A computational modelling approach to reconstruct the fluvial system of the floodplain of Lucca in the Roman period.

Article · January 2022

DOI: 10.13131/unipi/2785-0668/k5sw-3e98

2 authors:

Salvatore Basile
Università di Pisa

Francesco Carrer
Newcastle University

doi 10.13131/unipi/2785-0668/k5sw-3e98

Salvatore Basile*, Francesco Carrer**

* Dipartimento di Civiltà e Forme del Sapere – Università di Pisa.
** School of History, Classics and Archaeology – Newcastle University.

A COMPUTATIONAL MODELLING APPROACH TO RECONSTRUCT THE FLUVIAL SYSTEM OF THE FLOODPLAIN OF LUCCA IN THE ROMAN PERIOD

Abstract: During antiquity, the floodplain of Lucca was crossed by numerous branches of the *Auser* river. Scholars recognised several paleo-channels in the area and proposed different possible paths of the river during the Roman period, relying on the study of place names and documentary sources, on the observation of Lucca's urban plan, on remote sensing, and archaeological excavations. Far from proposing a conclusive reconstruction of Lucca's ancient landscape, this research complements other studies and contributes to strengthening our knowledge of the paleo-fluvial system of the plain. This study aims to test a new methodology – based on logistic regression modelling – for assigning traces of paleo-channels to a given period based on selected geomorphological and archaeological variables.

Keywords: Lucca, *Auser*, environmental archaeology, geostatistics, logistic regression modelling.

1. Introduction

Located in northern Tuscany, the floodplain around Lucca has always been characterised by the presence of the *Auser*/Serchio river with its several branches. Throughout the Etruscan age, small settlements developed in the plain, primarily located in the prominent morphologies between the numerous branches of the river (Ciampoltrini, 2005; 2010; Ciampoltrini & Zecchini, 2007). The foundation of the Latin Colony of *Luca* (180 BC) marks an important milestone in the history of the territory: the first centuriation of the *ager* can be dated to the first decades of the 2nd century BC, while a second one dates to the second foundation of the colony in the Triumviral or Augustan age (Ciampoltrini, 2009; 2016). Centuriation was not only used for agricultural purposes, but it was also aimed at regulating and containing the waterways. However, the alteration that centuriation brought to the delicate hydrogeological balance of the Plain contributed to triggering flooding events that deeply conditioned local settlement patterns (Bini et al., 2020). Starting from the 2nd century AD, the poor maintenance of irrigation canals might be related to the increased frequency of floods recorded by numerous archaeological research in the territory. This constant threat might have accelerated the depopulation phenomenon of the *ager*. Flood risk and depopulation, in turn, created a cause-effect relationship that determined the swamping of the territory during the Middle Ages (Basile, 2021) and the appearance of Lake Bientina in its southernmost part.

Even though neither the river nor the lake exists anymore, the plain's delicate hydrogeological balance is still observable in periods of intense rainfall, when the area of the Bientina riverbed turns back to swamp.

The study of the landscape and its features is fundamental to understanding settlement's dynamics and characteristics. Besides the numerous floods that determined the life and the abandonment of many settled areas, the *Auser* was mostly used for trading purposes, as it constituted the easiest and fastest way to connect Lucca, the city of Pisa, and the sea. Necessarily, its waters were also used for manufactures, agricultural, breeding and fishing purposes. Hence, reconstructing the *Auser* river path and correlating it with the settlement pattern can lead to a better comprehension of the complex interaction between humans and landscape during the Roman period.

Today, numerous paleo-channels are known in the Plain and several studies were conducted to reconstruct the ancient path of the river (Ciampoltrini et al., 2009; Cosci, 2005; Mencacci & Zecchini, 1981, pp. 10-32). Nevertheless, the analysis of aerial photographs and the study of place names alone do not enable a reliable identification of the paleo-riverbeds active during the Roman age. In fact, although the possible path of the river has already been approximately reconstructed, which of the many paleochannels belong to the Roman period is still unclear.

This study proposes a novel methodology for identifying Roman paleo-riverbeds using archaeological layers dated to the Roman period[1]. Sophisticated spatial interpolation techniques (Lloyd & Atkinson, 2004) are employed to create a palaeo-DEM from these data. A map displaying the probability of each location to be a riverbed is produced through logistic regression, using archaeological data as dependent variable and ancient geomorphological parameters (derived from the palaeo-DEM) as predictors. The use of logistic regression in landscape archaeology has significantly increased in recent years. It is primarily used to predict the spatial distribution of archaeological sites (Carrer, 2013) and to model the characteristics of past environments (Alberti et al., 2018). To the best of our knowledge, this is the first attempt to use this methodology to infer ancient watercourses. The modelling results will be compared with the palaeo-channels identified through remote sensing, to isolate those palaeo-channels that are more likely to date to the Roman period.

2. Background

2.1 Geology and geomorphology

The plain of Lucca (fig. 1) falls within a large tectonic depression, which began taking shape in the northern Apennines starting from the Upper Miocene (Ambrosio et al., 2010; Nardi et al., 1987). It is bounded by the Apuan Alps and the Apennines to the North, the Monti Pisani to the South-West, the Cerbaie upland to the South-East, and the Monte Albano to the East. The area was occupied by a large lake during the Pleistocene; the relationship among the lake deposits, layers referring to a fluvial system in the Cerbaie, and the general geological and geomorphological features suggest that the region of the ancient Pleistocene depression was affected by important and diversified periods of tectonic uplift and by differentiated cycles of erosion and fluvial sedimentation. These events outlined the essential geomorphological features of the region (Nardi et al., 1987, p. 133).

From the Würm period, the river *Auser* started shaping the morphology of the plain. Due to the tectonic movement, and the continuous episodes of flooding and swamping of the plain caused by the rising level of the river Arno, the *Auser* splitted. From the north, the watercourse flowed throughout the area of Bientina to reach the river Arno, while several minor branches detached from the main river with a North-East/South-West direction, reaching the sea through the areas of Ripafratta and Pisa (Salvini et al., 2006, p. 300).

[1] For this study, we have considered only data that date from the early 2nd century BC to the late 2nd century AD. This timespan provides a reasonable trade-off between the resolution of archaeological data and the variability of the fluvial system.

fig. 1. The floodplain of Lucca.

From a hydrological perspective, it is possible to distinguish three types of alluvial deposits in the plain: sandy-gravelly deposits with high permeability to the North, where the river enters the plain from the Apennines; silty-sandy deposits with medium permeability, which constitute the largest alluvial sediment and cover a large part of the plain; silty-clayey deposits with low permeability located in the South-eastern portion of the plain (Ambrosio et al., 2010, p. 217; Nardi et al., 1987, p. 135)

Local infiltrations on the Apennines and the sub-bed flow of the Serchio River recharge the aquifer system of the area; meteoric infiltrations also play the same role together with streams descending from the Monti Pisani and the Cerbaie hills (Ambrosio et al., 2010, pp. 217-218).

As far as the morphology of the plain is concerned, its average slope is approximately 2.4%, with a 4% slope in its northern area and a 1.5% slope in the southernmost part. The elevation varies from 31 m above sea level in the northernmost part to 4 m in the Bientina area.

The morphology combined with the several antropic interventions, as well as the abundance of both superficial and underground water, produced numerous digressions of the river-beds. These dynamics led to the formation of the different paleo-channels currently visible through remote sensing.

2.2 The Auser river

In order to reconstruct the main branches of the Auser river in the Roman period, many scholars relied on the study of place names and documentary sources (Barsocchini, 1853; Mencacci & Zecchini, 1981, pp. 18-32; Paderi, 1932; Repetti, 1839), on the observation of Lucca's urban plan (Sommella & Giuliani, 1974, p. 17), on remote sensing (Cosci, 2005; Salvini et al., 2006) and on archaeological excavations (Ciampoltrini & Andreotti, 2008, pp. 13-16).

The integration of all these approaches enabled some tentative reconstructions. Once it entered the plain at Ponte a Moriano (fig. 2), the *Auser* splitted into two branches: a smaller

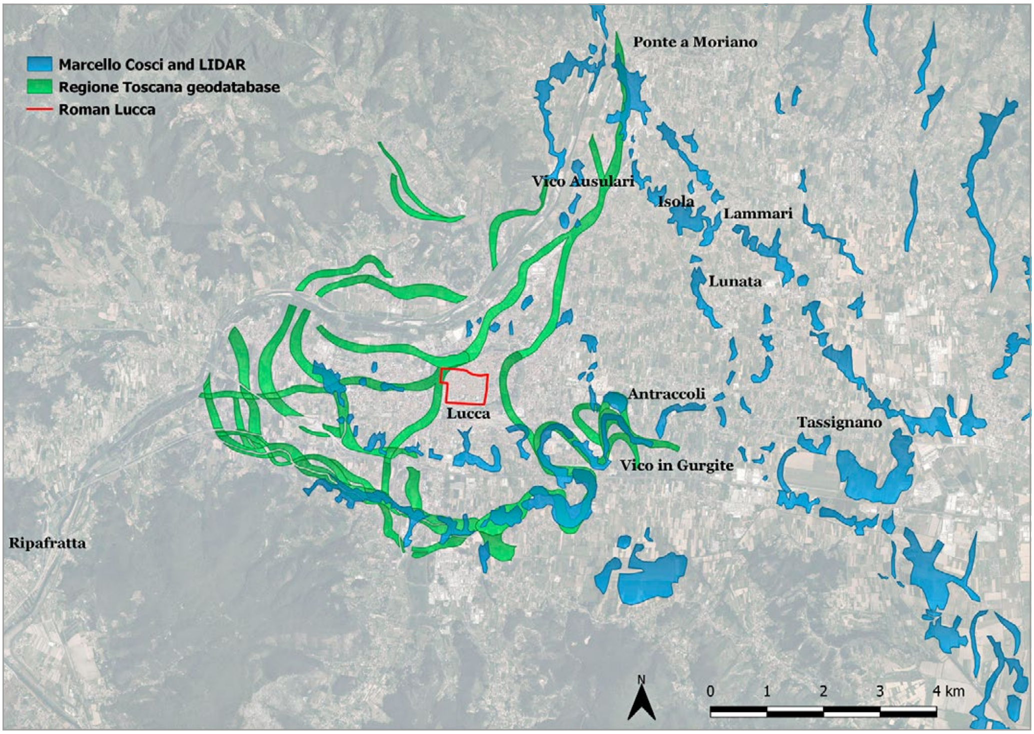

fig. 2. The westernmost part of the plain, with an indication of paleo-channels known for the area and place names.

one, the *Auserculus*, which bent towards the West entering the Plain of Pisa through Ripafratta and is today almost replicated by the Serchio River; and a bigger branch which, instead, continued towards the South. In the area of Lammari this second branch splitted into two further branches: an eastern one which will be discussed later, and a western one which the numerous toponyms allow to reconstruct. From North to South several locations refer to the presence of the river. The area of San Pietro a Vico, indeed, since the 8th century was known as "*Vico Ausulari*" (Mencacci & Zecchini, 1981, p. 23), while a little further to the South-East the hamlet of "*Isola*" suggests that this area was located between two branches of the river. Furthermore, according to Mencacci and Zecchini (1981, p. 24) the toponym "*Lunata*" owes its name to the proximity of a river bend. To the South-West, the river crossed the area today known as Antraccoli, a toponym usually interpreted as "*inter aquas*" (Giannoni, 2015, pp. 107-108). Finally, the area today known as Pieve di San Paolo used to be called "*Vico in Gurgite*" (Barsocchini, 1837, p. 1636). The last two places might suggest a meandering and swirling course of the river, which is indeed confirmed by the photo interpretation analysis conducted by Marcello Cosci (2005). The *Auser* thus continued West, passing the city of Lucca to the South and again merging to the already mentioned branch on the way to Ripafratta and Pisa.

Based on this evidence, it can be argued that the city of Lucca was located in the middle of a fluvial island which provided, alongside the Late-Republican walls, a natural defensive feature. Perhaps, the river flowed very close to the city, as the shape of the northern and western walls might suggest (Ciampoltrini, 2008, p. 26; Sommella & Giuliani, 1974, p. 17).

Unfortunately, this portion of the plain is one of the hardest to analyse due to the many paleo-channels investigated in the suburbs, which makes it difficult to discriminate between those active in the Roman period and those active earlier or later. Besides, the significant

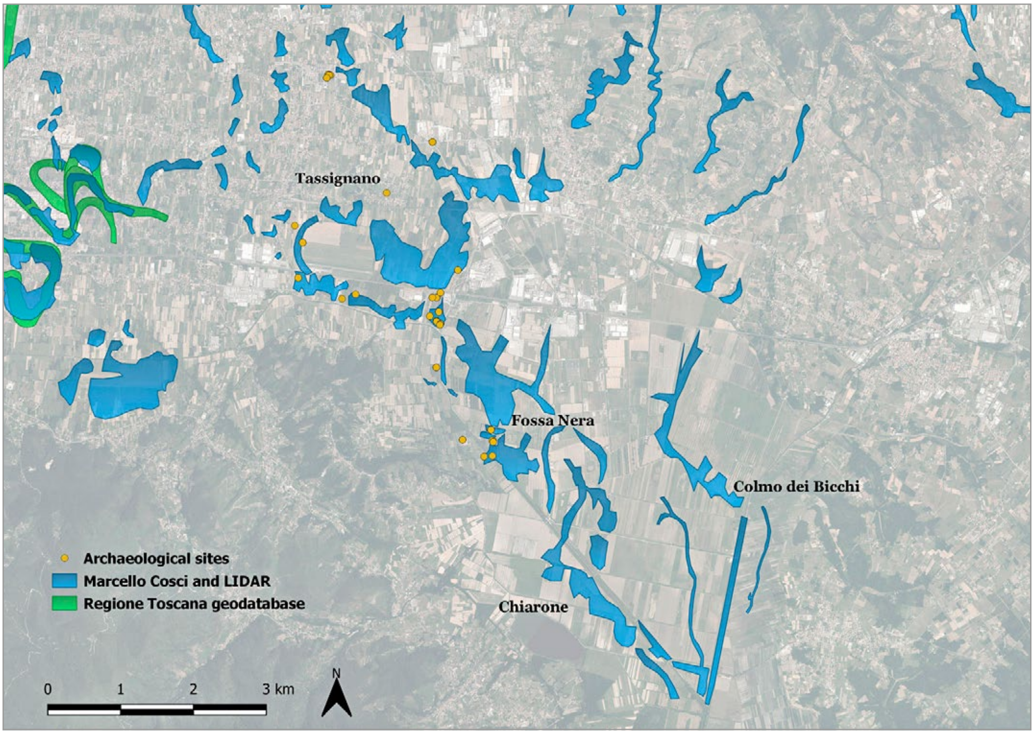

fig. 3. The southernmost part of the plain, with an indication of paleo-channels known for the area, archaeological sites between Tassignano and Fossa Nera, and place names.

urbanisation of the area around the city of Lucca prevents a reliable identification of the paleochannel associated with the perimeter of the Roman city. On the other hand, this is one of the areas with the highest number of stratigraphic analysis of the whole plain. Such interpretative challenges and the availability of stratigraphic data led to the selection of the plain around Lucca as a case study for this research.

From the Lammari bifurcation (fig. 3), the eastern branch of the river descended towards the South, with a large bend in the Tassignano area. The layout of the investigated settlements here seems to match the layout of the paleo-riverbed.

The passage of the *Auser* in this sector of the plain is also documented by written sources from the Early Medieval period. Emanuele Repetti (1839, p. 707) mentions a document dated November 18 956 which reports the exchange of some assets in the "*Canabbia*" area near Tassignano. The source defines *Cannabbia* as «next to the *Ausere*».

To clarify the location of this toponym, the same author refers to another document dated 28 March 953 in which a piece of land is cited «*in loco Quarto, ubi dicitur a Canabbia*»; it is thus deduced that the *Auser* passed through Quarto, an area today located between Tassignano and Capannori.

Continuing South, the river lapped the area where two Late Republican farms (known as "Fossa Nera A" and "Fossa Nera B") are located. The traces of paleo-riverbeds in this area make the location of the two farms uncertain with respect to the course of the river. The paleo-channel detected by Cosci is in fact located to the east of the two buildings, while Giulio Ciampoltrini (2009, p. 25) considers a second trace between the two buildings more reliable.

The different interpretations of this area combined with the relatively high density of stratigraphic research conducted in this part of the plain, led to the choice of the plain between Tassignano and the two Fossa Nera farms as the second case study for this research.

Finally, the river splitted into three further branches. Some of them, such as the one in the Chiarone area, have only been recognised through aerial photography, while others, such as the one located in Colmo dei Bicchi, were identified during archaeological excavations (Ciampoltrini & Andreotti, 2008).

These three branches merged further South, before flowing in the Arno river.

3. Materials and Methods

3.1 Dataset

The data used for this analysis represent a small portion of a larger dataset created for a research project that aims at studying the whole ancient territory of Lucca from the 2nd century BC to the late 6th century.

With this purpose in mind, all the available archaeological and environmental data for this territory were collected. These data derive from both published and archival material.

All the data collected were stored in the Pisa University MAPPA project database (Anichini et al., 2012; Fabiani & Gattiglia, 2012) and geo-localised in QGIS.

The dataset used for this research is composed of 427 archaeological interventions, with 1196 finds in total. A huge percentage of these interventions – 326 out of 427 – is located on the plain, 1013 finds in total. The non-spatial information includes chronology of the excavated context, sedimentological characteristics, and elevation. Unfortunately, elevation was recorded in a limited number of contexts, the majority of which are located in two areas: the urban/suburban area of Lucca (Area 1), and the area of Capannori (Area 2). Following a thorough assessment of the dataset, a sub-sample of 36 locations for Area 1 (*elevation_ancient_area1.shp*) and 13 for Area 2 (*elevation_ancient_area2.shp*) have been selected.

Besides, modern altitude values have been extrapolated from 120 random locations in each Area, to be used as a co-variable in the co-kriging interpolation (*elevation_modern_area1.shp*, *elevation_modern_area2.shp*).

All the paleo-channels identified by Marcello Cosci (Cosci, 2005) were digitised on QGIS together with those detected on LIDAR relief (*paleo_channels_cosci_lidar.shp*). Paleo-riverbeds available on Regione Toscana geodatabase were also used (*paleo_channels_geodb.shp*).

Finally, to perform logistic regression, a series of "event" locations have been sampled in Area 1 and 2 where archaeological excavations recorded layers suggesting the presence of riverbeds in the Roman period. These locations were given a value of 1. Several "non-event" locations were sampled where the presence of a river in the Roman period could be excluded, such as Roman sites with structures. A value of 0 was assigned to these locations. All these locations, indicating the presence/absence of the river, constitute the dependent variable in the logistic regression. For Area 1, 13 "event" and 37 "non-event" locations were selected (*events_area1*); for Area 2, 11 "event" and 55 "non-event" locations were selected (*events_area2*). The overall number of locations, the relatively balanced ratio between "events" and "non-events" and the reliability of the attribution to each binary category enabled the creation of a solid regression model.

3.2 Data quality

Since the paleo-DEM is produced using locations for which stratigraphic information is known, a careful assessment of the quality of this information is paramount. One of the main problems is the non-uniformity of the information available for each location. The available data are the result of different types of archaeological research (rescue or research excavations, surveys, fortuitous recoveries) carried out over more than a century. Besides, this research was undertaken by different subjects, including amateur archaeological groups, universities, professional archaeologists, and superintendents. Therefore, the excavation methods and the documentation quality are significantly different. In addition to this, data collection in the Archives of the Soprintendenza of Tuscany highlighted significant gaps in the documentation. For example, it is not uncommon that no archival documentation is available for published excavations. This means that in some cases the elevation values used for producing the paleo-DEM are obtained from publications instead of the original reports. Furthermore, publications often approximate layer's elevations or cannot be used to estimate the elevation above sea level, because the reference point from which the excavation depth was computed is not specified.

The inconsistent level of information available for each location affects the reliability of the paleo-DEM. Since the predictors for the logistic regressions are derived from this paleo-DEM, model outcomes are also affected. In order to mitigate the effect of poor data quality, co-kriging method has been used for spatial interpolation and current morphological parameters have been used as additional covariates in the regression (see below).

Similar biases affect the identification of the dependent variable for the regression model. Each sample location has been assigned to a "river" (event) or a "non-river" (non-event) category, based on the available description of the stratigraphy (sediment texture, fabric, etc.). However, relying on what has already been interpreted by the excavators, instead of directly consulting the excavation documentation, often implies that interpretations might be unreliable. Furthermore, especially in the case of older research, the description of non-anthropic layers identified during the excavation was not considered as important as the description of structures or other types of evidence. Therefore, only few locations can be reliably assigned to each category, thus reducing the sample available to train the model. However, the event/non-event ratio (more than the absolute number of measurements) represents the most relevant parameter for the creation of a solid binomial model (King & Zeng, 2001), and this precondition is met in the sampled locations. Besides, a small training sample contributes to mitigating the spatial dependence of events and non-events (second-order effect), related to the fact that nearby locations are more likely to belong to the same category than locations further apart, as Tobler's First Law of geography states (Alberti et al. 2018; Miller, 2004).

3.3 Co-Kriging

The first step was the reconstruction of the topography of the plain during the Roman period (paleo-DEM), using geostatistical interpolation (kriging). Although the area under investigation extended from the city of Lucca to the location of Botronchio in the south, the dishomogeneous distribution of the elevation checkpoints prevented a reliable interpolation for the whole plain. Therefore, the area was divided in two: Area 1 is a 870,5 ha rectangular area, from 700 m to the West of the Roman walls of Lucca to the area of Antraccoli with a East-West direction; Area 2 is a 1605 ha irregular area, from Tassignano to Fossa Nera in the municipalities of Capannori and Porcari (fig. 4).

Despite this mitigation, the elevation checkpoints turned out to be not dense enough. After assessing a significant correlation between the elevation of the Roman levels and the elevation of the current ground (0.648 Pearson's coefficient for Area 1, 0.954 for Area 2), a co-kriging interpolation approach was used. Roman elevation values were used as target variable and modern elevation values as co-variable. The latter were measured both at the same locations as the target variable (co-located samples), and in randomly selected locations. Exploratory analysis and interpolation were implemented in R using the *gstat* package (Pebesma & Wesseling, 1998).

A *gstat structure* containing the two sample sets of ancient and modern elevations was built (Rossiter, 2007). Variogram models were added to the *gstat* object and a linear model of coregionalization was fitted. Finally, the *predict* function was used to produce the co-kriging interpolation; the *gstat* object is used as the first argument, while the second is a 20×20 m grid for each of the two interpolated areas (Area 1 and 2). The grid is a Spatial Points Data Frame created using the *create grid* function on QGIS. A Spatial Pixel object, to be used for the prediction, was created applying the *gridded* function.

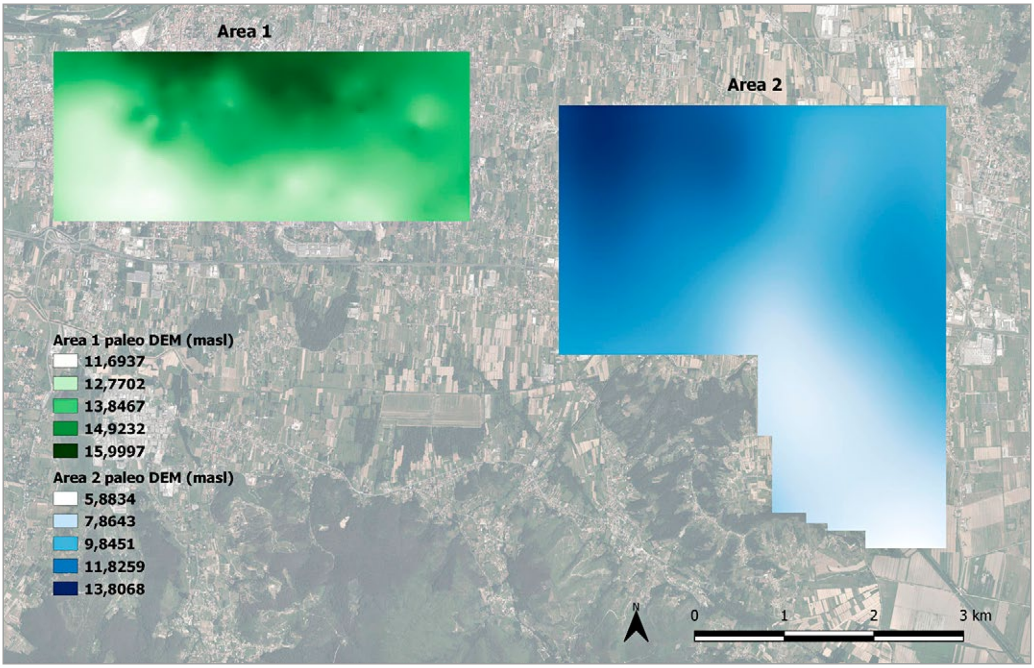

fig. 4. Paleo-DEMs of Areas 1 and 2.

3.4 Spatial covariates

Regression is a mathematical mapping of the relationship of two or more variables (Nakoinz & Knitter, 2018, pp. 87-97). Logistic regression, widely used for geospatial data, models the likelihood of each location to be an event or a non-event, based on the observed correlation between a training sample of events/non-events and a series of independent variables (or predictors). In this research the events are locations where a Roman riverbed has been identified and non-events are locations where the presence of a Roman riverbed can be excluded.

To model the probability of each location of the computation area to be a riverbed, a total of 18 predictors have been used. To overcome the plausible low reliability of the paleo-DEM, the same predictors derived from the ancient DEM have been derived from the modern elevation map as well.

Ancient and modern elevation are two key predictors, as landscape geomorphology is one of the main factors that influences the path of a watercourse (Kolka & Thompson, 2006). Variations in slope and aspect were also considered influential. Local variations of slope gradient and slope orientation (aspect) can have a profound influence on the directionality of surface water flows. Slope and aspect maps were created using GRASS 7 *r.slope.aspect* function. Sine and Cosine functions were applied to the aspect map through *r.mapcalc* to extract respectively the East-West and North-South aspect directions. Another predictor is the profile curvature, also created using *r.slope.aspect* function. Curvature indicates convergence and divergence of water flow, as they are determined by the terrain's convexity or concavity (Alberti et al., 2018, p. 10).

Other variables were produced in GRASS using *r.terraflow* (Flow Direction, Flow Accumulation, and Topographic Convergence Index) and *r.topidx* (Topographic Wettex Index) functions. Flow Accumulation and TWI are both used to describe the tendency of an area to accumulate water. The difference in these algorithms stands in the reckoning of the flow direction and in whether a single flow direction or multiple flow directions are computed (Ballerine, 2017; Mattivi et al., 2019).

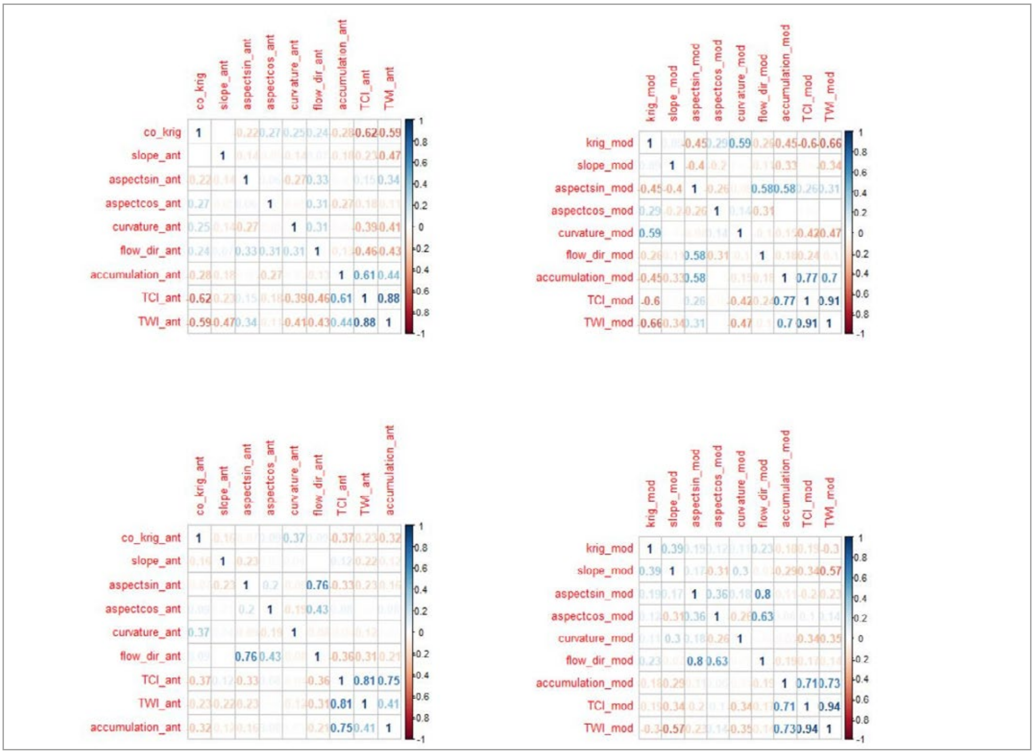

fig. 5. Corrplots: top left – Area 1 Model 1; top right – Area 1 Model 2; bottom left – Area 2 Model 1; bottom right – Area 2 Model 2.

3.5 Logistic regression modelling

Dependent variable (event/non-event) and predictors were estimated for 50 selected locations (training sample) in Area 1 and 66 in Area 2. The first step in the development of the regression model was the assessment of the correlation between predictors in the sampled locations, to prevent the negative effects of collinearity on coefficient estimation. Pearson's correlation index was calculated for each pair of predictors, discarding those variables with an index above 0.70 (Alberti et al., 2018, p. 13; Dormann et al., 2013). For Area 1, TCI and TWI derived from the modern and paleo-DEM were discarded, due to their strong collinearity with the DEMs and between themselves. For Area 2, TCI and flow direction derived from modern and paleo-DEM were discarded, alongside TWI from modern DEM (fig. 5). Two models were created for each area (*glm* function) using selected predictors generated from paleo-DEM (*Model1*) and modern DEM (*Model2*). Since multicollinearity among predictors may significantly inflate the variance of the estimated parameters, and the pair correlation matrix may not be able to identify it, Variance Inflation Factor (VIF) was also considered to increase the reliability of each model (Midi et al., 2010). As pointed out in literature (Alberti et al., 2018, p. 13; Midi et al., 2010, p. 261), a VIF above 10 indicates a significant multicollinearity. In Area1-Model 1, *aspectsin*, *curvature* and *flow_dir* had a VIF of 24.62, 16.83 and 18.38. *Aspectsin* was thus discarded, as it was the one variable with the highest coefficient. The procedure was then repeated without *aspectsin*, and the results showed no significant multicollinearity. Despite no predictors had a VIF greater than 10 in Area 1-Model 2, *flow_dir* was discarded, as it had a VIF of 9.83. None of the predictors used for Area 2 had VIF greater or close to 10.

The identification of the most parsimonious model was carried out using the *stepAIC* function, available in the *MASS* library of R, which performs a stepwise model selection by Akaike

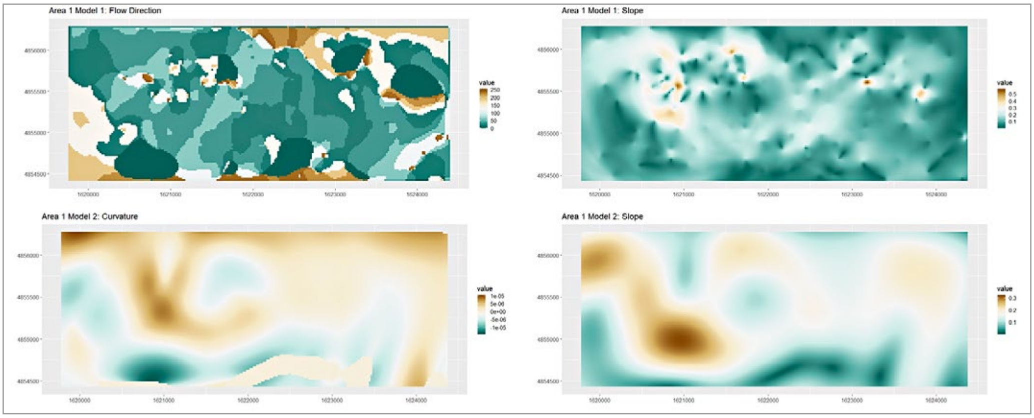

fig. 6. Spatial covariates selected by Models 1 and 2 for Area 1.

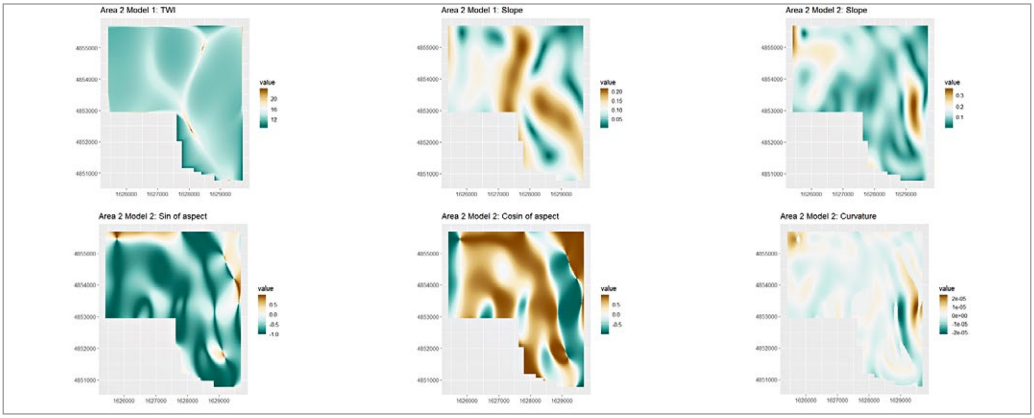

fig. 7. Spatial covariates selected by Models 1 and 2 for Area 2.

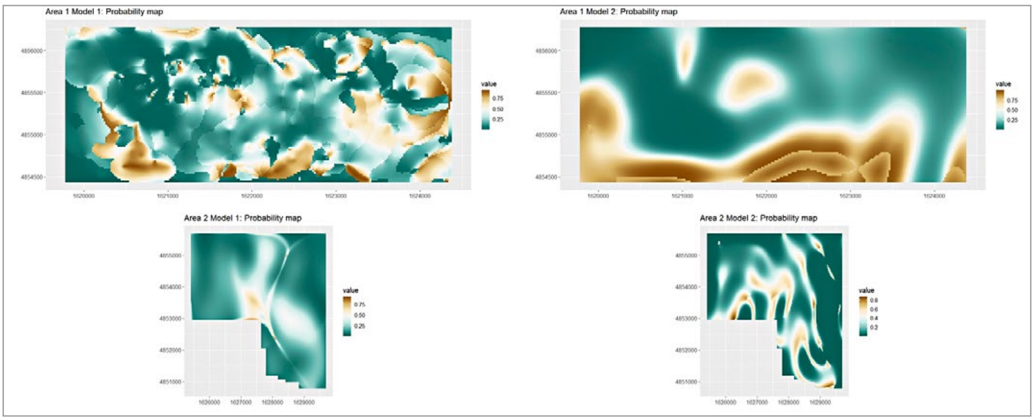

fig. 8. Probability maps.

Information Criteria (AIC). It is known that the best model is usually the one that provides the best fit using the lowest possible number of parameters; the stepwise function computes the different combinations of predictors in the model, returning the model with the lowest

AIC value, namely, according to the principle of parsimony, the model with the best balance between likelihood and number of predictors (Venables & Ripley, 2002, pp. 175-177). As a result of this process, flow direction and slope were selected as significant predictors for Area1-Model 1 (*Model1AIC*), while curvature and slope were selected for Area 1-Model 2 (*Model2AIC*) (fig. 6). For Area 2, only slope and TWI were retained for Model 1, while slope, sin of aspect, cosine of aspect and curvature were retained for Model 2 (fig. 7).

The AIC values of the two Area 1 models are almost equal (*Model1AIC*= 42.09; *Model2AIC*= 41.91), while for Area 2, *Model1AIC* has an AIC of 56.94 and *Model2AIC* has an AIC of 52.90. AIC weight was also computed (Wagenmakers & Farrell, 2004). The results indicates that Area 1's models perform almost equally (AIC weight Model 1= 0.478006; AIC weight Model 2= 0.521994), while for Area 2, Model 2 appears to be significantly more reliable than Model 1 (AIC weight Model 1= 0.117272; AIC weight Model 2= 0.882728).

To further test the performance of the models, Receiver Operator Characteristic (ROC) curve was plotted. Using *roc* and *ggroc* functions (*pROC* package), True Positive Rate (TPR or Sensitivity) was plotted against False Positive Rate (FPR or Specificity). Sensitivity is the percentage of correctly predicted events, while specificity is the percentage of events which were incorrectly predicted as non-events. The more efficient a model prediction is, the closer the curve is to the top left corner of the plot. The Area Under the ROC Curve (AUC) was also computed with *auc*. In a range from 0 to 1, it represents the odds of a random event to be closer to 1 than a random non-event. If a model prediction is not more likely than random probability AUC value approximates 0.5; a model with only correct predictions has an AUC value of 1.0. The AUC values show that all the models (in both areas) are fairly reliable (Area 1: Model 1 AUC= 0.88, Model 2 AUC= 0.88; Area 2: Model 1 AUC= 0.79, Model 2 AUC = 0.85).

The coefficients obtained through the *stepAIC* function were used to produce a predictive map representing the logarithm of the odds for each model. To create these maps, GRASS *r.mapcalc* was used. Raster maps of the selected covariates were weighted based on their regression coefficient and combined through map algebra; log-odds were then converted to probability using:

$$\text{PRED} = (\exp(\text{PROB MAP}))/(1+(\exp(\text{PROB MAP})))$$

Each 20×20 m cell of each map was assigned a value from 0.0 (non-event) to 1.0 (event) (fig. 8).

3.7 Evaluation of the paleo-riverbeds

The last phase in the analysis was to assess the probability of each of the many paleo-riverbeds known in the two areas to date back to the Roman period.

The first step was to identify the threshold value where the prediction of event and non-event in the models is exactly the same. Correct event and non-event predictions for the training sample (sensitivity and specificity of the ROC) were plotted against different predictive cutoffs (from 0 to the maximum predictive value). The cutoff where sensitivity and specificity are equal, represents the threshold above which each prediction can be classified as an event. Due to the small number of events and non-events in the two areas, the threshold values were found to be very low: 0.24 for both Area 1 models, 0.1 for Area 2-Model 1 and 0.16 for Area 2-Model 2. However, as discussed below, this issue does not seem to compromise the analysis results.

A polygonal shape file was created for each paleo-channel known in the two computational areas. A total of 10 polygons were created for Area 1 and 13 for Area 2 (fig. 9). Some of the polygons do not exactly correspond to paleo-channels recognised in the area, but were created by joining different traces probably pertaining to the same paleo-riverbed. For instance, Area 1-PC3 was created because a core sample identified layers referred to a Roman riverbed between two paleo-channels recognised by Marcello Cosci (Basile, 2021). Some others, such as Area 1-PC1, are composed of two separate polygons, as part of the paleo-riverbeds lies outside the computational area.

Subsequently, 120 locations were randomly created inside each paleo-channel and the corresponding probability map value was sampled for each location using GRASS *v.what.rast*

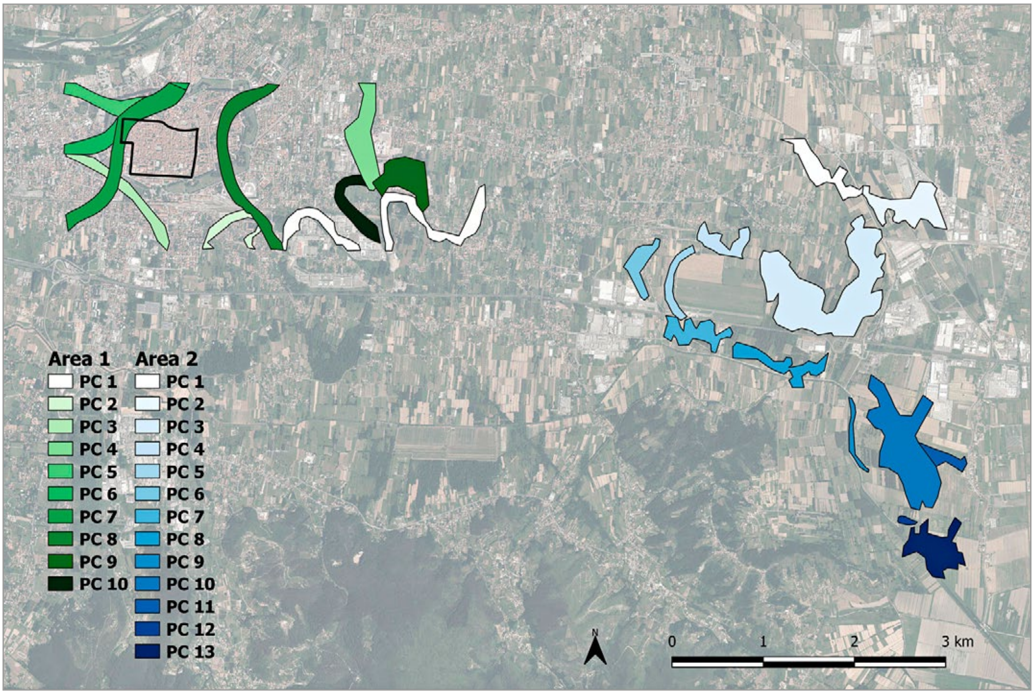

fig. 9. Areas 1 and 2 paleo-channels.

function. Median and mean values of each group of locations were then computed, and their relationship with the aforementioned threshold assessed. If they laid above the threshold, the corresponding paleo-channel was likely to be active during the Roman period, as most of the locations inside its area had values which correspond to events in the probability map. This step was thus fundamental in the analysis process, as it provided a semi-formal and generalisable approach to interpret the probability maps and address the main research questions of this paper.

4. Discussion

4.1 Area 1 (fig. 10)

Despite the lack of information and the above-mentioned issues with data quality, the analysis produced interesting results which provided a new perspective on the area.

For many of the analysed paleo-channels – 4, 5, 6, 7 and 8 – the median and mean values of the randomly sampled locations are clearly below the threshold value. Hence, it can be tentatively argued that they were not active during the Roman period. This inference is partially confirmed by the fact that most of these palaeo-channels are located in areas which were occupied by Roman structures or by part of the Roman city (fig. 11).

On the contrary, median and mean values of points sampled inside PC1 and PC3 are strongly above the threshold according to both Model 1 and Model 2. Although with smaller values, PC9 and PC10 values are also above the threshold.

Finally, PC2 values are below the threshold according to the map produced from Model 1, while they are above the threshold according to Model 2.

Based on these outcomes, it can be argued that the Antraccoli area was indeed crossed by the river, although it is hard to determine which of the paleo-channels were active during

Paleo-channel	Model 1			Model 2		
	Median	Mean	IQR	Median	Mean	IQR
1	0.5586364	0.4881346	0.4656892	0.8552318	0.7972158	0.6132091
2	0.1941294	0.2159363	0.1768567	0.4105249	0.4851939	0.1221414
3	0.7061375	0.5847845	0.4196028	0.7098862	0.5601318	0.7342603
4	0.2155044	0.1757747	0.4337822	0.1568109	0.1689215	0.04900467
5	0.0180024	0.0892193	0.08126488	0.0161265	0.0589166	0.04037971
6	0.0303235	0.1700721	0.2946486	0.1254997	0.2485901	0.4389702
7	0.0423789	0.166889	0.2199202	0.1453036	0.2676836	0.4781042
8	0.211145	0.2612404	0.391171	0.1420358	0.2668767	0.378515
9	0.5758731	0.4897426	0.2616781	0.395779	0.443923	0.1923091
10	0.2745388	0.3433584	0.4106928	0.5222946	0.5494321	0.5563728

fig. 10. Area 1: medians, means, and interquartile ranges of 120 randomly created locations inside each paleo-channel.

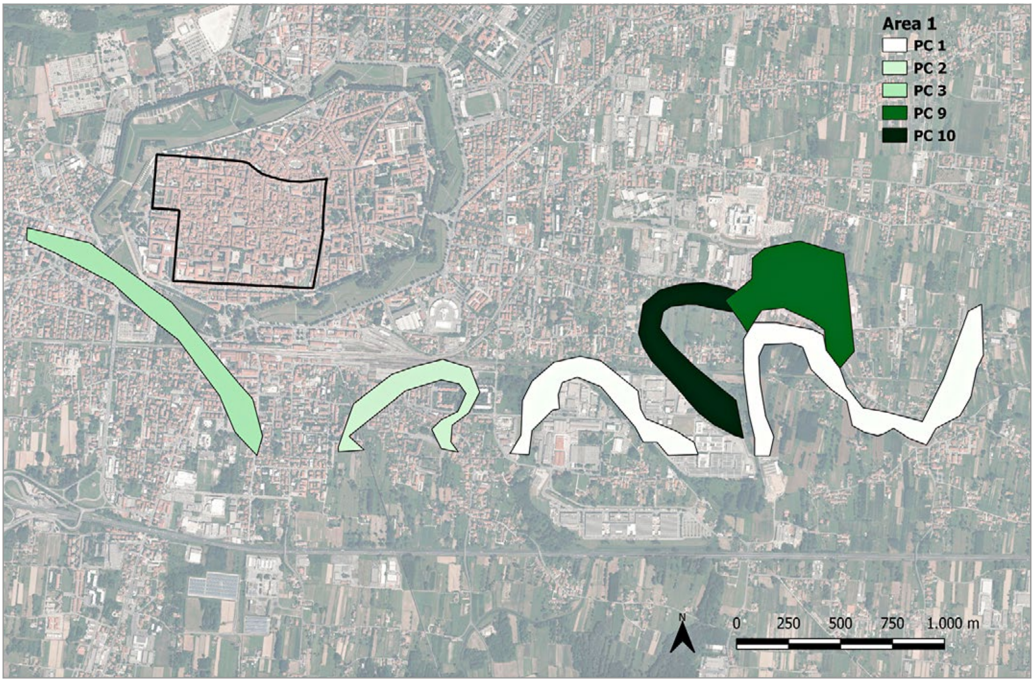

fig. 11. Area 1: paleo-channels that were active during the Roman period.

the Roman period. The tributary branch coming from the north (PC4), was probably active in other periods. Perhaps, it should be identified with the Etruscan age riverbed recently excavated in the area of San Filippo on the North-West of Antraccoli (Giannoni, 2015). Despite the difficulties of understanding whether PC2 should be dated to the Roman period, clearly the river continued to the West, following the path of one of the two paleo-channels which are unfortunately outside of Area 1. A branch of the river – possibly a smaller one – lapped the South-Western corner of the city. It can be suggested that this branch connected the northern and the southern branches of the *Auser*. A comparison may come from the nearby city of Pisa, where the *Auser* and Arno rivers had to be connected by a minor branch West of the city (Bini et al., 2015).

Paleo-channel	Model 1			Model 2		
	Median	Mean	IQR	Median	Mean	IQR
1	0.2423601	0.2485278	0.1468633	0.2954586	0.298737	0.0844781
2	0.04638543	0.0577693	0.03777614	0.00023287	0.01909217	0.008118179
3	0.250739	0.2352394	0.2402206	0.4353118	0.3611757	0.4381106
4	0.19654	0.1927269	0.03684189	0.2440297	0.2329417	0.1623897
5	0.1601578	0.165678	0.02755298	0.157643	0.2429819	0.2674533
6	0.0893022	0.0936842	0.03047885	0.1173395	0.1438265	0.1263995
7	0.1534095	0.1748355	0.06270533	0.5154961	0.4725372	0.2341242
8	0.4897526	0.4718929	0.111995	0.0781908	0.1875101	0.2748041
9	0.2099953	0.2416574	0.1887132	0.5345205	0.4812345	0.2016655
10	0.2527844	0.2759766	0.2007127	0.0243761	0.1144051	0.1802818
11	0.2312949	0.2162588	0.08910351	0.0092319	0.0001807	0.00009896
12	0.1015544	0.1037354	0.02166359	0.4115492	0.4109367	0.05733848
13	0.1941618	0.1867138	0.03533604	0.2790608	0.3184694	0.3051431

fig. 12. Area 2: medians, means, and interquartile ranges of 120 randomly created locations inside each paleo-channel.

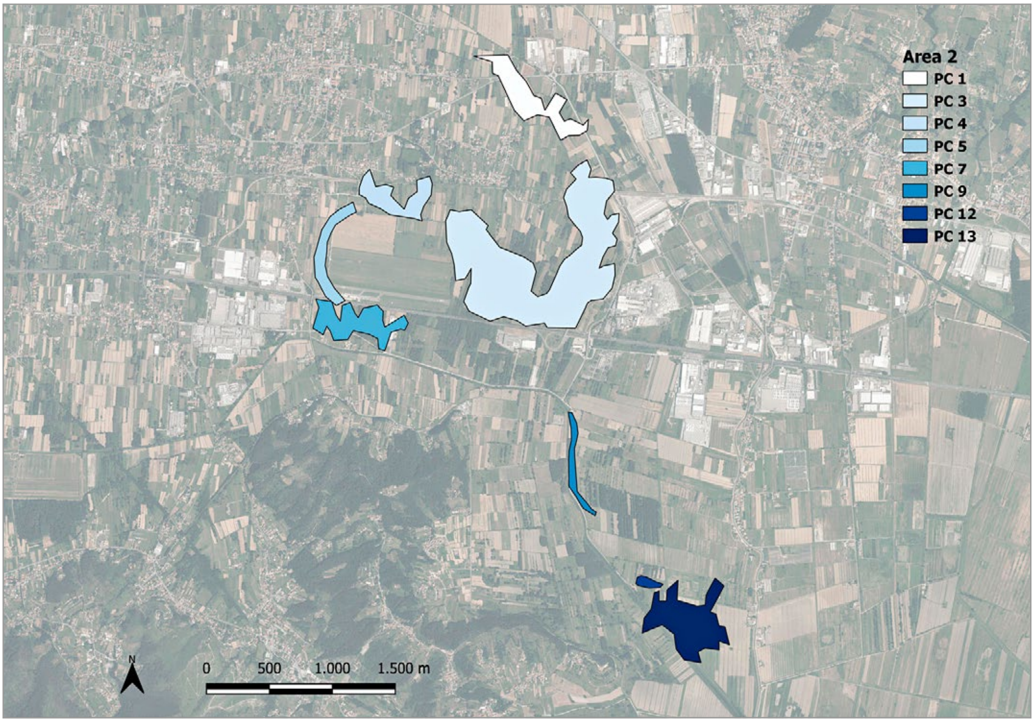

fig. 13. Area 2: paleo-channels that were active during the Roman period.

As far as the area north to the city is concerned, none of the examined paleo-riverbeds dates to the Roman period. It can be argued that the river path might have been similar to today. As both the shape of the defensive walls and the presence of sandy-gravelly layers North to the city suggest, it is possible that a minor branch of the river flowed in this area just before the foundation of the colony. This branch did not exist anymore during the Imperial period, as the construction of the amphitheatre (late 1st century AD) and of other structures suggest.

4.2 Area 2 (fig. 12)

As Model 2 proved to be more reliable than Model 1 for Area 2, data obtained from Model 2 were given a greater weight in evaluating whether the paleo-riverbeds can be dated to Roman times or not.

In the northernmost part of the area, the path of the river essentially corresponds to the reconstruction already proposed by other scholars (paragraph 2.2). It is thus not surprising that PC2's median and mean are both below the threshold values (fig. 13). Flowing towards South, the river describes a large bend in the Tassignano area (PCs 1, 3, 4, 5, 7). One of the research questions for this area was to assess whether the bend described a large curve (PC6) or a narrower one (PC5). Based on the results of the analysis, both models suggest PC5 as the most probable Roman riverbed. The river thus crossed the area of Tassignano airport, where a Late Republican farm was excavated approximately 20 m to the East of the paleo-channel (Giannoni, 2018).

Model 2 suggests that PC8 cannot be considered active in the Roman period. However, based on the available data, this is the only trace that suggests a connection between the northern and the southern path of the river. The outcome of the model can be tentatively interpreted as inaccurate prediction, perhaps related to recent transformations in the morphology of this sector of the plain. From here, the river descended towards the south to the Fossa Nera area. As mentioned above (paragraph 2.2), another research question of this study was to understand the relationship between the river and the two farms excavated in this area. According to Model 2, PC10 median and mean are strongly below the threshold, while PC9 and PC12 are above it. Therefore, it can be assumed that the river flowed between the two farms, as Giulio Ciampoltrini already proposed (Ciampoltrini, 2009, p. 25).

Both Tassignano and Fossa Nera areas show a strong relationship between the river and some of the Late Republican farms known in the plain. Nevertheless, several other buildings along the river path support this reconstruction (fig. 14). Due to the many factors that can

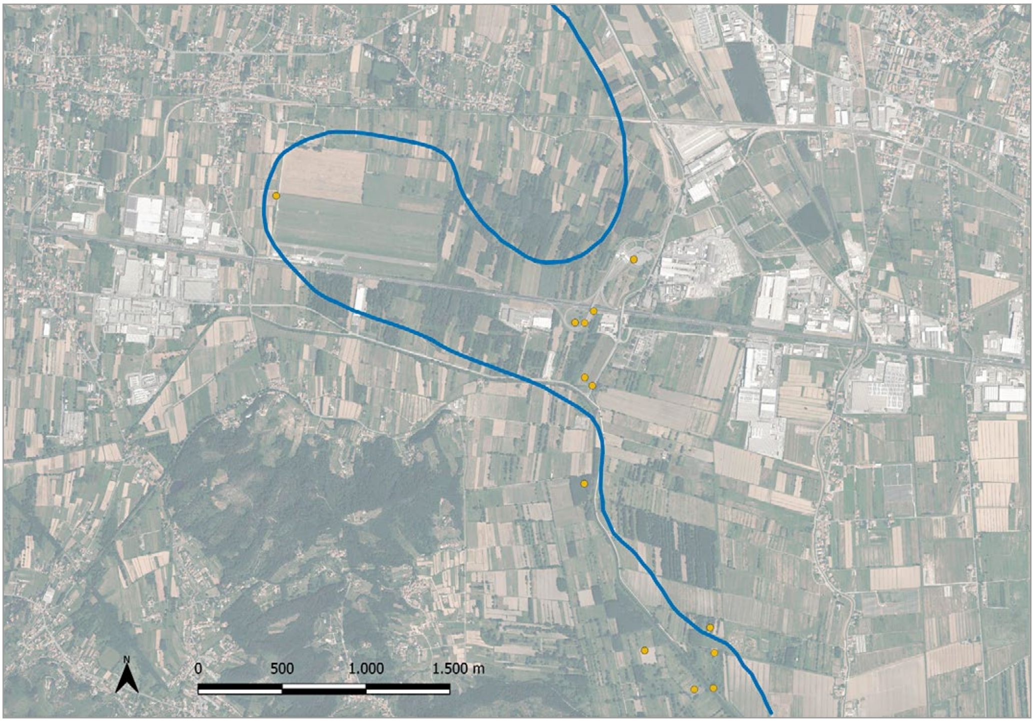

fig. 14. Area 2: reconstruction of the river Auser in the Roman period. The yellow dots represent Roman buildings inside Area 2.

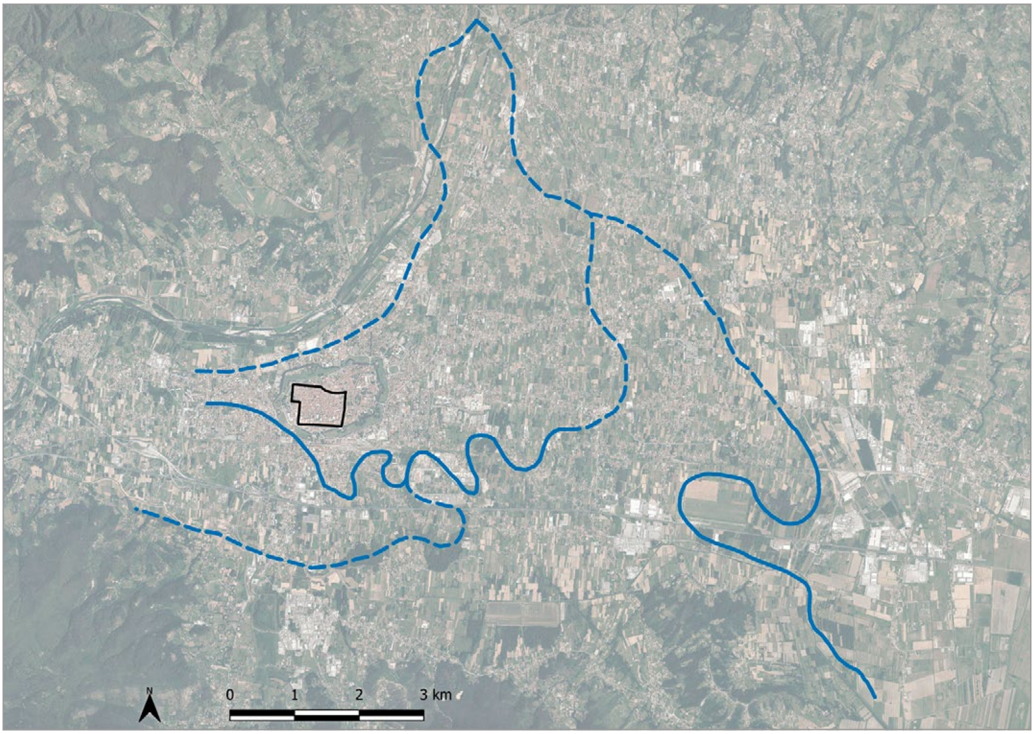

fig. 15. Reconstruction of the river branches during the Roman period. The blue solid lines indicate the paths reconstructed through this study. The blue dashed lines indicate the hypothetical paths outside the 2 study areas.

determine the distribution of farms during the Roman period, further studies need to be carried out to understand the role played by the river in this process.

5. Conclusion

This study was aimed at testing a new methodology for assigning traces of palaeo-channels to a given period based on selected geomorphological variables.

The accuracy of the analysis is affected by research intensity and research strategy biases in the area. The low reliability and non-standardization of data also contributed to reducing the analysis accuracy. Nevertheless, according to the analysis results, it was possible to identify those paleo-channels more likely to be active during the Roman period. The subsequent reconstruction of the possible *Auser* path excludes paleo-channels that were likely to be inactive in the Roman period, such as PC6 and 7 Area 1 (which cross part of the walled city) or PC 4 and 8 Area 1 (where Roman structures were identified).

Moreover, the paleo-channels that can be dated to the Roman period create a coherent river path (fig. 15).

Despite all the dataset quality issues, this methodology offers a new approach for investigating the spatial pattern of landscape features in the past, using archaeological and geomorphological data. Far from proposing a conclusive reconstruction of Lucca's ancient landscape, this research complements other studies and contributes to strengthening our knowledge of the paleo-fluvial system of the plain. Despite the large amount of research conducted in the last 40 years, the absolute lack of GIS-based studies and spatial analysis and modelling is noticeable. In fact, the use of GIS was openly criticised in a recent paper by Giulio Ciampoltrini himself (Ciampoltrini, 2020, p. 11). Nevertheless, these approaches are today widely employed in landscape archaeology, and their crucial role in archaeological research is widely acknowledged (Gillings et al. 2020). These methodologies should be used together with more classical ones, to obtain new and more complex pictures of the past interactions between humans and landscapes and to analyse archaeological and environmental data from a different (more formal and larger-scale) point of view. In conclusion, by recognising Roman period's paleo-riverbeds, this study allows to gain new insights into the analysis of human behaviour and settlement strategies in this area, and to expand the use of spatial analysis in the archaeology of fluvial landscapes.

References

Alberti, G., Grima, R., & Vella, N. C. (2018). The use of geographic information system and 1860s cadastral data to model agricultural suitability before heavy mechanisation. A case study from Malta. *PloS ONE*, *13*(2). https://doi.org/ 10.1371/journal.pone.0192039

Ambrosio, M., Del Sordo, A., Fagioli, M. T., Giannecchini, R., Puccinelli, A., & Sartelli, A. (2010). Analisi dell'evoluzione piezometrica nel periodo 2007-2008 nel settore meridionale della Piana di Lucca (Toscana), area di crisi idrica conclamata. In A. Lenzi, L. Leoni, C. Baldacci et al. (Eds.), *Codice Armonico 2010, Terzo Congresso Di Scienze Naturali Ambiente Toscano*, 224-234. Edizioni ETS.

Anichini, F., Fabiani, F., Gattiglia, G., & Gualandi, M. L. (2012). Un database per la registrazione e l'analisi dei dati archeologici. *MapPapers*, *2*, 1-20. https://doi.org/10.4456/MAPPA.2012.01

Ballerine, C. (2017). *Topographic Wetness Index Urban Flooding Awareness Act Action Support, Will & DuPage Counties, Illinois*. Illinois State Water Survey. http://hdl.handle.net/2142/98495

Barsocchini, D. (1837). *Memorie e documenti per servire all'istoria del Ducato di Lucca: Raccolta di documenti per servire alla storia ecclesiastica lucchese. Vol. V*. Pacini Fazzi.

Barsocchini, D. (1853). Sull'antico corso del Serchio. *Atti Della Reale Accademia Lucchese*, *14,* 393-487. Accademia Lucchese.

Basile, S. (2021). L'Auser e l'*ager Lucensis*: Analisi spaziali per una ricostruzione delle dinamiche tra fiume e insediamento. In F. Fabiani & G. Gattiglia (Eds.), *Atti Della Giornata Di Studi Paesaggi Urbani e Rurali in Trasformazione. Contesti e dinamiche dell'insediamento alla luce del dato archeologico* (Pisa, 22 novembre 2019) (pp. 35-46). Archaeopress.

Bini, M., Rossi, V., Amorosi, A., Pappalardo, M., Sarti, G., Noti, V., Capitani, M., Fabiani, F., & Gualandi, M. L. (2015). Palaeoenvironments and palaeotopography of a multilayered city during the Etruscan and Roman periods: Early interaction of fluvial processes and urban growth at Pisa (Tuscany, Italy). *Journal of Archaeological Science*, *59*, 197-210. https://doi.org/10.1016/j.jas.2015.04.005

Bini, M., Zanchetta, G., Regattieri, E., Isola, I., Drysdale, R. N., Fabiani, F., Genovesi, S., & Hellstrom, J. C. (2020). Hydrological changes during the Roman Climatic Optimum in northern Tuscany (Central Italy) as evidenced by speleothem records and archaeological data. *Journal of Quaternary Science*, *35*(6), 791-802. https://doi.org/10.1002/jqs.3224

Carrer, F. (2013). An ethnoarchaeological inductive model for predicting archaeological site location: A case-study of pastoral settlement patterns in the Val di Fiemme and Val di Sole (Trentino, Italian Alps). *Journal of Anthropological Archaeology*, *32*, 54-62. https://doi.org/10.1016/j.jaa.2012.10.001

Ciampoltrini, G. (2008). La porta e la torre: Nuovi materiali per le mura (e l'urbanistica) di Lucca romana. *Rivista di Topografia Antica*, *XVIII*, 23-33.

Ciampoltrini, G. (2020). *L'area urbana di Lucca. Repertorio illustrato dei contesti archeologici d'età romana.* I Segni dell'Auser.

Ciampoltrini, G. (2009). Paesaggi e comunità di una colonia latina. Liguri, Etruschi, Romani nel territorio di Capannori fra III e I secolo a.C. In G. Ciampoltrini & A. Giannoni (Eds.) *La terra dell'Auser. II. Le ricerche archeologiche in località Frizzone e il territorio di Capannori in età romana.* (pp. 13-46). I segni dell'Auser.

Ciampoltrini, G. (2016). La griglia di Igino: Nuovi materiali per la centuriazione di Lucca. *Atlante tematico di topografia antica, 26*, 233-242. https://doi.org/10.1400/260106

Ciampoltrini, G. (2005). Culture in contatto. Etruschi, Liguri, Romani nella valle del Serchio fra IV e II secolo a.C. In G. Ciampoltrini (Eds.), *I Liguri della Valle del Serchio tra Etruschi e Romani. Nuovi dati e prospettive di valorizzazione* (pp. 15-66). Tipografia Menegazzo.

Ciampoltrini, G.. (2010). *Edilizia rurale tra Valdarno e Valle del Serchio: La colonizzazione etrusca tra VI e V secolo a.C. e le deduzioni coloniali d'età tardorepubblicana*. In M. Bentz, Ch. Reusser (Eds.), *Etruskisch-italische und römisch-republikanische Häuser, Atti del Convegno (Bonn 2009)* (pp. 135-144). Reichert Verlag.

Ciampoltrini, G., & Andreotti, A. (2008). *Tra «ager centuriatus» e «silva». Ricerche sul «decumanus» del Colmo dei Bicchi-Botronchiò nella piana di Lucca*. I segni dell'Auser.

Ciampoltrini, G., Cosci, M., & Spataro, C. (2009). I paesaggi d'età romana tra ricerca aerotopografica e indagini di scavo. In G. Ciampoltrini, A. Giannoni (a cura di), *La Terra dell'Auser I. Lo scavo di Via Martiri Lunatesi e i paesaggi d'età romana nel territorio di Capannori* (pp. 13-62). I segni dell'Auser.

Ciampoltrini, G., & Zecchini, M. (2007). *Gli Etruschi della Piana di Lucca. La via del Frizzone e il sistema di insediamenti fra VIII e V secolo a.C.* San Marco Litotipo.

Cosci, M. (2005). Dal monte al mare: Evoluzioni idrografiche dell'antico fiume Auser rivelate dai sensori satellitari. In G. Ciampoltrini, M. Zecchini (Eds.), *Le Dimore dell'Auser. Archeologia architettura ambiente dell'antico Lago di Sesto* (pp. 9-16). San Marco Litotipo.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27-46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Fabiani, F., & Gattiglia, G. (2012). The digital archiving structure. In Anichini F., Fabiani F., Gattiglia G., Gualandi M. L. (Eds), *Mappa. Vol. I. Methodologies applied to archaeological potential predictivity* (pp. 43-72). Nuova cultura.

Giannoni, A. (2015). Il suburbio orientale di Lucca in età romana: Nuovi dati dalle attività di archeologia preventiva per la viabilità dell'Ospedale San Luca. *Notiziario della Soprintendenza per i Beni Archeologici della Toscana*, *11* (saggi), 107-122. https://doi.org/10.1400/244080

Gillings M., Hacigüzeller P., & Lock G. (2020) (Eds.). *Archaeological Spatial Analysis. A Methodological Guide.* Routledge.

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. *Polytical Analysis*, *9*(2), 137-163. https://doi.org/10.1093/oxfordjournals.pan.a004868

Kolka, R., & Thompson, J. (2006). Wetland geomorphology, soil, and formative processes. *In Batzer D. & Sharitz R. (Eds.), Ecology of Freshwater and Estuarine Wetlands* (pp. 7-42). University of California Press.

Lloyd, C. D., & Atkinson, P. M. (2004). Archaeology and geostatistics. *Journal of Archaeological Science*, *31*, 151-165. https://doi.org/10.1016/j.jas.2003.07.004

Mattivi, P., Franci, F., Lambertini, A., & Bitelli, G. (2019). TWI computation: A comparison of different open source GISs. *Open Geospatial Data, Software and Standards*, *4*. https://doi.org/10.1186/s40965-019-0066-y

Mencacci, P., & Zecchini, M. (1981). *Lucca romana*. Pacini Fazzi.

Midi, H., Sarkar, S. K., & Ranaa, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, *13*(3), 253-267. https://doi.org/10.1080/09720502.2010.10700699

Miller, H. J. (2004). Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers*, *94*(2), 284-289.

Nakoinz, O., & Knitter, D. (2018). *Modelling human behaviour in landscapes: Basic concepts and modelling elements.* Springer.

Nardi, R., Nolledi, G., & Rossi, F. (1987). Geologia e idrogeologia della Piana di Lucca. *Geografia Fisica e Dinamica Quaternaria*, *10*, 132-160.

Paderi, E. (1932). Variazioni fisiografiche del bacino del bientina e della pianura lucchese durante i periodi storici. *Memorie Della Società Geografica Italiana*,*18*, 7-17.

Pebesma, E. J., & Wesseling, C. G. (1998). Gstat: A program for geostatistical modelling, prediction and simulation. *Computers & Geosciences*, *24*(1), 17-31. https://doi.org/doi.org/10.1016/S0098-3004(97)00082-4

Repetti, E. (1839). *Dizionario geografico fisico storico della Toscana, contenente la descrzione di tutti i luoghi del Granducato, Ducato di Lucca, Garfagnana e Lunigiana: Vol. III*. L'autore.

Rossiter, D. G. (2007). *Co-kriging with the gstat package of the R environment for statistical computing.* https://www.itc.nl/rossiter/teach/R/Rck.pdf.

Salvini, R., Guastaldi, E., Coscini, N., & Del Seppia, N. (2006). Ricostruzione del paleoalveo del fiume Serchio (Lucca, Italia) tramite rilievi LIDAR, foto aeree e immagini Quick Bird. *Il Quaternario. Italian Journal of Quaternary Science*, *19*(2), 299-310.

Sommella, P., & Giuliani, C. F. (1974). *La pianta di Lucca romana*. De Luca Editore.

Venables, B., & Ripley, B. (2002). *Modern Applied Statistics With S* (4th ed.). Springer. https://doi.org/10.1007/b97626

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192-196. https://doi.org/10.3758/BF03206482