Archeologia e Calcolatori
34.2, 2023, 87-104
doi 10.19282/ac.34.2.2023.05

INTEGRATING POINT PATTERN ANALYSIS AND LOGISTIC REGRESSION APPROACHES FOR EXPLORING THE SETTLEMENT PATTERN OF THE VERSILIA AND GARFAGNANA MOUNTAINS IN ROMAN TIMES

1. Introduction

Mountain archaeology has a long tradition of study and in recent years new methodologies and theories for studying these landscapes have emerged, also thanks to the widespread development and use of digital technologies for the management of big datasets. Archaeologists have regained interest in this type of environment (Tzortzis *et al.* 2010a) and, over the years, the number of studies investigating highlands with multiscale and interdisciplinary approaches has increased considerably, shedding new light on the dynamics of mountain's communities (Della Casa 2010; Carrer 2013; Migliavacca *et al.* 2021; Visentin, Carrer 2017). Roman archaeology has long underestimated the phenomenon of organising, managing, and settling highland areas, focusing more on the rural settlements in the lowlands, such as farms and *villae*. Nevertheless, interest in settlement dynamics of mountain territories has increased in recent years (Mocci *et al.* 2010).

Renewed efforts aimed at systematically acquiring new data using traditional methodologies or developing new collection methodologies have been more rarely accompanied by a review and digitization of legacy data. Since 2011, the MAPPA Laboratory of the University of Pisa has developed a set of tools for digitizing and managing archaeological legacy data (Anichini *et al.* 2012), starting from the urban area of Pisa and then progressively expanding the area of interest to cover a large part of northern Tuscany. Two recent projects focused on the analysis of the cities and territories of Pisa and Lucca in Roman times have systematically collected, digitised, and managed legacy data, both published and preserved in the Superintendencies' archives. On each side, the lack of systematic investigation in the vast mountainous territories of Versilia and Garfagnana emerged, despite traces of evanescent frequentations seem to indicate a settlement pattern of undoubted interest.

Nevertheless, some studies have had the great merit of recovering and contextualising isolated finds – often the result of random discoveries – in a broader framework, relating them to major urban centres, road infrastructures, and silvo-pastoral agricultural practices (Menchelli 1991; Ciampoltrini 2003, 2006; Fabiani 2006). Similarly, human-environment relationships in historical times have been studied, especially focusing on the plains (Bini *et al.* 2020). So far, however, this territory has never been analysed as a whole, a geo-referenced picture of the archaeological record was missing, and the lack of specific analyses of the settlement-environment relationship prevented from identifying and explaining large-scale settlement patterns. It is therefore this gap that the present study will attempt to fill, integrating the results of Point Pattern Analysis and Logistic Regression approaches to evaluate settlement dynamics in relation to this specific environment, and finally create a predictive map. Indeed, spatial and computational analysis help to identify patterns in big and diverse datasets and are used to assess representativity and biases in data. Statistical methods and predictive modelling can be used to mitigate these biases and restore realistic images of the human-environment dialectic in the formation of mountain landscapes (Kempf, Weaverdyck 2023).

2. Background

2.1 *Study area*

The study area comprises the mountain district of Versilia and Garfagnana (Lucca, north-western Tuscany) (Fig. 1). Versilia is a territorial district in north-western Tuscany between the Apuan Alps mountain ridge and the Tyrrhenian coast, bordered to the N by the Seravezza River and to the S by Forte del Motrone, although usually the territory also includes the Camaiore basin and the coastal plain extending to Viareggio. The narrow plain is morphologically homogeneous and gently sloping towards the sea, generally standing at elevations slightly above or below zero. The coastal ridge of the Apuan Alps and Mount d'Oltre Serchio reaches higher altitudes in the N, up to 900-1000 m, and progressively less marked towards the S, where it is around 200-400 m. The reliefs delimit the Versilia plain with steep slopes or wide alluvial conoids at the downstream outlet of the numerous streams. Most of the watercourses are torrential (Versilia, Cinquale, Camaiore Ditches, Motrone, Viareggio Canal), and relatively few are not overly affected by seasonal variations in rainfall, such as the Frigido River, Serchio River, and Magra River (Devoti *et al.* 2003, 73-76; Fabiani 2006, 19). Garfagnana stretches from the eastern ridge of the Apuan Alps to the crest line of the Apennine and between the Magra valley to the N and the plain of Lucca to the S.

From a geomorphological point of view, the eastern Apuan ridge is characterised by extensive rock formations of terrigenous nature that have produced less marked slopes than the Versilia side (Carmignani *et al.* 1978). The hydrographic evolution of the Serchio River passed through a Plio-Pleistocene fluvial-lacustrine phase with the development of intravalley plains where the resumption of erosive phenomena on alluvial deposits gave rise to vast terraced forms (Boccaletti *et al.* 1980; Castaldini *et al.* 1998, 416-417).

The highest peaks are concentrated in the northern part: to the W, Mt. Pisanino (1947 m), Mt. Cavallo (1895 m), and Mt. Tambura (1891 m), while

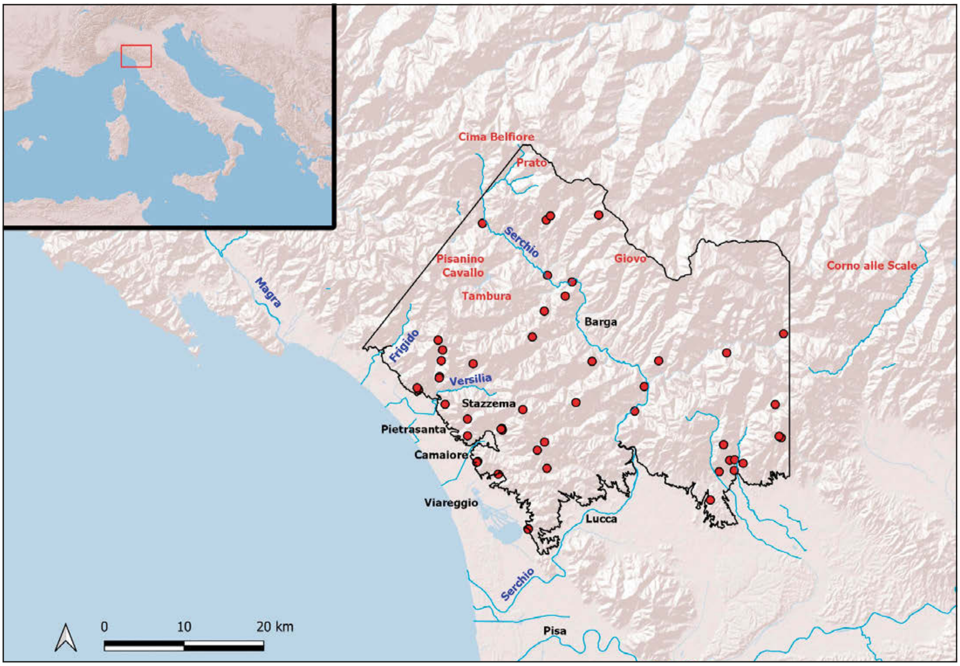

Fig. 1 – Study area and archaeological finds used in this research work.

to the E, several peaks rise above 1800 m (Cima Belfiore 1840 m, Mt. Prato 2054 m, Mt. Giovo 1991 m, Corno alle Scale 1945 m) distributed along the entire length of the watershed with the Po side. Between the two mountain ranges lies the valley of the Serchio River. The shaft of the valley up to the confluence of the Lima stream – the Serchio's main tributary – runs NW-SE, thus parallel to the coastline and the Apuan massif.

The altitude of the valley varies from 450 m a.s.l. at Piazza al Serchio – where the various branches into which the river divides (Serchio di Soraggio, di Sillano, di Gramolazzo) converge – to 220 m a.s.l. at Castelnuovo di Garfagnana, and 100 m a.s.l at Bagni di Lucca.

2.2 *Archaeological background*

The geomorphological conformation of the mountainous terrain and the dense ground cover make the internal areas difficult to access. Furthermore, the extensive abandonment of agricultural practices after World War II (Arnaez *et al*. 2011; Modica *et al*. 2017) makes ploughing rare, limiting the possibility of recognizing surface scattering. Over time, several archaeological features have been reported in these territories; however, most of these are the result of random discoveries or the activity of local enthusiasts and often limited to out-of-context pottery fragments. Studies on long-term landscape and settlement dynamics of these mountains are rare (Ciampoltrini 2003) and archaeological evidence of the Roman period, often related to seasonal activities, has often found only marginal space in the study of general settlement patterns (Giannini 2005; Ciampoltrini 2006).

Frequently, random finds are in fact the only evidence upon which a synthesis of human presence in the area can be based (Menchelli 1991, 387-388). In these cases, the partial or sketchy documentation often prevents a precise location of the finds, the lack of stratigraphic data frequently makes it impossible to specify the contexts of discovery and, in some cases, the materials found are lost without any proper study.

The non-verifiability of much of the available data entails an undeniable reliability issue of the archaeological information, leading to a sampling bias that must necessarily be considered in the analysis and interpretation processes. Nevertheless, the georeferencing of all known archaeological data for these territories provides, for the first time, an assessment of the actual density of the archaeological record in the highlands, which appears to be much higher than expected for marginal areas. These endemic uncertainties of the archaeological record led previous studies to abandon any attempts to read the settlement pattern of the mountain territory in its informative completeness and its connection with the other landscape constituents. This study stems from these reflections and aims to, at the very least, begin to fill this gap.

3. Materials and methods

3.1 *Archaeological dataset and co-variables*

The dataset derives from the collection, digitisation and systematisation of data performed as part of broader projects to analyse the settlement patterns of the cities and territories of *Pisae* and *Luca* in the Roman period. During these projects, 1026 archaeological intervention records and more than 1420 finds were catalogued for the *Pisae* territory (Campus 2022) and 427 intervention records, for a total of 1196 finds for *Luca*'s territory (Basile 2022a). An 'intervention' is every single action carried out in a specific location, from excavations and surveys to remote sensing, cores, random findings, and inspections by Superintendence. 'Finds' are classified with an increasing level of abstraction, from the traces in the field to their categorization into functional macro typologies (for a more detailed description of 'interventions' and 'finds' see Anichini, Gattiglia 2012). Given the purpose of the research, only records falling in the hilly and mountainous areas of Versilia and Garfagnana were considered in this study. Therefore, a target area was prepared using the contour line tool, selecting the portion of the territory within the isohypse of 50 m a.s.l.

The study area thus selected covers approximately 1574 km² and includes 78 finds, then reduced to 55 by using the R function remove.duplicates to find and delete points with the same coordinates (Fig. 1). The choice of independent variables is a crucial preliminary step in the creation of the model. Considering that this study focuses on the distribution of settlement in a specific mountainous environment, we deliberately decided to select as predictor for our models only those environmental variables that could have an influence on the settlement pattern, leaving out cultural variables such as toponymy or proximity to mountain passes and routes.

Among those selected are geomorphological and pedological variables. Geomorphological variables include Digital Terrain Model (*DTM*), Slope (*Sl*) Sine and Cosine of the Aspect (*sin_aspect*, *cos_aspect*), Profile Curvature (*curv*), and Tangent Curvature (*curv_tan*). Pedological variables derived from the Tuscany Region Pedological Database[1] include land use capacities such as Drainage (*Dre*), Erosion (*Eros*), Chemical Fertility (*Fert*), Landslide (*Slide*), and the percentage composition of soils: Clay (*Cla*), Sand (*San*), Silt (*Sil*), Pebbles (*Peb*), Rockiness (*Rock*), Organic Substrate (*Sostorg*). Furthermore, we also considered Distance from major waterways (*St_dist*), Total Solar Exposure, and Exposure Time measured at the summer (*Sotosu* and *Sotisu*) and winter (*Sotowi* and *Sotiwi*) solstices as possible predictors.

In a first step, the interaction between archaeological finds and variables were analysed studying the first-order effects of the point pattern. In a second phase, two regression models were constructed with selected variables using different approaches to create a predictive map of the area: a first model (Model_1a) only considers variables selected after the Point Pattern Analysis; a second (Model_2) uses variables selected and combined manually through an assessment of P-values and Akaike Information Criterion (AIC) of various models.

3.2 *Point Pattern Analysis*

Point Pattern Analysis is a method increasingly used in archaeology to provide a reliable statistical assessment of landscape and settlement dynamics (Eve, Crema 2014; Brandolini, Carrer 2020; Costanzo *et al.* 2021; Basile 2022b; Campus 2022). Point Pattern Analysis generally refers to a suite of statistical methods designed to assess the potentially complex spatial relationships that might exist among entities that can be described as points. The underlying processes behind a given point pattern are determined either by interaction with a range of exogenous influences (*first-order effects* or *induced spatial dependency*) or by intrinsic factors to the phenomena of interest (*second-order effects* or *inherent spatial dependency*) (Crema 2020, 158). In this paper, Point Pattern Analysis performed in R using the spatstat package (Baddeley *et al.* 2021) will focus on attempting to formally

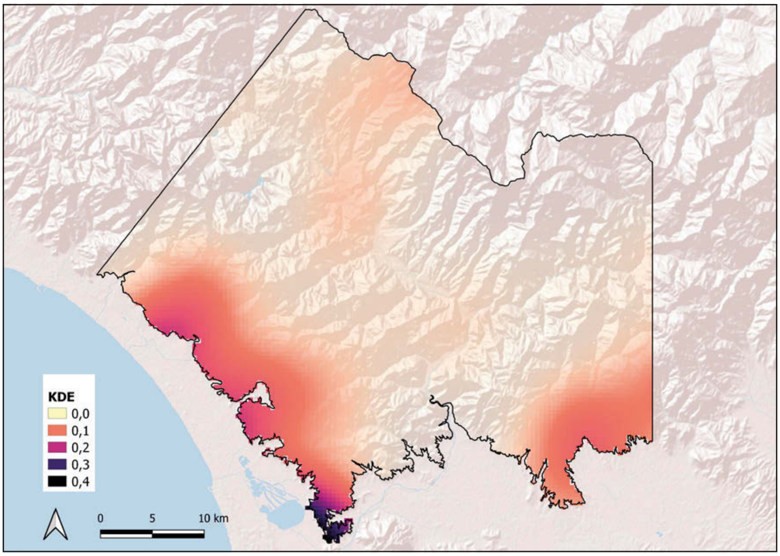

Fig. 2 – Kernel Density Estimation of the Roman period archaeological finds within the study area.

model the exogenous environmental variables that may have induced spatial dependence of the study area evidence. In other words, we will study the large-scale spatial interaction to determine whether the density of the point pattern in the study area, proportional to the intensity of the underlying process, is stationary and isotropic (Homogeneous Poisson Process - HPP) or spatially variable (Inhomogeneous Poisson Process - IPP), assessing whether an inhomogeneous model describes spatial variability more accurately than the stationary homogeneous Poisson model by fitting external covariates that might influence the distribution of spatial events (Brandolini, Carrer 2020, 5). A popular method for summarizing the first-order intensity of a point pattern is to create a density surface using Kernel Density Estimation, which computes a continuous approximation of the distribution by weighting events relative to their distance from the point from which the intensity is estimated (Conolly, Lake 2006, 175-177; Bevan 2020, 63-64).

As is clearly shown in the figure (Fig. 2), the density of points in the study area is variable and inhomogeneous, showing a greater number of points in particular areas, such as in the foothills. Following the principle of parsimony, the model was created by selecting the combination of covariates that minimises the Akaike Information Criterion (AIC) values to simplify the model without affecting performance. In this way, model 1 was created, the results of which are presented in the table (Tab. 1).

model	Covariate	Estimate	S.E.	CI 95% lo	CI 95% hi	Z test	Z-value
1	Intercept	-9,074267965	3,377345034	-15,693740000	-2,454793000	**	-2,686805000
1	DTM	-0,000901334	0,000465451	-0,001813602	0,000010934		-1,936474000
1	Cos aspect	-0,469826506	0,205848019	-0,873281200	-0,066371800	*	-2,282395000
1	Tangent Curvature	11,331528590	3,925082210	3,638509000	19,024550000	**	2,886953000
1	Slope	-0,074725602	0,023749200	-0,121273200	-0,028178030	**	-3,146447000
1	Distance from streams	0,000235373	0,000113971	0,000011993	0,000458752	*	2,065193000
1	Erosion	-0,161707472	0,109905088	-0,377117500	0,053702540		-1,471337000
1	Landslide	0,201112572	0,141027606	-0,075296460	0,477521600		1,426051000
1	Organic substrate	0,179231863	0,040165758	0,100508400	0,257955300	***	4,462305000
1	Solar total summer	-0,000881674	0,000398235	-0,001662199	-0,000101148	*	-2,213956000
1	Solar total winter	0,000351656	0,000187218	-0,000015284	0,000718596		1,878328000

Tab. 1 – Coefficients of the Point Pattern Analysis.

3.3 *Logistic Regression Modelling*

Regression models are among the most widely used approaches in archaeology to predict the relationship between the probability of encountering an archaeological site and several independent variables. Logistic regression is used to specify a binary outcome (event/non-event, present/absent, large/small, etc.): hence, logistic regression has mainly been used for the development of predictive models on the location of archaeological sites, estimating the probability that a site is present in a particular study area (Carlson 2017, 235; Nakoinz, Knitter 2018, 87-97). In our case, logistic regression was used to model the probability of encountering a frequented (event) or a non-frequented (non-event) area and to test the relationship between the presence of settlement elements and the covariables already considered for the Point Pattern Analysis.

The 55 points of our dataset were considered as 'events', assigning each of them numerical value of 1. Considering the low density of investigation in the area, it was not possible to establish points or areas that could represent 'non-events' with certainty. Therefore, 350 'non-event' points representing a process of complete spatial randomness were obtained with the QGIS Random Points tool, resulting in a ratio of 1:6 between 'events' and 'non-events' entities, as already tested in archaeology (Wachtel 2018; Li *et al*. 2022). Although we are aware of the reduced representativeness of the dataset, the event/non-event ratio, more than the total number of measures, is the most relevant parameter for structuring a robust binomial model (King, Zeng 2001).

To avoid collinearity effects in the estimation of the coefficients, the correlation between the predictors was assessed before the models were created by calculating the Pearson correlation index for each pair of variables (Fig. 3). A strong collinearity above the 0.7 threshold value (Alberti *et al*. 2018, 13) was thus established between some variables (positive: *Sotiwi-Sotowi*, *Sotiwi-Sotosu*; negative: *Sin-aspect-Sotiwi*, *Sin-aspect-Sotowi*, *Slope-Sotosu*). Even though the two variables Slope and Total Summer Solar Exposure (*Sotosu*) have a coefficient of 0.71 – just above the threshold value – we still decided

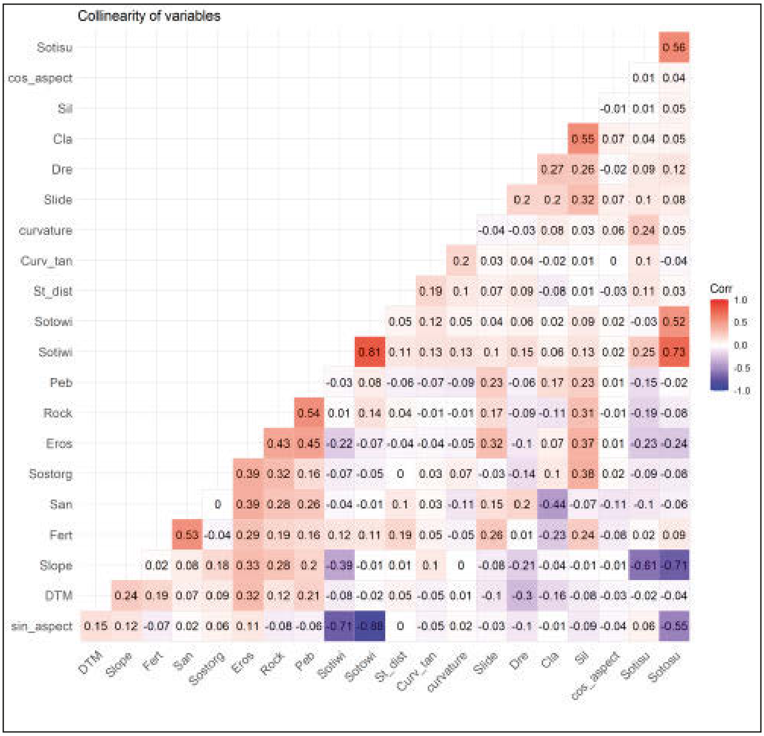

Fig. 3 – The Corrplot which shows collinearity among variables.

not to exclude them from the model. In fact, in our interpretation, these could be two of the most influential factors in establishing the settled areas. Therefore, we were interested in assessing their true impact in predicting sites.

The first model (Model 1a), created using the R glm function, only considers the variables already selected by the Point Pattern Analysis. Using the stepAIC function, a stepwise selection of these variables was performed to obtain the model with the best balance between likelihood and predictors, according to the principle of parsimony (for a similar procedure see: Basile, Carrer 2022, 71-73). The obtained model preserved 6 out of 10 variables: DTM, Slope, Cosine of Aspect, Tangent Curvature, Distance from main streams, and Organic Substrate (Tab. 2). Then, the influence of the interaction between predictors on the model was tested by multiplying two or more variables to obtain two types of information: assessing the possibility of a strong dependence between variables – in which case one of the predictors would lose its importance in the model – and assessing whether the interaction *per se* could be relevant for the model. In this way, we established that

model	Covariate	Estimate	Std. Error	Z-value	Pr(>|z|)	Z test
1a	Intercept	-0,75460	0,45830	-1,64700	0,09964	.
1a	DTM	-0,00135	0,00047	-2,86600	0,00416	**
1a	Slope	-0,02686	0,01363	-1,97100	0,04871	*
1a	Tangent Curvature	19,28000	12,86000	1,49900	0,13390	
1a	Distance from streams	-0,00014	0,00023	-0,60500	0,54498	
1a	Organic substrate	-0,08210	0,10950	-0,75000	0,45331	
1a	Cosine Aspect	-0,63890	0,23330	-2,73800	0,00618	**
1a	Dist_from_st:Org_subst	0,00013	0,00007	2,02200	0,04321	*
2	Intercept	3,19275	2,01486	1,58500	0,11306	
2	DTM	-0,00137	0,00048	-2,86500	0,00417	**
2	Slope	-0,06414	0,01900	-3,37700	0,00073	***
2	Tangent Curvature	14,54413	10,17958	1,42900	0,15308	
2	Distance from streams	0,00025	0,00014	1,86500	0,06215	.
2	Organic substrate	0,12635	0,05251	2,40600	0,01612	*
2	Cosine Aspect	-1,19060	0,35566	-3,34800	0,00082	***
2	Curvature	34,20600	15,59248	2,19400	0,02825	*
2	Landslide	0,26463	0,15306	1,72900	0,08384	.
2	Exposure Time summer	-0,36481	0,14406	-2,53200	0,01133	*
2	Org_subst:Cos_aspect	0,14131	0,06918	2,04300	0,04110	*

Tab. 2 – Coefficients of the Logistic Regression Models.

the interaction between Distance from main streams and Organic Substrate (*St_dist:Sostorg*) was particularly influential.

For comparison purposes, a second model (Model 2) was created by reconsidering all variables without the selection applied by the Point Pattern Analysis. After discarding Sin of Aspect (*sin_aspect*) and Solar Time and Total Exposure in winter (*Sotiwi, Sotowi*) due to their high collinearity, the following predictors were selected by applying the stepAIC function: DTM, Slope, Curvature, Tangent Curvature, Distance to main streams, Organic Substrate, Cosine of Aspect, Landslide, and Solar Exposure Time on the summer solstice (Tab. 2). As for Model 1a, we inspected the interaction among variables by multiplying them. In this case, the interaction between Organic Substrate and Cosine of Aspect (S*ostorg:cos_aspect*) is significant, therefore it was considered among the predictors.

As a first step in validating both models, using roc and ggroc functions (pROC package), Receiver Operator Characteristic (ROC) curves were plotted, to display the True Positive Rate (or Sensitivity) against False Positive Rate (or Specificity) (for a similar procedure see: Basile, Carrer 2022, 73). The former is the percentage of correctly predicted events, while the latter is the percentage of events incorrectly predicted as non-events. Furthermore, using the auc function, we calculated the Area Under the ROC Curve (AUC) to assess the reliability of the predictions. In a range from 0 to 1, the AUC value represents the probability that a random event is closer to 1 than a random non-event. Hence, values close to 1 indicate a high predictive capacity of the model, whereas values below 0.5 indicate low prediction reliability (Li *et al*. 2022, 8).

A chi-square test was then performed subtracting the residual deviance from the null deviance and the residual degrees of freedom from the null degrees of freedom in order to compare the response of the models with only the intercept (null deviance), against the models that include the independent

Model 1a	NO-Event	Event	Sum
NO-Event	224	126	350
Event	17	38	55
Sum	241	164	405
Model 2	**NO-Event**	**Event**	**Sum**
NO-Event	231	119	350
Event	14	41	55
Sum	245	160	405

Tab. 3 – Cross tables of the observed values against the fitted values of the Logistic Regression Models.

variables (residual deviance) and thus determine whether the fitted models represent an improvement over the null hypothesis that the decrease in deviance is not significantly different from zero. Subsequently, a second test was performed comparing the residual deviance and the residual degrees of freedom, with the null hypothesis that the observed values differ significantly from fitted values (Carlson 2017, 237-238).

A series of automated and manual comparisons between true positive (*tpr*) and true negative (*tpn*) prediction rates were performed to establish a threshold value, to balance as best as possible the prediction for non-events and events. Finally, a cross table (Tab. 3) was created to compare the predictions of the regression model with observed values (for a similar procedure: Campus 2022, 391-396).

The performance of the two models was finally compared, using Akaike information criterion (AIC), Schwarz's Bayesian Information Criterion (BIC), and BIC weights. The AIC is normally used to compare alternative models, generally selecting the one with the lowest AIC value (Carlson 2017, 238). The BIC is calculated as the difference between the maximum likelihood of the model and the product of the covariates for the number of observations (points), so the lower the BIC, the better the model performance. The BIC weights are used to provide a normalised estimate of the relative performance of the two models (Brandolini, Carrer 2020, 8).

3.4 *Predictive map*

Using GRASS r.mapcalc function, raster maps of the selected covariates were weighted based on their regression coefficient and combined to produce a probability map representing the logarithm of the odds for each model:

Prob = Intercept+(coef. Covar1×covar1)+(coef.Covar2×Covar2)…

The probability map was then used for both models to obtain a predictive map in which a value from 0.0 (non-event) to 1.0 (event) was assigned to each 20×20 m cell:

Pred = (exp(Prob))/(1+(exp(Prob)))

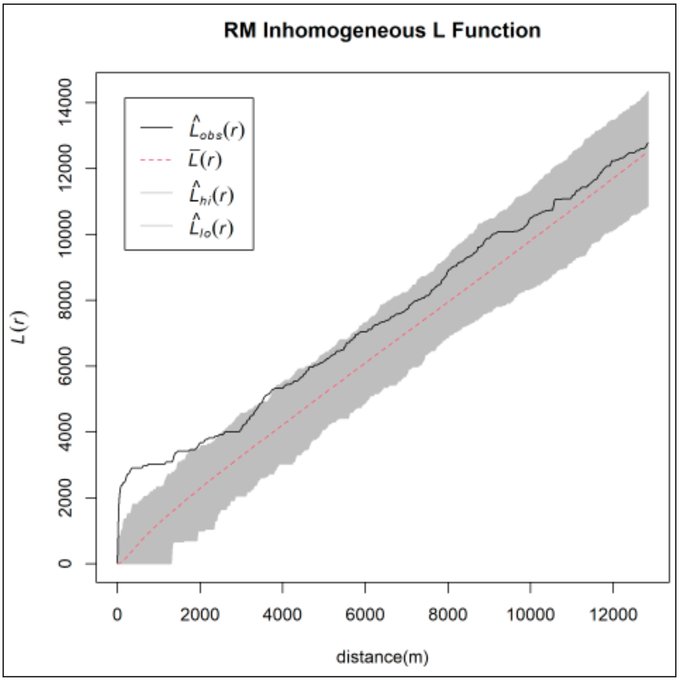

Fig. 4 – Inhomogeneous L-function.

4. Results and discussion

Exploratory analysis such as Kernel Density Estimation (Fig. 2) demonstrates the inhomogeneity of the distribution of archaeological evidence, concentrated on the foothills near the plain, where urban centres and road network were located. Point Pattern Analysis assesses the spatial dependence among points at various scales. The inhomogeneous L-function (Fig. 4) shows a significant aggregation of points – beyond the envelope interval – within 2 km, probably due to the existence of areas with a higher density of finds, resulting in clusters that do not necessarily reflect the actual settlement distribution. On a larger scale, first-order processes – namely environmental features – explain the distribution of points (Model 1). Point Pattern Analysis coefficients show a significant direct correlation with Organic substrate, Landslides and Tangent Curvature, while Distance from streams and Total Solar Exposure in winter display a weaker direct correlation. A very weak inverse correlation with DTM and Total Solar Exposure in summer and a weak one with slope and erosion are observed. In contrast, the most significant inverse correlation is with Aspect Cosine.

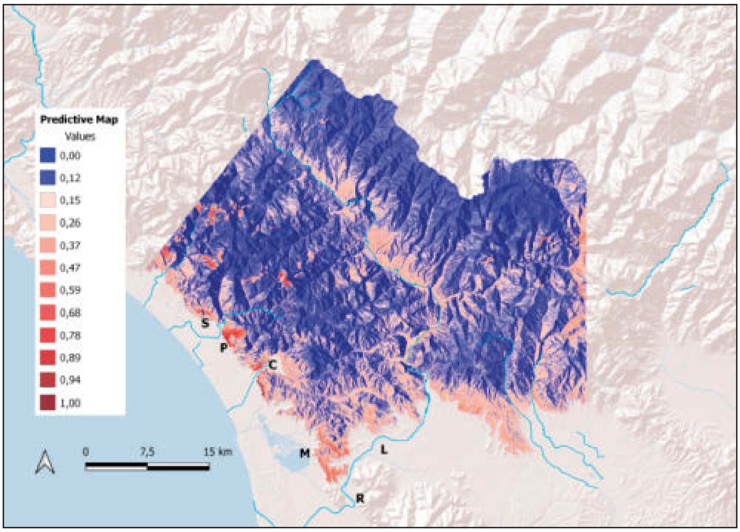

Fig. 5 – Predictive map derived from Model 1a with locations mentioned in the text: L) Lucca; R) Ripafratta; M) Massaciuccoli; C) Camaiore Valley; P) Pietrasanta; S) Strettoia.

As far as the regression models are concerned, in both models, variables with a significant inverse correlation and low P values are DTM, Slope, Cosine of Aspect (Tab. 2); while Distance from Streams and Organic Substrate have higher P values. Among the selected variables Tangent Curvature presents very high estimate coefficients, particularly affecting the predictions. Only Model 2 considers Solar Exposure Time in summer, Landslide and Curvature – the latter of which has a particularly high estimate coefficient. Furthermore, Model 1a considers the interaction between Distance from Streams and Organic Substrate (*St_dist:Sostorg*), and Model 2 the one between Organic Substrate and Cosine of Aspect (*Sostorg:cos_aspect*).

Chi-square tests prove that both models are acceptable and exhibit a significant correlation between the probability of encountering an event and the independent variables. In the first test, comparing the null deviance against the residual variance the p-values for both models are close to 0 (Model 1a: 2.174182e-08; Model 2: 5.176584e-09); similarly, the second chi-square tests return values close to 1 (Model 1a: 0.9999996; Model 2: 0.9999999). We can therefore conclude that the fitted models provide a significant improvement over the null models. For both models, the Area Under the ROC Curve (AUC) values are around 0.75 (Model 1a: 0.754961; Model 2: 0.7695584), thus showing a similar predictive capability. Threshold values were selected by comparing true positive (*tpr*) and true negative (*tpn*) prediction rates, seeking

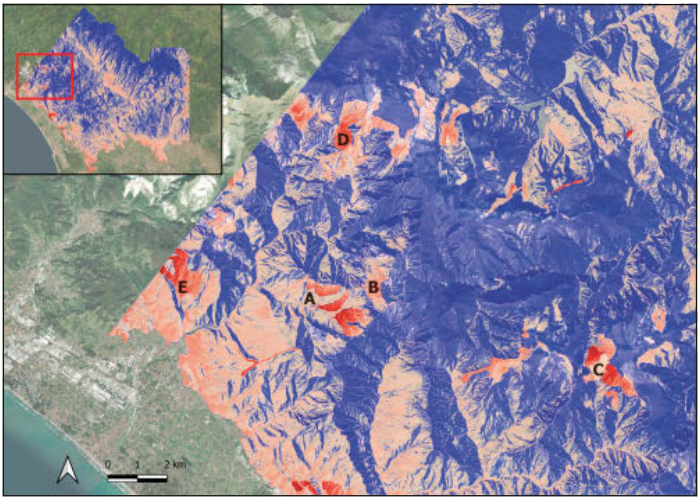

Fig. 6 – The north-western area of Model 1a predictive map with locations mentioned in the text: A) Antona; B) Mt. Altissimo; C) Foce di Mosceta; D) Vergheto; E) Mt. Brugiana.

a value that balanced the predictions between 'event' and 'non-event' (Model 1a: 0.08054071; Model 2: 0.1155059) (Tab. 3).

Finally, both models have similar AIC values (Model 1a AIC = 292.7; Model 2 AIC = 296.1) and show similar reliability and prediction capability. In fact, the BIC scores show almost identical values (Model 1a BIC score = 316.747; Model 2 BIC score = 316.092); the BIC weight is slightly higher for Model 2 (Model 1a BIC weight = 0.418; Model 2 BIC weight = 0.58); however, there are no substantial differences in reliability between models.

The application of computational models and predictive analyses in archaeology – especially when human behaviour factors are involved – should always be paired with a contextual interpretation of the multiple factors interrelated. In the case of two equally reliable models with the same predictive capacity, the choice of the best model is a crucial step that falls on the researchers who read and interpret the results (Gillings *et al.* 2020, 13). Although Model 1a has slightly lower performance, it was considered as more suitable for the construction of the predictive map. In fact, Model 1a produces a more readable and higher detailed map (Fig. 5), and the selected variables were derived from an additional Point Pattern process step that ensures the spatial dependence of the archaeological features with the environmental variables. Based on the considered variables, the predictive map confirms that the foothill area has suitable characteristics for permanent settlement, despite the study area excludes urban settlement and centuriated plain areas. The entire foothill strip tends to be above the threshold, as in the N-E of the plain of Lucca and between Ripafratta and Massaciuccoli. Values close to 1 are attested in Piana di Mommio and Santa Lucia at the Camaiore Valley entrance and the foothill to the N of Pietrasanta up to Strettoia at Lago di Porta.

Inland, the entire Serchio Valley area is at medium-low values but still above the threshold. Going up in altitude and away from main waterways, values above the threshold decrease and the non-event areas increase proportionally, especially along ridges and steepest and most exposed to the N areas. Nevertheless, event values are especially recorded in areas characterised by gentler slopes and exposed to the E-SE. Values close to 1 are located on the NE slope of Monte Altissimo at approximately 1000 m a.s.l. However, in this case, a prediction bias must be considered due to quarries that regularised slopes and mountain profiles over the centuries; therefore, predictions are probably overestimated. Plateaus with gentle slopes, such as the western side of Mt. Pania della Croce, near Foce di Mosceta, the area of Vergheto to the W of Mt. Tambura, the area of Antona, and the Mt. Brugiana at approximately 800 m a.s.l, show more reliable values toward 1 (Fig. 6).

5. Conclusion

Predictive archaeological models are tools for projecting known patterns to different, unexplored locations in the landscape (Warren, Asch 2000, 6), with the aim of generating a spatial pattern that has predictive implications for future observations and, especially in archaeology, for predicting the location of evidence not yet observed (Wheatley, Gillings 2002, 161). In this study, we aimed to explore the settlement pattern of the Versilia and Garfagnana mountains provided by previous investigations in relation to environmental variables with the purpose of integrating the archaeological framework, clarifying human-environment dynamics and past landscape use, and directing new research in the area, such as the survey campaigns currently conducted by the MAPPA Lab (Dept. of Civilisations and Forms of Knowledge, University of Pisa) within the ARAM (ARcheologie dell'Abbandono sulla Montagna di Mezzo) project. After investigating the spatial dependence of finds with environmental variables through Point Pattern Analysis, two predictive models were created. Given their equal robustness and reliability, the choice of the most fitting model to create the prediction map was then guided by the readability and interpretability of the output.

The suitable characteristics of the foothills for permanent settlement are thus confirmed; as clearly evidenced by Kernel Density Estimation, it is here, moreover, that the greatest archaeological concentration occurs, close to the plain where urban centres and the main road network were located. In any case, event values are also present in the innermost areas and at higher altitudes, especially on the gentler, E-SE-facing slopes. Usually, predictive models in archaeology rest on the assumptions that settlement choices were strongly influenced or conditioned by the characteristics of the natural environment and that these environmental factors are represented, at least indirectly, in contemporary maps (Warren, Asch 2000, 6). For these reasons, the major criticism of predictive models is environmental determinism and predictive modelling of archaeological patterns is regarded as one of the most controversial applications of computational archaeology (Wheatley, Gillings 2002, 161). Critical issues and opportunities in the archaeological modelling process have long been discussed and emphasised (Kvamme 2006) and the importance of the researcher's contextual interpretation and cross-validation of results are established.

The probably overestimated predictions at the quarry areas, due to the regularisation of slopes and mountain profiles over the centuries, are a clear example of the possible distortion due to the use of environmental factors recorded in contemporary cartography. Particularly in the mountainous area, larger and permanent settlements are often more visible and therefore often represent the only features that allow for the development of settlement location models. A predictive model that can generate a spatial pattern with predictive implications provides greater awareness of the possible extension of the settlement pattern, human-environment dynamics, and past landscape use to direct future research and field validation on lesser-known areas.

Author statement

Conceptualization and methodology: S. Basile and A. Campus; Data collection and curation: S. Basile (Garfagnana), A. Campus (Versilia); Data analysis: S. Basile and A. Campus; Writing and review of the original draft: S. Basile and A. Campus; Visualization: S. Basile and A. Campus.

Salvatore Basile, Antonio Campus
Laboratorio MAPPA
Dipartimento di Civiltà e Forme del Sapere
Università degli Studi di Pisa
salvatore.basile@cfs.unipi.it, antonio.campus@cfs.unipi.it

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, e0192039 (https://doi.org/10.1371/journal.pone.0192039).

Anichini F., Fabiani F., Gattiglia G., Gualandi M.L. (eds.) 2012, *Mappa. Metodologie Applicate alla Predittività del Potenziale Archeologico*, Roma, Edizioni Nuova Cultura (https://doi.org/10.4458/0917).

Anichini F., Gattiglia G. 2012, *Urban Archaeological Information System. Riflessioni e criticità*, in Anichini *et al.* 2012, 31-41.

Arnaez J., Lasanta T., Errea M.P., Ortigosa L. 2011, *Land abandonment, landscape evolution, and soil erosion in a Spanish Mediterranean mountain region: The case of Camero Viejo*, «Land Degradation & Development», 22, 6, 537-550 (https://doi.org/10.1002/ldr.1032).

Baddeley A., Turner R., Rubak E. 2021, *Spatstat: Spatial Point Pattern Analysis, Model-Fitting, Simulation, Tests*, «R Foundation for Statistical Computing», Vienna (https://cran.r-project.org/web/packages/spatstat/spatstat.pdf).

Basile S. 2022a, *Lucca e l'*ager Lucensis *dall'età tardo repubblicana al tardoantico: le trasformazioni di una città e del suo territorio*, unpublished Doctoral Thesis, Università di Pisa (https://etd.adm.unipi.it/t/etd-06272022-165434/).

Basile S. 2022b, *The funerary landscape of late antique and early medieval Lucca through Point Pattern Analysis*, «European Journal of Post-Classical Archaeologies», 12, 87-106.

Basile S., Carrer F. 2022, *A computational modelling approach to reconstruct the fluvial system of the floodplain of Lucca in the Roman period*, «ArcheoLogica Data», 2, 63-80 (https://doi.org/10.13131/unipi/2785-0668/k5sw-3e98).

Bevan A. 2020, *Spatial point patterns and processes*, in M. Gillings, P. Hacıgüzeller, G. Lock (eds.), *Archaeological Spatial Analysis: A Methodological Guide*, Oxon, Routledge, 60-76.

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 (https://doi.org/10.1002/jqs.3224).

Brandolini F., Carrer F. 2020, Terra, Silva et Paludes. *Assessing the role of alluvial geomorphology for Late-Holocene settlement strategies (Po Plain – N Italy) through Point Pattern Analysis*, «Environmental Archaeology», 7, 5, 511-525 (https://doi.org/10.1080/14614103.2020.1740866).

Boccaletti M., Coli M., Decandia F.A., Giannini E., Lazzarotto A. 1980, *Evoluzione dell'Appennino settentrionale secondo un nuovo modello strutturale*, «Memorie della Società geologica italiana», 21, 359-373.

Carlson D.L. 2017, *Quantitative Methods in Archaeology Using R*, Cambridge, Cambridge University Press.

Campus A. 2022, *L'*ager Pisanus *tra romanizzazione e tardoantico: continuità, discontinuità e trasformazioni*, unpublished Doctoral Thesis, Università di Pisa (https://etd.adm.unipi.it/theses/available/etd-09262022-155452/).

Carmignani L., Giglia G., Kligfield R. 1978, *Structural evolution of the Apuane Alps: An example of continental margin deformation in the Northern Apennines, Italy*, «The Journal of Geology», 86, 4, 487-504.

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.

Castaldini D., Genevois R., Panizza M., Puccinelli A., Berti M., Simoni A. 1998, *An integrated approach for analysing earthquake-induced surface effects: A case study from the Northern Apennines, Italy*, «Journal of Geodynamics», 26, 2-4, 413-431.

Ciampoltrini G. 2003, *La valle del Serchio fra I e VI secolo d.C. Aspetti della dinamica dell'insediamento*, in G. Roncaglia, A. Donati, P. Giuliano (eds.), *Appennino tra antichità e Medioevo*, Città di Castello, Petruzzi Editore, 209-223.

Ciampoltrini G. 2006, *La colonia e la montagna. Archeologia d'età augustea a Lucca e nella Valle del Serchio*, San Miniato, Tipografia Toscana.

Conolly J., Lake M. 2006, *Geographical Information Systems in Archaeology*, Cambridge, Cambridge University Press.

Costanzo S., Brandolini F., Idriss Ahmed H., Zerboni A., Manzo A. 2021, *Creating the funerary landscape of Eastern Sudan*, «PLOS ONE», 16, 7, e0253511 (https://doi.org/10.1371/journal.pone.0253511).

Crema E.R. 2020, *Non-stationarity and local spatial analysis*, in M. Gillings, P. Hacigüzeller, G. Lock (eds.), *Archaeological Spatial Analysis. A Methodological Guide*, Oxon, Routledge, 155-168.

Della Casa P. 2010, *Recherche interdisciplinaire en archéologie et écologie humaine dans la vallée centre-alpine de la Léventine, Tessin (Suisse)*, in Tzortzis, Delestre, Greck 2010a, 23-32.

Devoti S., Nisi M.F., Silenzi S. 2003, *Caratteri geologici ed evoluzione paleogeografica della pianura versiliese*, «Studi Costieri», 6, 73-89.

Eve S.J., Crema E.R. 2014, *A house with a view? Multi-model inference, visibility fields, and Point Process Analysis of a Bronze Age Settlement on Leskernick Hill (Cornwall, UK)*, «Journal of Archaeological Science», 43, 267-277 (https://doi.org/10.1016/j.jas.2013.12.019).

Giannini M. 2005, *I Romani in Versilia: dinamiche di popolamento e organizzazione del territorio*, «Orizzonti», 16, 81-91.

Gillings M., Hacigüzeller P., Lock G. 2020, *Archaeology and spatial analysis*, in M. Gillings, P. Hacigüzeller, G. Lock (eds.), *Archaeological Spatial Analysis. A Methodological Guide*, Oxon, Routledge, 1-16.

Fabiani F. 2006, "*… stratam antiquam que est per paludes et boscos…*". *Viabilità romana tra Pisa e Luni*, Pisa, Edizioni plus.

Kempf M., Weaverdyck E.J.S. 2023, *Modelling bias and environmental preferences in archaeological spatial analysis*, «Vegueta. Anuario de la Facultad de Geografia e Historia», 23, 1 (https://doi.org/10.51349/veg.2023.1.03).

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).

Kvamme K.L. 2006, *There and back again: Revisiting archaeological locational modeling*, in M.W. Mehrer, K.L. Wescott (eds.), *GIS and Archaeological Site Location Modeling*, New York, CRC Press, 23-55.

Li L., Li Y., Chen X., Sun D. 2022, *A prediction study on archaeological sites based on geographical variables and logistic regression. A case study of the Neolithic era and the Bronze age of Xiangyang*, «Sustainability», 14, 15675 (https://doi.org/10.3390/su142315675).

Menchelli S. 1991, *Materiali per la storia della Versilia in età romana*, «Studi Classici e Orientali», 40, 387-429.

Migliavacca M., Bandera S., Bezzi J., Casarotto A., Pisoni L. 2021, *Archeologia delle alte quote sulla montagna veneta: la campagna di ricognizione di superficie 2019 a Recoaro Terme (Vicenza)*, «Fasti On Line Documents & Research», 14, 1-30.

Mocci F., Segard M., Walsh K. 2010, *Données récentes sur l'occupation humaine dans les Alpes méridionales durant l'Antiquité*, in Tzortzis, Delestre, Greck 2010a, 309-324.

Modica G., Praticò S., Di Fazio S. 2017, *Abandonment of traditional terraced landscape: A change detection approach (a case study in Costa Viola, Calabria, Italy)*, «Land Degradation & Development», 28, 8, 2608-2622 (https://doi.org/10.1002/ldr.2824).

Nakoinz O., Knitter D. 2018, *Modelling Human Behaviour in Landscapes: Basic Concepts and Modelling Elements*, Switzerland, Springer.

Tzortzis S., Delestre X., Greck J. (eds.) 2010a, *Archéologie de la montagne européenne, Actes de la table ronde internationale de Gap (2008)*, Paris, Édition Errance.

Tzortzis S., Delestre X., Greck J. 2010b, *Avant-propos*, in Tzortzis, Delestre, Greck 2010a, 9-10.

Visentin D., Carrer F. 2017, *Evaluating Mesolithic settlement patterns in mountain environments (Dolomites, Eastern Italian Alps): The role of research biases and locational strategies*, «Archeologia e Calcolatori», 28, 1, 129-145 (https://doi.org/10.19282/AC.28.1.2017.08).

Wachtel I., Shelach-Lavi G., Garti S. 2018, *Predictive modelling for archaeological site locations: Comparing logistic regression and maximal entropy in North Israel and North-East China*, «Journal of Archaeological Science», 98, 28-36 (https://doi.org/10.1016/j.jas.2018.02.001).

Warren R.E., Asch D.L. 2000, *A predictive model of archaeological site location in the Eastern Prairie Peninsula*, in K.L. Westcott, R.J. Brandon (eds.), *Practical Applications of GIS for Archaeologists: A Predictive Modelling Kit*, London, Taylor&Francis, 5-32.

Wheatley D., Gillings M. 2002, *Spatial Technology and Archaeology. The Archaeological Applications of GIS*, London, Taylor & Francis.

ABSTRACT

Mountain archaeology has a long research tradition and in recent years the number of studies on this topic has increased considerably, shedding new light on the dynamics of mountain's communities. Versilia and Garfagnana districts (Lucca, north-western Tuscany) largely fall between the Apuan Alps and the Apennine ridge. Although these territories have never been systematically investigated, the collection of all available archaeological legacy data indicates a settlement pattern of undoubted interest for the Roman times. This paper aims at exploring the settlement pattern of these mountain territories, integrating Point Pattern Analysis and Logistic Regression to achieve a predictive map of archaeological presences and to analyse their interrelations with the environment. Analyses prove the spatial dependence of finds with geomorphological and pedological variables, but also with the distance to major watercourses and solar irradiation. Based on the considered variables, the predictive map confirms that the foothill and gentler slopes facing E-SE areas have suitable characteristics for permanent settlement. Moving towards the more inland and higher altitude territories, the non-event areas increase proportionally, especially along the ridges, and the steeper, north-facing areas. Thus, the results make it possible to integrate the archaeological framework, clarifying human-environment dynamics, and directing new studies.