Y component ($RMSE_x$ = 8.8 m; $RMSE_y$ = 9.5 m), the COS90 dataset produced by the second adjustment tested 22.4 meters horizontal accuracy at a 95% confidence level.

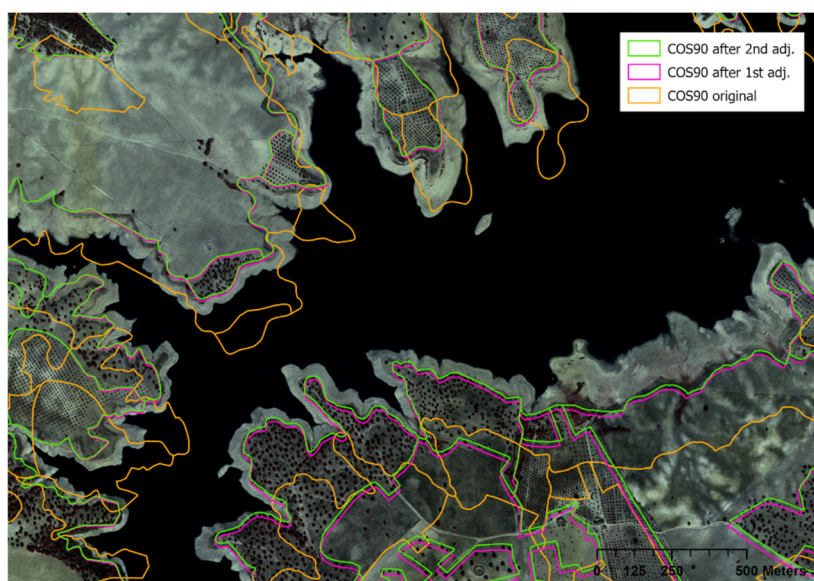Figure 3 illustrates the results of the adjustments applied to COS90 for an area with a water body.

Figure 3. Excerpt of the COS90 map near a water body (before and after each adjustment), superimposed with the aerial images of 1995.

Table 3 presents statistical descriptors on the horizontal errors of the COS90 map before and after each geometric transformation. It shows that the first adjustment has reduced the COS90 horizontal mean error from 203.3 to 19.0 meters, while the second adjustment lowered it to 11.9 meters. The maximum horizontal error also decreased from 246 to 81.7 meters with the first adjustment, dropping to 42.3 meters with the second adjustment.

Table 3. Statistical descriptors of the horizontal errors of the COS90 map before and after each geometric correction (first adjustment: NTv2 0.02o IDW (power = 1) transformation; second adjustment: Rubber-sheeting—Linear interpolation transformation).

Horizontal error	Original COS90 (m)	COS90 after the first adjustment (m)	COS90 after the second adjustment (m)
Minimum	142.4	0.7	0.9
Mean	203.3	19.0	11.9
Maximum	246.0	81.7	42.3
Standard Deviation	14.7	10.7	5.1

In what concerns the thematic, or semantic, harmonization of the map COS90, our analysis identified 201 river sections (with a total surface of 27.2 km^2) that were represented in the COS95 map but were missing in COS90. This thematic content was imported to COS90.

4. Discussion

This research was motivated by the need to quantify the changes that have occurred in the landscape since a specific date, which falls within an epoch in which digital maps were produced by methods that have now fallen into disuse. This paper describes a methodological approach for the geometric and thematic harmonization of past or historical maps with similar current maps. The maps concerned are vector thematic (or categorical) maps that present incompatible classification systems. Furthermore, the early map has an unknown geometric accuracy, which seems to be very low.

The main objectives of the geometric and thematic harmonization are the reduction of both, positional errors of the early map, and semantic disagreements between the classification system of the old map and that of the modern maps. The geometric harmonization comprised the application of two consecutive geometric transformations to the old map. For the thematic harmonization, a table of correspondences between maps nomenclatures was created, which served to reclassify the old map. Next, by comparison of corresponding classes in the two maps, the contents represented in the modern map, which are missing in the old map, were identified. Some of these contents (e.g. small river sections) were imported to the old map because they existed at the time, but the map technical specifications did not enable their representation.

This approach was applied to the Portuguese LCLU map for 1990 (COS90), using the 1995 orthorectified aerial images and the 1995 LCLU map (COS95) as reference sources of higher positional accuracy. Its application comprised the transformation of the COS90 map by two consecutive adjustments, the assessment of its positional accuracy before and after these transformations, and the analysis of the thematic coherence between the COS90 map and the ensuing COS map (COS95).

As mentioned in section 1.1, the choice of a harmonization approach should rely on the type of maps to compare (e.g. topographical, categorical, cadastral) and their attributes (reference date, geometric accuracy, scale, amongst others). We select the above-described approach, according to the availability of data and their specifications, for enabling the assessment of landscape changes.

The geometric harmonization of the COS90 map by adjustment transformations was decided after the first evaluation of its planimetric positional accuracy. The low map accuracy (RMSE = 204 meters) did advise against its overlay with other maps. On the other hand, the reference data available to harmonize the COS90 map (the 1995 orthorectified aerial images and the 1995 LCLU map: COS95) are relative to a time frame that is five years after that of COS90.

The COS90 map was firstly adjusted using the 1995 orthorectified aerial images. The selection of 1995 orthorectified aerial images for the first adjustment of the COS90 map, was grounded on the fact that these images are the single available reference dataset which is independent of COS90. As the map COS95 portrays a temporal evolution of COS90 (i.e. it is not independent of CO90) and presents less positional accuracy than the orthorectified aerial images, we have decided not to use it in

the first adjustment of COS90. Furthermore, the low positional accuracy of the COS90 map advised against its overlay with COS95.

The decision to apply two consecutive geometrical adjustments to the COS90 map was based on previous research, which defends that the first transformation produces a globally adjusted map, while the second "refines the results of the former" by correcting local distortions [19,25]. In these studies, the map is first aligned by a global transformation and next by a local transformation. In our study, the selection on the type of transformation (global or local) to be used in the first map adjustment, was determined by the method that had the lowest RMSE estimate, which ended up to be a local transformation method. Therefore, the COS90 map was aligned with the 1995 images, by the NTv2 0.02° IDW (power = 1) transformation. Considering that the RMSE estimates of most alternative transformations were very close to the chosen one, which was implemented in Python, we do not recommend its use. Thus, we admit that the first adjustment could have been performed by much simpler transformations, as the Affine, which would have caused a negligible decrease (34 cm) of the positional accuracy of the transformed map.

Our results show that most of the global and local transformations tested in the first map adjustment had a similar performance. This may be due to the high number of control points used, as well as to their proper spatial distribution. In line with our outcome, other researchers [12] showed that the differences in accuracy, between global and local transformation models, are greater for small sets of control points than for large sets.

As the accuracy of the COS90 map resulting from the first adjustment was considered very low compared to the positional accuracy of the modern COS maps, we decided to apply a second transformation to the already transformed COS90, using the 1995 equivalent map as a reference. To perform this second adjustment, we opt to choose a specific local transformation (Rubber-sheeting using linear interpolation), instead of comparing several transformations. We chose this local transformation based on the following:

(a)   We had a large and reliable set of control points;

(b)   This transformation enables keeping the map features that are already aligned, while locally adjusting the remaining map features by piecewise transformations that preserve straight lines;

(c)   This transformation has been often used to align historical maps with good results [10,12,30–32].

Our results indicate that the geometric harmonization of COS90 was very successful, which can be confirmed by visual comparison of the original map with the final one. The transformations applied to the early map did enable to decrease its RMSE from 203.8 meters to 13 meters, as well as a reduction of the mean and maximum horizontal errors of the map by 94% and 83%, respectively.

These results are comparable to those obtained in the detection and correction of positional errors and geometric distortions of topographic data, based on rigorous Synthetic Aperture Radar (SAR) image simulation and mathematical modelling of SAR imaging geometry [10]. By aligning the topographic dataset of the Dry Valley Region (at scale 1:50,000), using a piecewise linear transformation based on Delaunay triangulation, the researchers [10] were able to reduce its RMSE from 200 m to 50 m.

The implementation of the proposed approach relies on the availability of data sources for use as a reference in the harmonization process and their technical characteristics. In this study, the 1995

images and the COS95 map were used as reference datasets. Concerning the 1995 aerial images, we found that many did not have sufficiently good quality for the identification of control points. As regards the COS95 map, its main limitation is its smaller thematic detail than the 1990 map and the COS maps after 1995. In the thematic harmonization process, the COS90 map was reclassified using the COS90–COS95 correspondence table. This reclassification led to a substantial loss of the semantic detail of the original map, which nomenclature was composed of 942 classes. To mitigate the loss of the semantic content of the COS90 map, we could have used the 2007 map (COS2007) as a reference instead of the 1995 map (COS95). However, we avoid using it because the larger time lag between maps would cause a reduction of the areas that did not change between the two dates, which would hinder the geometric harmonization process.

Our study required an exhaustive gathering of sets of control point pairs, which linked COS90 with the datasets used as a reference. It is believed that both, the spatial distribution of the collected control points and the size of the control point sets, contributed to the good performance of the applied transformations. Although the placement of the control points for the second map adjustment, was based on a semi-automated procedure that has speeded up the process significantly, almost half of the control point positions were latter manually adjusted. Therefore, we admit that the control point gathering methods used are time-consuming, and may constitute an obstacle to the reproduction of our methodology in other studies.

The main recommendations arising from our study concern the selection of reference datasets and the compilation of control points to perform the map adjustment. The reference datasets should be independent of the historical map and preferably of a period close to that of the historical map. Although the size of the sets of control points used to align a map is very important, it is desirable that the positioning of the control points on the map is rigorous to ensure that their spatial distribution is homogeneous.

5. Conclusions

This paper describes an approach for harmonizing early or historical vector categorical maps with related modern maps. The approach aims at the correction of geometric distortions of the old map by its alignment with reference datasets of higher positional accuracy. It also aims to reduce semantic discrepancies between the old map and the related modern maps, by harmonization of their classification systems. These procedures, enable the inclusion of specific thematic contents of the modern map into the old map, as the specifications used for the production of the old map did not allow their representation, despite their presence at the time. The map resulting from the described harmonization process has a stated positional accuracy, that is higher than that of the original map, and a classification system that is compatible with that of modern related maps.

This approach was applied to the Portuguese Land Cover/Land Use map for 1990 (COS90), using the 1995 orthorectified aerial images and the 1995 related map (COS95), as reference sources of higher positional accuracy. While the geometric harmonization enabled to substantially increase the positional accuracy of the COS90 map (its Root Mean Square Error dropped from 204 to 13 meters), the thematic harmonization made its comparison with modern related maps possible, as well as the integration of some thematic contents that were missing (201 small river sections).

The harmonization approach presented in this study, although designed to enable the comparison of old categorical vector maps with related modern maps, can be applied in diverse contexts that require the integration of historical maps with modern maps, such as the inclusion of historical datasets in a GIS or Spatial Data Infrastructure. Therefore, it is of interest to users and producers of geographic information, namely those developing GIS analysis.

Extension of this work planned for the short-term is an additional harmonization of the map produced by this study. For this purpose, we intend to test different fuzzy tolerances in the overlay of the harmonized COS90 map with the COS95 map. The several change maps produced by this test will enable the identification of additional positional and classification errors in the COS90 map.

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. Mas JF (2005) Change Estimates by Map Comparison: A Method to Reduce Erroneous Changes Due to Positional Error. *Trans GIS* 9: 619–629.
2. Luz A, Nunes A, Caetano M (2008) Proposta metodológica para a correcção das distorções geométricas da COS'90: resultados preliminares. In: Rocha JG, editor. Proceedings of the 10th ESIG—X Encontro de Utilizadores de Informação Geográfica. Oeiras Portugal, Universidade do Minho, 14.
3. Chrisman NR, Lester M (1991) A diagnostic test for error in categorical maps. In Proceedings of the Tenth International Symposium on Computer-Assisted Cartography—Auto-Carto 10. American Congress on Surveying and Mapping and American Society for Photogrammetry and Remote Sensing, Baltimore Maryland, 330–348.
4. Ware JM, Jones CB (1998) Matching and aligning features in overlayed coverages. In Laurini R, Makki K, Pissinou N, editors. Proceedings of 6th ACM International Symposium on Advances in Geographical Information Systems—ACM-GIS'98. Association for Computing Machinery, 28–33.

5. Gaeuman D, Symanzik J, Schmidt JC (2005) A Map Overlay Error Model Based on Boundary Geometry. *Geogr Anal* 37: 350–369.

6. Veregin HA (1989) Taxonomy of Error in Spatial Databases. Technical Report 89-12; Santa Barbara: U.S. National Center for Geographic Information and Analysis, University of California, 109.

7. Harvey F, Vauglin F (1996) Geometric match processing: Applying multiple tolerances. In Kraak MJ, Molenaar M, editors. Proceedings of the 7th International. Symposium on Spatial Data Handling: Advances in GIS Research II—SDH'96. Delft University of Technology, Delft, The Netherlands, 1: 13–29.

8. Galanda M, Koehnen R, Schroeder J, et al. (2005) Automated generalization of historical U.S. census units. In: Proceedings of the 8th ICA Workshop on Generalisation and Multiple Representation. International Cartographic Association, A Coruña, Spain, 10.

9. Schorcht M, Krüger T, Meinel G (2016) Measuring Land Take: Usability of National Topographic Databases as Input for Land Use Change Analysis: A Case Study from Germany. *ISPRS Int J Geo-Inf* 5: 134.

10. Liu H, Zhao Z, Jezek KC (2004) Correction of Positional Errors and Geometric Distortions in Topographic Maps and DEMs Using a Rigorous SAR Simulation Technique. *Photogramm Eng Remote Sens* 70: 1031–1042.

11. Schuurman N, Grund D, Hayes M, et al. (2006) Spatial/temporal mismatch: a conflation protocol for Canada Census spatial files. *Can Geogr/Le Géogr Can* 50: 74–84.

12. Herrault P, Sheeren D, Fauvel M, et al. (2013) A comparative study of geometric transformation models for the historical "Map of France" registration. *Geogr Tech* 1: 34–46.

13. Schulze MJ, Thiemann F, Sester M (2014) Using Semantic Distance to Support Geometric Harmonisation of Cadastral and Topographical Data. *ISPRS Ann Photogramm Remote Sens Spatial Inf Sci II-2*, 15–22.

14. Brovelli MA, Minghini M (2012) Georeferencing old maps: a polynomial-based approach for Como historical cadastres. *e-Perimetron* 7: 97–110.

15. Barazzetti L, Brumana R, Oreni D, et al. (2014) Historical Map Registration via Independent Model Adjustment with Affine Transformations, In: Murgante B, Misra S, Rocha AMAC, et al. (Eds.), *Computational Science and Its Applications—ICCSA 2014*, Portugal: Springer, 44–56.

16. Boutoura C, Livieratos E (2006) Some fundamentals for the study of the geometry of early maps by comparative methods. *e-Perimetron* 1: 60–70.

17. Balletti C (2006) Georeference in the analysis of the geometric content of early maps. *e-Perimetron* 1: 32–42.

18. Bitelli G, Cremonini S, Gatta G (2009) Ancient map comparisons and georeferencing techniques: A case study from the Po River Delta (Italy). *e-Perimetron* 4: 221–233.

19. Dai Prà E, Mastronunzio M (2014) Rectify the river, rectify the map. Geometry and geovisualization of Adige river hydro-topographic historical maps. *e-Perimetron* 9: 113–128.

20. Caetano M, Pereira M, Carrão H, et al. (2008) Cartografia temática de ocupação/uso do solo do Instituto Geográfico Português. *Mapp-Rev Int Cienc Tierra* 126: 78–87.

21. Caetano M, Araújo A, Nunes V (2009) Proposta de conversão da nomenclatura da COS'90 para a da COS2007. Lisboa: Instituto Geográfico Português, 51.

22. Caetano M, Nunes A, Dinis J, et al. (2010) Carta de Uso e Ocupação do Solo de Portugal Continental para 2007 (COS2007v2.0): memória descritiva. Lisboa: Instituto Geográfico Português, 97.

23. DGT—Direção-Geral do Território (2018) Especificações técnicas da Carta de uso e ocupação do solo de Portugal Continental para 1995, 2007, 2010 e 2015. Lisboa: Direção-Geral do Território, 103.

24. FGDC—Federal Geographic Data Committee (1998) Geospatial Positioning Accuracy Standards. Part 3: National Standard for Spatial Data Accuracy. Available from: https://www.fgdc.gov/standards/projects/accuracy/part3/chapter3.

25. Guerra F (2000) 2W: new technologies for the georeferenced visualization of historic cartography. *Int Arch Photogramm Remote Sens* 33: 339–345.

26. Junkins D, Farley S (1995) NTv2 National Transformation Version 2—Developer's Guide. Canada: Geomatics Canada—Geodetic Survey Division, 27.

27. White MS, Griffin P (1985) Piecewise linear rubber-sheet map transformations. *Am Cartographer* 12: 123–131.

28. Saalfeld A (1985) A fast rubber-sheeting transformation using simplical coordinates. *Am Cartographer* 12: 169–173.

29. ESRI—Environmental Systems Research Institute (2016) ArcGIS for Desktop documentation About spatial adjustment rubbersheeting. Available from: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/editing-existing-features/about-spatial-adjustment-rubbersheeting.htm.

30. Fuse T, Shimizu E, Morichi S (1998) A study on geometric correction of historical maps. *Int Arch Photogramm Remote Sens* 32: 543–548.

31. Balletti C (2000) Analytical and quantitative methods for the analysis of the geometrical content of historical cartography. *Int Arch Photogramm Remote Sens* 33: 30–37.

32. Shimizu E, Fuse T (2005) A Method for Visualizing the Landscapes of Old-Time Cities Using GIS. In: Okabe A (Eds.), *GIS-based Studies in the Humanities and Social Sciences*, CRC Press Taylor & Francis group, 265–277.