

Machine Learning for the Built Heritage Archaeological Study

AMAIA MESANZA-MORAZA, ISMAEL GARCÍA-GÓMEZ, and AGUSTÍN AZKARATE,
Built Heritage Research Group, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain

The presence of artificial intelligence in our lives is increasing and being applied to fields such as medicine, engineering, telecommunications, remote sensing and 3D visualization. Nevertheless, it has never been used for the stratigraphic study of historical buildings. Thus far, archaeologists and architects, the experts in archaeology of architecture, have led this research. The method consisted of visually—and, consequently, subjectively—identifying certain evidence regarding the elevations of such buildings that could be a consequence of the passage of time. In this article, we would like to present the results from one of the research projects pursued by our group, in which we automated the stratigraphic study of some historic buildings using multivariate statistic techniques. To this end, we first measured the building using surveying techniques to create a 3D model, and then, we broke down every stone into qualitative and quantitative variables. To identify the stratigraphic features on the walls, we applied machine learning by conducting different predictive and descriptive analyses. The predictive analyses were used to rule out any blocks of stone with different characteristics, such as rough stones, joint ashlar, and voussoirs of arches; these are irregularities that probably show building processes and whose identification is crucial in ascertaining the structural evolution of the building. In supervised learning, we experimented with decision trees and random forest—and although the results were good in all cases, we ultimately opted to implement the predictive model obtained using the last one. While identifying the evidence on the walls, it was also very important to identify different continuity solutions or interfaces present on them, because although these are elements without materiality, they are of great value in terms of timescale, because they delimit different strata and allow us to deduce the relationship between them.

CCS Concepts: • **Applied computing** → *Arts and humanities; Architecture (buildings)*;

Additional Key Words and Phrases: Archaeology of Architecture, building archaeology, built heritage, machine learning, multivariate analysis, stratigraphic analysis, data mining

ACM Reference format:

Amaia Mesanza-Moraza, Ismael García-Gómez, and Agustín Azkarate. 2020. Machine Learning for the Built Heritage Archaeological Study. *J. Comput. Cult. Herit.* 14, 1, Article 10 (December 2020), 21 pages.
<https://doi.org/10.1145/3422993>

This contribution is part of the research project “The Archeology of Architecture: new methodological challenges and their application in Europe and Latin America” (HAR2015-64439-P), financed by the Spanish Ministry of Economy and Competitiveness and the project “Archaeology of Architecture in the old and the new world: from the stratigraphy of the buildings to the stratigraphy of the urban fabric” (PID2019-109464GB-I00), financed by the Spanish Ministry of Science and Innovation.

Author’s addresses: A. Mesanza-Moraza (corresponding author), I. García-Gómez, and A. Azkarate, Built Heritage Research Group, University of the Basque Country UPV/EHU, Micaela Portilla Research Center, Justo Velez de Elorriaga Street, 1st. 01006 Vitoria-Gasteiz, Spain; emails: {amaia.mesanza, ismael.garcia, agustin.azcarate}@ehu.eus.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs International 4.0 License.

© 2020 Copyright held by the owner/author(s).

1556-4673/2020/12-ART10

<https://doi.org/10.1145/3422993>

1 INTRODUCTION

In archaeology, the main dating method is stratigraphic, while the rest is structured around the following: documentary, typological, formal, structural, or archaeometric sources [1–3]. This stratigraphic analysis, which is diachronic in nature, requires the identification of all stratigraphic units above ground archaeology—deposits and surfaces [4]—as well as existing links between them to be able to interpret the historical evolution of a building. Together with stratigraphic analysis, there is a second method that is typological in nature. In this case, it is a synchronic analysis whereby the formal characteristics of a series of elements or material remains can be compared, making the formation of clusters possible, i.e., groups in which its individuals share very similar characteristics, but that are very different from those of the rest of the components of other groups.

Archaeological typologies in architecture are created with the objective of dating buildings. Thus, while there is a series of chronological instruments borrowed from other disciplines (Carbon 14, thermoluminescence, electron spin resonance, etc.), there are other types of indicators, developed in the field of archaeology of architecture, which are applied on a local basis, i.e., that are useful in more or less limited geographic contexts. This takes the form of the chronology of structural elements and mensiochronology [5]. This second technique—mensiochronology—attempts to connect the dimensions of bricks to different historical periods, and to this end, analyses them statistically [6, 7]. It fundamentally studies the bricks rather than any other type of construction material due to its being a question of mass-produced products, and in which dimensional variability is an important feature [8].

The procedure involves collecting a sufficiently large sample of bricks, measuring them, analysing them statistically, and comparing them with a series of known cases that have been previously dated with certainty based on other absolute chronological indicators, such as written sources, inscriptions, ceramics, and so on. In all cases, this is fundamental to ensure a good definition of the area of application, because it is a method that works very well in local areas [9].

Although these studies have mostly been carried out with bricks only, other types of tests have been carried out along the same lines, by changing the construction material, for example, in adobe [10, 11], limestone [12], or lithic materials [13–21]. However, the result obtained has not proved particularly satisfactory. Occasionally, the size of the stones with which a building is built may be limited by the architect’s inspiration, but in most cases, demand will be marked by the maximum size of the block of rock available, which is the main impediment according to specialists in the subject in terms of the development of mensiochronology projects in another material other than brick [7, 22].

Nevertheless, even though the few projects carried out so far with stone material have not generated optimism in terms of their value, the project carried out by Tiziano Mannoni in 1974 [13] will serve as a good starting point. In it, following on from a statistical study on the dimensions of the ashlar blocks from the “Castelo” de Genoa, Mannoni confirms the existence of different groups of pieces that, in the words of the researcher, “could constitute chronological evidence.”

Following in the Italian researcher’s footsteps, we applied the statistic to the historical-archaeological study of historical buildings, and the results obtained thus far [23] have confirmed what Mannoni said, enabling a possible chronological sequence for those buildings to be formulated. This reinforced the importance of the typological strategy for the archaeological study of stone elevations. But for such reconstruction, it was necessary to collect more evidence than the groups that had been formed. There are pieces of evidence that are more typical of the stratigraphic strategy of the analysis of walls that may include, among other things, interfaces or joint ashlar. And stratigraphy and typology do not have to be separated, but rather, can and should complement each other [1], just as they did in the course of the restoration project of the Santa María Cathedral of Vitoria-Gasteiz (Spain) in 1996 [24, 25], in the early medieval churches of Alava (Spain) in 2007 [26], and in the Salt Valley of Salinas de Añana (Alava, Spain) one year later [27, 28]. To the aforementioned multi-layered nature of historical buildings should now be added their pluritipologized nature [29].

However, our proposal is to tackle the archaeological study of elevations by applying only the typological methodology, albeit not in the way it has been done up to now, but rather, in an automated way. Of the two methodologies described above, this is the only one that can be automated. To this end, a multivariate analysis will be used that focuses on applying different techniques, such as Principal Components Analysis (PCA), regression analyses, discriminant analysis, cluster analysis, among others [30]. This methodology breaks with the traditional way of handling this type of project, and it is necessary to implement quantitative and multivariate methods accordingly [31], moving from the habitual way of proceeding—qualitative, subjective, and intuitive—to a more quantitative and objective one [32]. Together with this more quantitative archaeology and the incorporation of statistics into archaeological work, machine learning and expert systems—two branches of artificial intelligence—will also be included.

There are hardly any projects where multivariate analysis has been applied to the historical-archaeological study of buildings—perhaps the project carried out in 2013 by the Argentine architect G. Rolón [31, 33], although as it is a completely different area of application, it is difficult to find what to compare the obtained result to. While it is true that there are increasingly more projects that are researching into how to automatically identify or classify particular architectural elements or styles [34–38], it is not easy to find one that focuses on the archaeological study of historical buildings.

This is precisely the main innovation of this work that we present here, as it is the first time a historical building has been studied entirely and independently by a computer following a period of supervised learning, based on the evidence detected in the walls. The system has learned to identify particular signs that involve structural changes in a building, which enables a hypothesis to be formulated about its historical evolution throughout history.

Predictive and descriptive techniques were applied as part of the learning process that the computer is put through. With the former, the goal was to try and identify and characterise patterns and/or behaviour from the data that would enable key evidence to be pinpointed about the elevations to interpret the structural evolution of the building. At the same time, descriptive techniques would provide the solutions necessary to produce clusters, i.e., homogenous groups of ashlar that, together with the previously categorised evidence, would enable a hypothesis to be formulated about the structural sequence of the building being studied.

Throughout this article, we will explain how to apply predictive techniques to identify evidence about the elevations of historical buildings, using different constructions that have been appropriately selected in each of the phases. The second part of the research, the cluster or descriptive analysis, is included in the article published by the authors of Reference [23].

2 MATERIALS AND METHODS

2.1 Background

2.1.1 Selected Buildings. The proposed research work required having buildings from which to obtain valid information for autonomous learning about the system and other different buildings where this knowledge could be applied. Both sets had to include a series of common features, such as the lithology of stones or the type of bonding, although it was not necessary for them to be of similar architectural styles. This issue was important for the reproducibility of the experiment.

The lithology of the walls will be the common link between all the buildings selected to take part in the project, as in all of them limestone or Palaeocene limestone (Lumaquelas de Ajarte), worked mostly in the shape of ashlar, appear as the preferred rock for their construction—an aspect that is quite common in other churches in Alava (Spain) [39].

We will therefore carry out our research on three historic buildings: (1) the Hermitage of the Purísima Concepción, built at the end of the 12th century and located in the town of San Vicentejo 14 km south of Vitoria-Gasteiz (Spain), in Burgos province. For many researchers it is an “exceptional” temple within the late

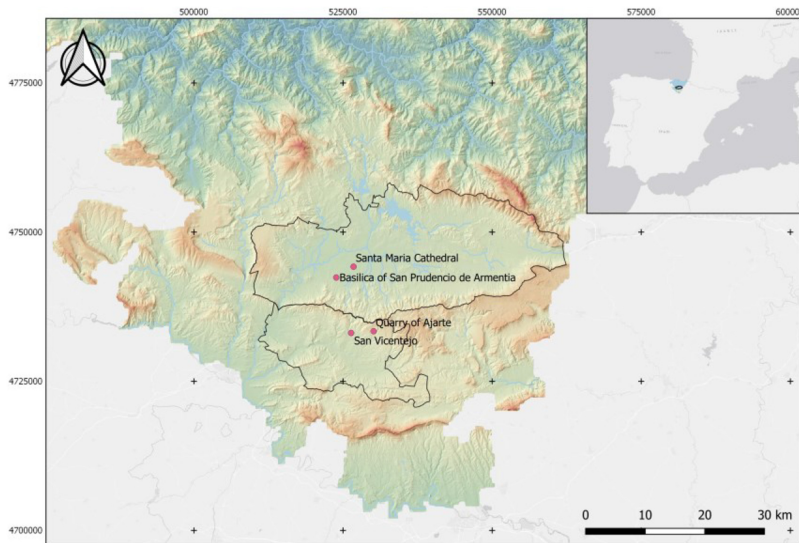


Fig. 1. Location of the buildings selected and the Ajarte quarry.

peninsular Romanesque style due mainly to its extraordinary capitals [40]. The church consists of a single nave with two sections and circular chancel, all built with very well-cut ashlar made from Palaeocene limestone from the nearby Ajarte quarry. (2) The Basilica of San Prudencio de Armentia, one of the best examples of the Romanesque style in the Basque Country. Its geographical location, just 3 km southwest of Vitoria-Gasteiz, at a crossroads of ancient roads and its link to the patron saint of the province, as well as its significance as the most important spiritual centre of Alava in the High Middle Ages, make it especially interesting from a historical point of view [41]. And (3) Santa María Cathedral, located in the most elevated part of the old city of Vitoria-Gasteiz, within its original walled enclosure. It is a Latin cross temple, with three naves, a chancel, an ambulatory, and side chapels, built using techniques common to different times during the Gothic period [24] (Figure 1).

2.1.2 The Breaking Down of Buildings into Variables Based on Their Geometric Documentation. In the course of reading elevations, it is essential to have geometric documentation where the evidence discovered and the results of the research can be captured [6, 42, 43]. The record that is made must be a complete representation of the building and its parts, this being all the more useful and facilitating archaeological work insofar as it will be of greater quality and precision [44]. L. Caballero [45] said that “the quality of the documentation does not necessarily have to affect the degree of validity of the final result of the archaeological analysis,” but the more detailed the documentation, the more perfect the result.

It should be noted that, in the study of the historical-structural evolution of a building, this is considered a single element, so the graphic model obtained must comply with this determining factor. In other words, regardless of whether the final product is a collection of drawings, the record should be put together in such a way that the different parts of the building may be linked [46].

When the 3D model is ready, carried out using any of the most widely used techniques in built heritage (laser scanner, automatic photogrammetry, stereophotogrammetry, rectified photography or topography), it is usual to develop 2D drawings that may serve as a tool where the results of archaeological and/or architectural research can be captured more visually. However, for our purpose of trying to produce automatic learning, of trying to programme a computer so it may be able to simulate the way in which archaeologists of architecture find solutions to problems, it is necessary to go one step further by making all this geometric information an active

part of the research project itself [47]. To do this, it will be necessary to find a way of extracting all this data and transforming it, dividing it into simpler forms that help us to understand it computationally.

After much reflection as to whether to adopt one element or another as a minimum unit of study, as this choice conditions the rest of the process and the result obtained from it, it was decided to determine that the minimum unit of study would be each block of stone that makes up a building. This is because the latter would be able to provide sufficient information when it came to “automating” the process of the stratigraphic reading of elevations, being relatively simple to break down into variables.

These variables, in which each block of stone is going to be broken down, must be able to define them unambiguously and unequivocally. They are a fundamental part of the research, as they are present in all phases of the process—observation and measurement, classification, analysis, prediction [48]. They need to be set out clearly, by being defined conceptually at first, and then moving on to the operational phase where the procedures are established that allow measurements to be obtained.

In the same way as when it comes to bringing together the geometric documentation of a building, it is necessary to consider it in its entirety, because it is also interesting to study the variability in all its stones or at least the great majority of them. If this is done partially, it will be difficult to identify structural moments or chronologies on the elevations.

Initially, for a typological study, neither the type nor the number of variables to be used is established, as these depend on the circumstances and needs attached to each context of application [49]. However, it is advisable to take all options into consideration in this first phase of the work as until the whole process is finished, it is unknown which one or ones will be the greatest determining factors [26]. It will be the same numerical methods that indicate those variables that carry most weight [32]. Generally speaking, in the study of historical buildings, the variables are usually of a technical-structural type (dimensions, types of materials, types of bonding, types of instruments used in sculptures, stonemason marks, mortar, etc.), formal type (types of capitals, arches, triforiums, windows, brackets, etc.) or spatial type (location, level, gradient, etc.) [50].

When analysing the aforementioned variables, practically all of them turned out to be qualitative except for those that evidenced a geometric or spatial component. Therefore, it was necessary to add to those qualitative variables others of a quantitative nature that could be obtained automatically. Finally, in the case of the stone blocks, 33 attributes were extracted in total, of which only 9 were qualitative, and the rest, quantitative. Thus, there will be a first group where 6 contextualisation variables are located (building, type of building, location, orientation, exposition, lithology), the second of which would study 3 dichotomous variables (arch, joint ashlar, and type of stone), and it will be these that will be used to classify the stones within the supervised learning phase. For the third block, there will be the quantitative variables that describe both the geometric characteristics of the stone blocks (height, length, area, perimeter, centroid (X and Y coordinates), diagonal 1, diagonal 2, height to length ratio), those of their coverings¹ (area, perimeter, and centroid coordinates) and existing relations between the stone blocks and their coverings (area difference (% and absolute), perimeter difference (% and absolute), centroid difference (% and absolute and coordinates) and diagonal difference (% and absolute)).

We automated the process of generating the coverings and creating and calculating the direct and indirect variables as much as possible to avoid subjectivity. We used the “Model Builder” ArcGis application to avoid data collection errors. This was a fundamental part of the research process. In addition, the models created were exportable directly to Python in case modification was necessary. Each stone had a unique identification code.

2.2 Supervised Classification

With this proposal approach—documenting the building in its entirety and dividing it into minor elements defined by multiple variables—the information exceeds the capacity for manual management and analysis, as has

¹In this context, one synonym of “covering” could be “bounding box” or perhaps better “minimum bounding rectangle.” This is an expression of the maximum extents of a two-dimensional object (that is, the stones of the buildings) within its coordinate systems (min(x), max(x), min(y), and max(y)).

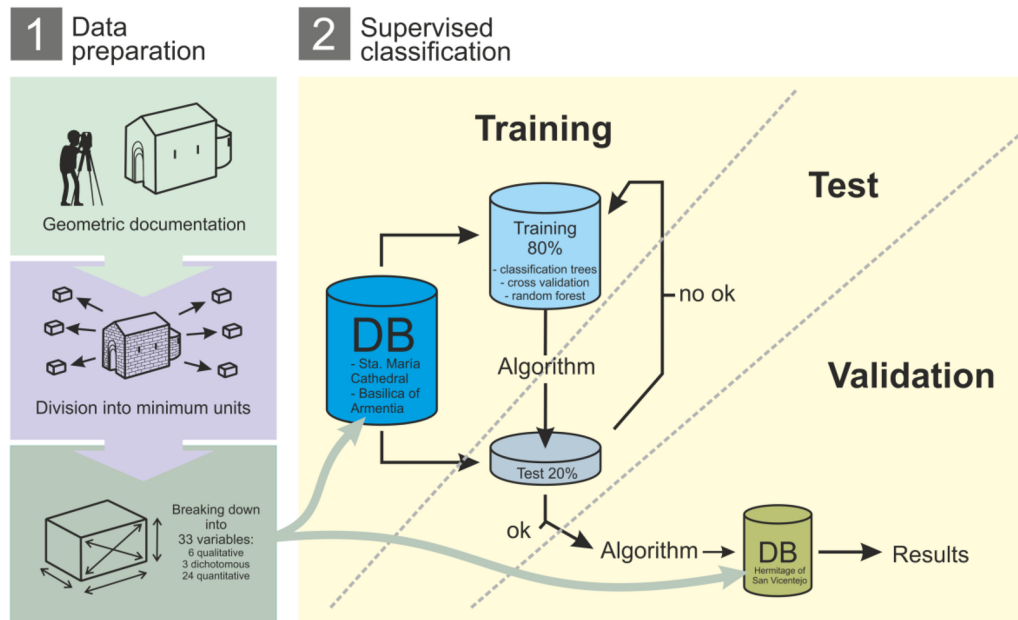


Fig. 2. Flowchart of the project.

been the case until now. Faced with such a volume of data, the growth in recent years of big data, data mining, machine learning, and artificial intelligence and its presence in almost all scientific fields makes the possibility of applying some of these concepts to the archaeology of architecture a very interesting option (Figure 2).

One of the purposes of using predictive techniques is to find patterns or behaviour within a broad set of data. In our research, these methodologies will be used with two very specific objectives in mind:

- the classification of stone blocks into two broad categories: ashlar and rough stones.
- the analysis of the continuity of coursed ashlar, which involves identifying stones with certain irregularities, such as is the case with joint ashlar.

In fact, what is proposed is the positioning of outliers within the configuration of the elevations, i.e., stones that, for various reasons, do not have the regular shape of an ashlar and that may be an indication of possible structural action. Normally, the construction of great historical buildings tends to be prolonged several years or decades, and this makes the wall heterogeneous. Or the simple passage of time makes it necessary to redo or modify the initial design. All these actions are reflected in the elevations in the type of evidence and it is precisely this that the system should autonomously detect [51].

2.2.1 Classification of Ashlar and Rough Stones. When building in stone, there are three basic types that differ depending on the degree of the cut and size: rough stones, smaller ashlar, and ashlar blocks. In other words, ashlar are those blocks of stone in the shape of a parallelepiped that are well cut and whose corners are perfectly squared, in contrast to rough stones that are completely irregular in shape. The smaller ashlar come somewhere in between.

The research will be based on individualisation of the ashlar in contrast to the rough stones, because, as can be seen, they have very uneven features that are relatively easy to detect deductively. The smaller ashlar compared to the ashlar does not have a clear nuance that distinguishes it. In the literature consulted, there is hardly any



Fig. 3. North elevation of the Basilica of San Prudencio de Armentia. The left part is made using ashlars, while the right part corresponds to rough ashlars.

work in this area, except that carried out by Osés [52] in which an initial attempt was made to classify the stone blocks of walls based on a series of outlines extracted from a series of images. In the aforementioned project, the blocks were classified into three groups, roughly equivalent to the three defined by us: masonry, small ashlar, and ashlar. Very good classifications were obtained for both extremes but not such good ones for the intermediate group, which does not have the characteristics of either one or the other. This classification of the stone blocks, apparently simple to undertake visually, is not so easy to automate. However, seeing that the researchers can indeed catalogue them manually in an agile and accurate way, we opted for supervised learning as being the most appropriate methodology for the purpose of trying to build predictive models that could classify them automatically.

To be able to look for patterns or behaviour, it is necessary to have a sample in which the data has been previously classified manually and is guaranteed—it will be this data that “teaches” the computer. This first part of the process is fundamental, as the results obtained will be closely conditioned by the researcher’s prior knowledge of this.

In this research, for the learning and/or test phases, the stones from the walls of the Basilica of San Prudencio de Armentia will be used, whereby both elements—ashlars and rough stones—are distinguished from each other without any doubt (Figure 3). In this way, the database will contain nearly 9,000 stones, all of them of the same lithology, limestone and from the same quarry in Ajarte, and spread out across the elevations of the north and south exterior walls of the aforementioned building.

The underlying model in the discriminant analysis requires prior verification of the supposition of normality, homoscedasticity, and linearity. These requirements are not necessary when logistic regression or decision trees are applied, but, even so, statistical analysis of the data can provide useful information that may enable its structure to be understood, as well as the results obtained.

One of the first suppositions to be proven will be multivariate normality. In the absence of a statistical test that would allow such proof, this verification will be carried out individually in the case of all variables.² In this case, after the Kolmogorov-Smirnov and Anderson-Darling tests were carried out for the groups of rough stones and ashlar, none of the variables used proved to be in accordance with a normal distribution. The lack of normality is due largely to the existence of some exceptionally high values, fundamentally in the “high” and “long” variables that make up the primary variables, which will have to be analysed in case of any anomalous values that would need to be removed from the sample. Even so, as the sample is large enough, we could turn to the theorem of the central limit to be able to assume this normality [53].

Analysing these two variables in detail, in the case of the category defined as rough stones, both attributes—high ($\mu = 0.085\text{m}$ $\sigma = \pm 0.050\text{m}$ $N = 5,123$) and long ($\mu = 0.190\text{m}$ $\sigma = \pm 0.115\text{m}$ $N = 5,123$)—are biased to the right, i.e., they have positive symmetry that comes from an accumulation of values in the lower area of the histogram, which is logical given the characteristics of this type of stone. The same analysis for the group of ashlar reflects a somewhat different situation, as none of the high ($\mu = 0.267\text{m}$ $\sigma = \pm 0.080\text{m}$ $N = 3,557$) or long ($\mu = 0.421\text{m}$ $\sigma = \pm 0.177\text{m}$ $N = 3,557$) variables appears to be biased.

In terms of homoscedasticity, the Levene test was applied, the result of which allows us once again to reject the null hypothesis of equality of variances for both groups—a condition already observed in the boxplots and the corresponding histograms.

As for multicollinearity, this occurs when one of the independent variables can be explained as a linear combination of one or more variables producing a redundancy of information. If this phenomenon occurs, it is normal that the predictive models subsequently obtained are not very explanatory, giving rise to major errors. It is recommended that those variables whose Pearson regression index is equal to or greater than 0.80 [54] be removed from the sample. However, special care must be taken not to remove such many variables from the sample, as there is a danger of later creating models that are not very predictive. After analysing the covariance and correlation matrix and following the above indications, the area and perimeter variables of all coverings were removed from the database.

The result obtained—neither normality nor homoscedasticity is achieved—obliges us to cast aside the discriminant analysis as a technique for classification of ashlar and rough stones and to opt, therefore, for another type of tool.

Not all variables will have the same importance or influence on the predictive models that are obtained. Therefore, before starting the first of the analyses, a separability study was carried out (Figure 4) that enabled us initially to ascertain which would be the most significant—not only to categorise the type of stone, but also to classify the joints. To do this, we used the normalized difference formula based on mean values and standard deviations [56].

As data mining experts point out [57–60], the percentage of records used in the training phase needs to be considerably higher than that used for the testing phase, its being recommended at 60%–40%. After calculating the models with different percentages, and seeing the results obtained, the predictive models were generated based on 80% of the data while the remaining 20% was left for the testing phase. The elements involved at each moment were randomly selected by the software used.³

In this training phase, we experimented with classification trees, cross-validation, and random forest (Table 1). However, to try to optimise the models, we finally opted for the random forest algorithm where, contrary to what happens for example with classification trees, as more trees are introduced into the calculation, the algorithm does not over-adjust, but, rather, sets a generalisation error limit [61]. Different models were created using

²The assumption of normality can be verified in various ways: graphically (e.g., histograms, box-plots, Q-Q charts), using statistical tests (e.g., Kolmogorov-Smirnov, Shapiro-Wilks, Anderson-Darling), or by numerical methods (e.g., central tendency measurements, coefficients of symmetry, kurtosis). For our research, we used the Kolmogorov-Smirnov and Anderson-Darling tests [54].

³To generate the predictive models, both in the training-test phase and in the subsequent validation phase, we used the free software KNIME Analytics Platform, version 2.11.2.

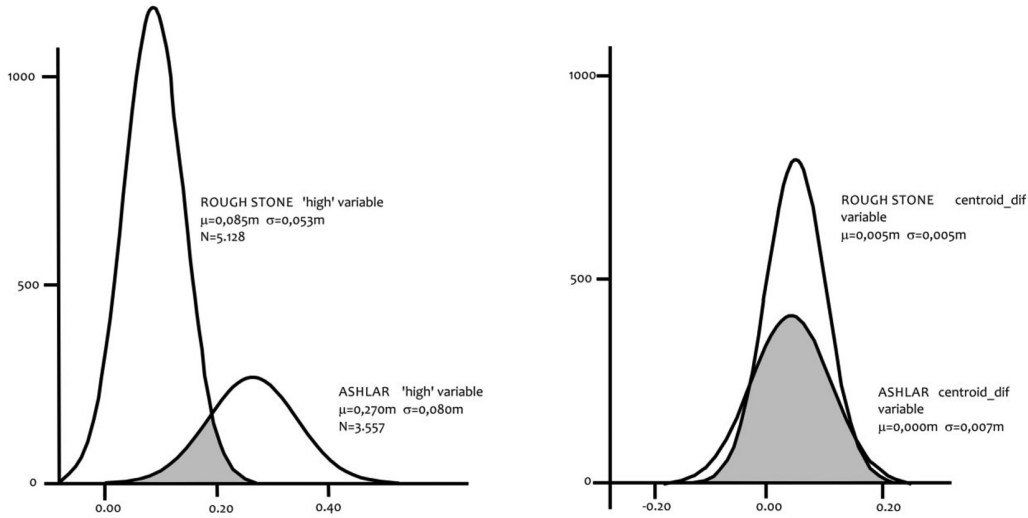


Fig. 4. Histograms with the mean and standard deviation of “high” and “centroid difference” variables. On the left, the histograms are quite separate, and this is because the “high” variable has a high separability in distinguishing between ashlar and rough stones. However, in the second graph, the histograms practically coincide, and so this means that the “centroid difference” variable is not a good discriminating variable for differentiating between ashlar and rough stones.

Table 1. Results of the Training Phase for Categorisation of Ashlars and Masonry Using Classification trees, Cross-validation, and Random Forest

	Overall reliability (%)	Cohen’s kappa	f1-measure
Classification trees	97.335	0.947	0.960
Cross-validation	97.552	0.951	0.965
Random forest	99.662	0.993	0.999

10-25-100 trees and the results obtained were very similar in all cases. All the tests carried out with this algorithm produced less than 1% of errors, achieving an almost perfect classification, and reaching the out-of-bag (OOB) statistic that indicates the prediction error values in the test set [62] were less than 3%.

However, there is a problem with overall reliability or accuracy, and this can hide important differences between categories with varying levels of accuracy, which is why it is important to analyse not only the diagonal values of the confusion matrix, but also the marginal values. We analysed the error of omission-EO (related to the precision or reliability of the producer-RP) and the error of commission-EC (related in this case to the recall or reliability of the user-RU). For our predictive models, these precision and recall values were 0.99, which indicated a very good classification; values close to one imply that the discriminant variables used allow for good differentiation between groups, while those close to zero—the extreme opposite—reflect the fact that the classification model is not valid. The Cohen’s kappa coefficient attempts to delimit the degree of adjustment due only to the accuracy of the classification, dispensing with that caused by random factors. In this case those values exceed 0.99, implying almost perfect agreement [64].

2.2.2 Joint Ashlars and Vertical Interfaces. Once the ashlar work and masonry had been separated, the database was much more homogeneous, and all the elements were ashlar. However, there were several blocks that had a slight rabbet, usually at a right angle in one of the corners, which ended up breaking that uniformity. These

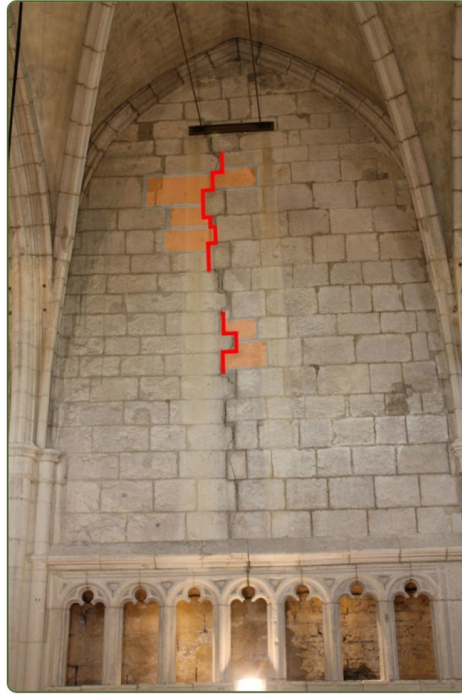


Fig. 5. Joint ashlar (orange blocks) and a zipper (term used colloquially to define the interface defined by the union of the joint ashlar) located in the northern transept of Santa Maria Cathedral (Vitoria-Gasteiz).

are the ashlar that we call joint ashlar (Figure 5). Identifying these blocks is important when undertaking the stratigraphic study of an elevation, as their appearance usually marks the presence of some type of intervention. In other words, ashlar tend to be made in the shape of a parallelepiped with four squared corners. When one whose angle has been lowered appears, that implies that it has been made this way to be able to adapt it to another block that was not initially in that position [63]. The presence of these special ashlar also enables a relative chronological component to be introduced in the area where they are located, as they will always be from a time prior to the ashlar that fits into them.

In the same way that the analysis of the ashlar work was carried out, here, we will also use, for training and testing, the database with the records of the elevations from the Basilica of San Prudencio de Armentia, where all those elements categorised as rough stones were removed, significantly reducing its size.

Another aspect that changed substantially was the percentage of stones included in each of the categories to be predicted. It is usual for a building to show several joint stones, but this is minimal in relation to the total count. Thus, in the Armentia database there were only 122 elements classified as such.

In view of this circumstance, the data from the ashlar of some of the elevations of the Santa Maria Cathedral were incorporated to enlarge the sample, because, like the rest of buildings involved in the research, they have similar characteristics. In this way, 2,695 more stones were included, 114 of which had been included in the group comprising joint ashlar. Thus, a database was finally obtained for training and testing with 5,954 ashlar, 236 of which (approximately 4%) were manually classified as being joint.

Nevertheless, with regards to optimisation in terms of creating the predictive model, it is always advisable to work with a sample that contains the same proportion of joint ashlar as normal ashlar. The fact that the complete database (5,954 elements) contained 25 times more information than was actually used (approximately

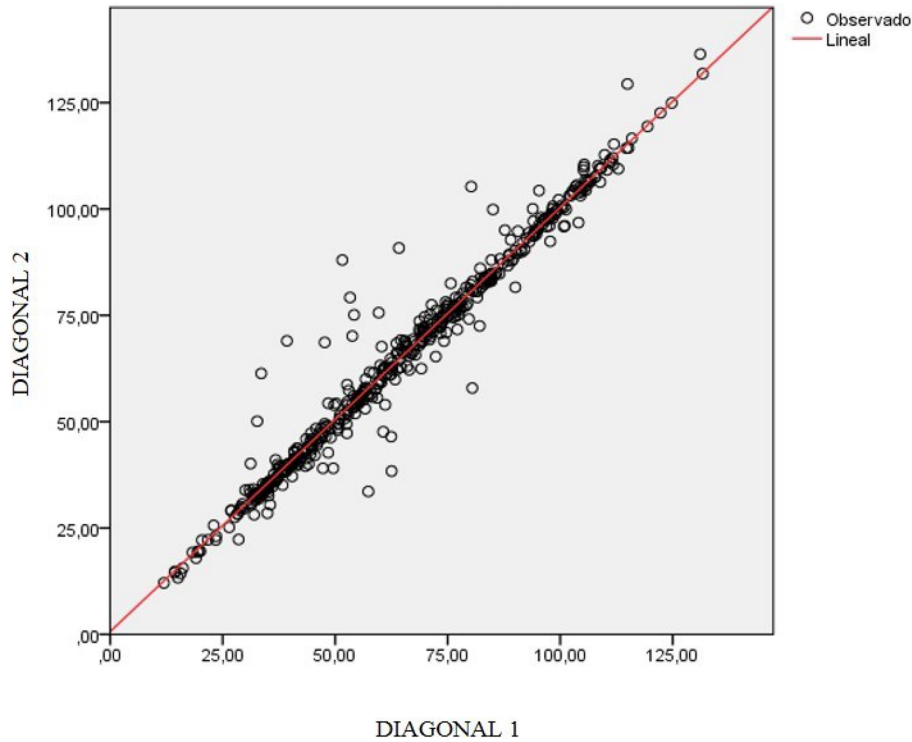


Fig. 6. Scatter plot with the dimensions of the ashlar diagonals. Most of the blocks appear to be located on the regression line. However, around a score of them appear quite far from that line. Initially these points could be considered outliers or anomalous values, but when identifying them in the corresponding elevation, we ascertained that many of them correspond to joint ashlar. This is a clear example of multivariate outliers, since the value of each of the diagonals separately does not involve anything strange, while combining both is where the possible anomalous data emerges.

500 blocks divided into two categories, ashlar and joint ashlar) allowed for experimenting with training samples that were always different, since the ashlar that were not joint could be selected randomly each time a new test was carried out.

Statistically speaking, the two main variables do not evidence any considerable asymmetries in either of the two established categories, which is logical given the homogeneity of the whole sample. The average dimensions⁴ of the joint ashlar stones are 0.311 ± 0.079 m x 0.479 ± 0.172 m (N = 236), while for regular ashlar they are 0.277 ± 0.079 m x 0.443 ± 0.174 m (N = 5,718).

As for the most significant variables in identifying these types of ashlar, taking into account the previously calculated separability values, it will be the difference in centroids between the delineated ashlar and its covering that predominantly provides the distinction between the two groups. This is followed by the difference in areas and diagonals.⁵ The latter is very interesting in terms of discrimination of both groups, as can be seen in Figure 6.

To create predictive models in the joint ashlar, the three aforementioned algorithms were applied, reaching an overall level of precision about 92%. However, the joint ashlar were not classified as well as the ashlar, since while on average 92% of the ashlar were well catalogued, only 85% of the joint ashlar were correctly classified.

⁴Also observed in this set of ashlar is homogeneity in the standard deviations of the high and long variables. Since the most irregular stone blocks have been removed from the sample, the remaining mass is much more uniform.

⁵If the ashlar has one of its corners cut out, the values of its two diagonals cannot be the same, as can be seen in Figure 6.

Table 2. Results of the Training Phase for the Categorisation of Joint Ashlars Using the Random Forest Algorithm and Taking into Account the Standardization or Not of the Variables

Number of trees	Normalized data			Unnormalized data		
	Precision (%)	Error (%)	Cohen's kappa	Precision (%)	Error (%)	Cohen's kappa
10	91.11	8.89	0.816	88.19	11.81	0.763
25	92.59	7.41	0.844	90.28	9.72	0.806
100	91.85	8.15	0.829	92.36	7.64	0.847

Again, given the characteristics of the random forest algorithm, we ultimately chose to keep the classification made by the latter, random forest algorithm. At that moment, a new component appeared that corresponds to the normalization of the variables. Until then, the tests carried out had not shown any significant differences when working with the original variables or with the transformed ones. However, it is at that moment when the databases are already more homogeneous that differences can be seen not so much in the most discriminating attributes but in the precision of the classification. For this reason, both possibilities were considered in all the tests carried out for the purpose of identifying joint ashlar.

As can be seen in Table 2, in general, the overall reliability of the adjustment is slightly better, i.e., around 2–3 points, when the variables involved are standardised beforehand, as well as the Cohen's kappa coefficient values that in all cases exceed 0.80, implying a substantial degree of classification [64]. It will be in the validation phase (Results section) where the goodness or otherwise of the predictive models obtained in this phase can really be appreciated.

At the same time as identifying the joint ashlar on the elevations of a building, it is also very interesting to be able to identify different continuity solutions or interfaces present in them. The interfaces or continuity solutions are elements that have neither volume nor materiality, but whose interest is even greater than that of the strata [4]. If they stand out for something, it is because of their great value in terms of timescale, because by delimiting two strata they allow the before and after relationship that exists between them to be deduced [45]. In conjunction with the delineation of the interfaces, for the joint ashlar, we also considered it interesting to indicate the ashlar they supported (Figure 7, left), because the upper ashlar will always be the most modern, indicating the chronological sequence. Therefore, this facilitates the understanding of the building's historical evolution.

To develop the research, we considered that this information should appear together with the identification of the joint ashlar in the walls. This was programmed in MATLAB and with the help of the DXFLib library, a routine that would allow for the automatic delineation of these lines. This was possible because, in our previous classification work, we delineated the contours of all the stone blocks that made up the building, obtaining a vectorial DXF file. Because that file contains the coordinates of all the points that form those outlines, it facilitates the analysis of the vertical continuity of the stones' edges, marking those areas where two or more ashlar coincide. In the construction of a building, it is common to position the blocks in an alternating fashion to ensure the cohesion and resistance of the facing and to coincide with openings (doors or windows) or the ends of the walls.

3 RESULTS AND DISCUSSION

Application of the predictive techniques on the elevations of the churches had two very specific purposes: the differentiation between ashlar work and masonry, and the continuity analysis of the courses of stone. This evidence, together with identification of the vertical interfaces, are key indicators when undertaking a study of the historical-structural evolution of a building in accordance with the stratigraphic analytical method used.

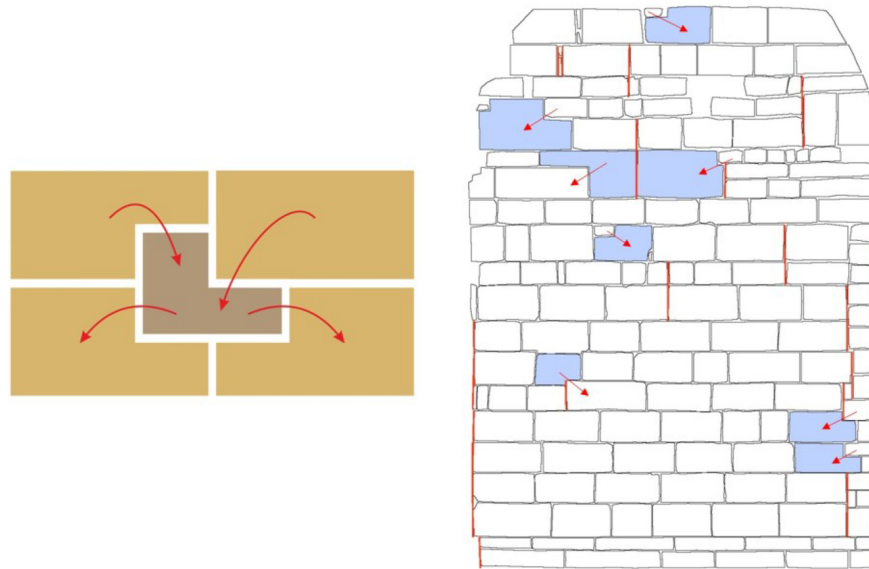


Fig. 7. At the left of the figure, chronological relative relationship of several ashlars. The red arrow indicates the relative chronological relationship with the stone that fits on them. At the right, elevation of San Vicentejo Hermitage with the joint ashlars automatically identified (blue blocks), their chronological relation (red arrows) and the programmed interfaces (red lines).

We will apply the classification rules obtained in the experimental stage to the stones of the San Vicentejo hermitage and we will analyse their behaviour as well as their usefulness in reading elevations.

The church, which is small in size—approximately 18×8 m—is made up of some 4,550 stone blocks. Statistically speaking, the masonry blocks of this temple ($N = 222$) are 0.131 ± 0.075 m \times 0.183 ± 0.099 m in size while the dimensions of the ashlar work are ($N = 4,316$) 0.264 ± 0.111 m \times 0.527 ± 0.284 m. While the former are considerably higher in this building than those belonging to the training sample, they coincide in length. In the case of the ashlars, the exact opposite is true: In height, the dimensions coincide in both churches, but it is in terms of length that San Vicentejo has higher values.⁶

3.1 Discriminating between Ashlar work and Masonry

The predictive models obtained using the random forest were applied, as were those that created the best classification models in the learning phase. Until that point, the classification had been quite efficient, based on the results obtained, but it is now that they really need to be corroborated, when applying the knowledge generated.

For validation of the classification of ashlars and rough stones, the overall reliability obtained (94.87%) was slightly worse than in training (99.90%) (Table 3). This is logical, as they constitute completely different data, although despite this the reliability is very satisfactory, which leads us to think of good categorisation for the

⁶Although it is not the specific objective of this article, we believe it is interesting to point out a fact that is highlighted by the data available to us: The greater the dimensions of the ashlar blocks, the closer the church is to the quarry [65]. This is logical, considering that minimizing transportation has also been a priority for stone used in dressed ashlar work and sculpture [65]. As an example, while the stones of the Basilica of Armentia and the church of Treviño (both located about 13 km from the Ajarte quarry, where the stone comes from) have an average length of 42 cm, for San Vicentejo, located just 6 km from the quarry, the ashlars are almost 10 cm longer on average, with 53 cm, and the church of the village of Ajarte itself, located a few metres from the quarry, has even longer ashlars. This is only a line of research that requires more detailed tests, but it is important, because it can provide a lot of information about building techniques used by stonemasons in the Middle Ages.

Table 3. Results of the Training and Validation Phase for the Categorisation of Ashlars and Masonry Using Classification Trees, Cross-validation, and Random Forest

	TRAINING PHASE						VALIDATION PHASE					
	Classification trees		Cross-validation		Random forest		Classification trees		Cross validation		Random forest	
	Ashlar	Masonry	Ashlar	Masonry	Ashlar	Masonry	Ashlar	Masonry	Ashlar	Masonry	Ashlar	Masonry
Precision	0.97	0.95	0.97	0.96	1.00	0.99	0.99	0.47	0.99	0.47	0.99	0.49
Recall	0.95	0.97	0.96	0.97	0.99	1.00	0.95	0.86	0.95	0.85	0.95	0.85
f1-measure	0.96	0.96	0.97	0.97	0.99	0.99	0.97	0.61	0.97	0.61	0.97	0.62

Table 4. Confusion Matrix as a Result of Applying the Predictive Models Obtained Using the Random Forest Algorithm for Categorisation of Ashlars and Rough Stones

		Prediction	
		Ashlars	Rough stones
Reality	Ashlars	4,117	199
	Rough stones	34	188

stones. Although it is estimated that validation precision may be less than that of training, this difference should not be greater than 2%–3%. For our data purposes, it was slightly higher, around 5%, but it should be noted that the results obtained in the first phase with this algorithm were surprisingly good.

The average value of the kappa index is 0.59, a record that, although it allows us to affirm that the classification is “moderate” according to the Landis and Koch table [64], is considerably lower than that obtained in the training phase, where values of around 0.99 were obtained. The kappa value will be higher as long as the values that appear on the diagonal of the confusion matrix are similar [67]. This is not the case here, as, contrary to what happened in the learning phase, the number of ashlar in the elevations of San Vicentejo is much higher than that of rough stones: 4,305 as opposed to only 233 (Table 4).

In this case, as the data was so unbalanced, i.e., most of the elements belong to one of the categories, the overall reliability values, while good, do not faithfully reflect the efficiency of the classifier. It is necessary to go back to calculation of producer and user reliability to analyse the results as in the training phase. In terms of labelling rough stones, the number of errors is now somewhat greater (FP = 0.49 as opposed to a value of practically 1 in the previous phase). This disparity is fundamentally due to the characteristics of the new database with which it is being validated, as the rough stones clearly represent the minority class, accounting for only 5% of the total.

If, instead of trying to classify all the blocks from the hermitage of San Vicentejo, a sub-sample was extracted in which there were the same number of ashlar as there were rough stones—these figures would change considerably, acquiring very similar values between them, equal to those obtained in the training phase. However, in this phase, such a division is not possible, as this would mean classifying the entire building.

Thus, while in the categorisation of ashlar there are only around 30 errors out of a total amount of more than 4,000 blocks—which means a producer reliability of 99%—with rough stones, the 199 badly classified elements imply a reliability of barely 48%. However, user reliability, i.e., the percentage of each real category that has been properly classified, is quite high in both classes; namely, 99% for the ashlar work and 85% for the masonry.

In addition to the analytical results, we can analyse the result obtained graphically by visualizing on the hermitage’s elevations which elements are poorly classified and researching the cause of such errors (Figure 8).

The stones that correspond to masonry are generally well classified; the problem lies with the ashlar. While the rough stones evidence total irregularity that allows them to be easily differentiated from the rest, in the

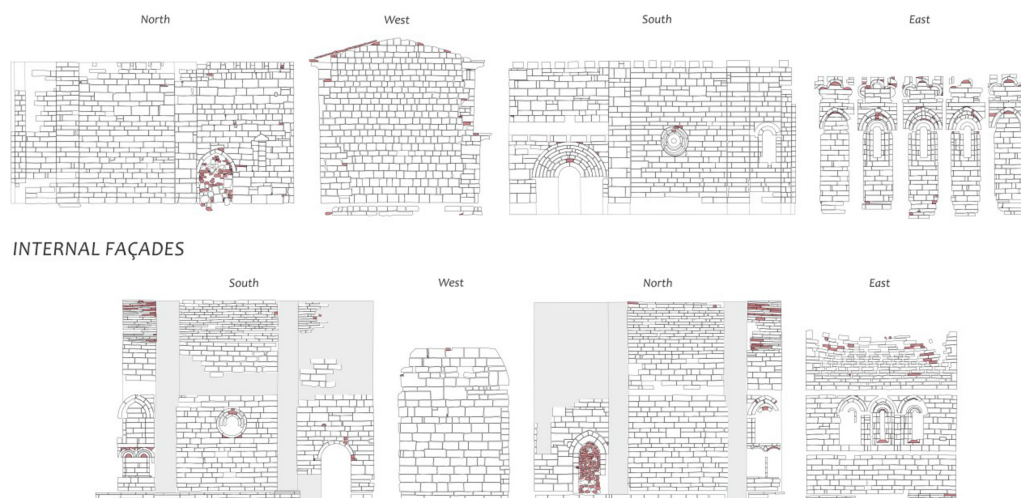


Fig. 8. Rough stones automatically identified in the Hermitage of San Vicentejo. Most of these blocks would have been located in the covered-up doors (north facades). Although there are some stones that have been poorly identified, all of them account for only 5% of the total.

ashlars there is a progression that goes from the most regular block, with all its perfectly squared corners, to the block that has rounded angles, which sometimes leads to poor classification.

Focusing on the ashlar that have been poorly classified, the majority of them correspond to irregular blocks of stones, such as the voussoirs of arches, the stones that surround those arches or the blocks that are not in a strictly horizontal position, as is the case with those arranged in the vaults. All these examples for one reason or another have an irregular shape in common. It is worth remembering that between the attributes that enabled distinction to be made between ashlar and rough stones, a highly influential one was the difference in surface area between the real shape of the stone and its covering. It is therefore logical that these elements should be identified as rough stones.

An error when classifying rough stones occurs when the block is of large dimensions, because these types of elements are usually small in size and easy to use by hand. In San Vicentejo, in the walled-up bay window on the north wall, although the great majority of the stones that appear are small to medium-sized rough stones, it can be seen that pieces from larger stones were reused—very likely deriving from another part of the construction and that were not classified by the model as rough stones.

3.2 Validation of the Identification of Joint Ashlars

In this trial, the aim of which was to identify joint ashlar, from the training phase it was seen that the percentage of success and error varied according to whether transformed data was used or not, most likely due to a greater homogeneity in the database, as by this point all the rough stones had been removed from the sample, and the remaining ones were mostly regular ashlar.

Therefore, in this validation phase, the predictive models obtained with the standardised and non-standardised data were applied, and they show quite different behaviour. Comparing the values obtained in this validation with those recorded in the training phase, classification precision is similar, with hardly any notable differences. Where an important decrease can be observed is in the value of the kappa coefficient, which ranges from 0.80, implying a substantial classification, to values of just 0.33. Although this is due to the inequality between the two categories, a fact that has also occurred in previous analyses, it is now when the difference between both groups becomes greater and directly leads us to very low values of this coefficient.

Table 5. Confusion Matrix as a Result of Applying the Predictive Models Obtained Using the Random Forest Algorithm for the Categorisation of Joint Ashlars with Unnormalized Data

		Prediction	
		Ashlars	Joint ashlar
Reality	Ashlars	3,458	266
	Joint ashlar	6	74

Table 6. Confusion Matrix as a Result of Applying the Predictive Models Obtained with Random Forest Algorithm for Categorisation of Joint Ashlars with Normalized Data

		Prediction	
		Ashlars	Joint ashlar
Reality	Ashlars	3,711	13
	Joint ashlar	49	31

The confusion matrix (Table 5) also produced some very interesting conclusions. In this first block, what was deduced is that the classifier worked exceptionally well in identifying the joints that appear in the San Vicentejo database; because of the possible 80, it correctly identified 74 of them—a success percentage of 92% (RU). But what must also be taken into consideration is the fact that it incorrectly labelled 266 blocks in this category, resulting in an error of omission close to 80% (RP), a dreadful percentage. To sum up the joint ashlars in the validation sample were very accurately identified, but many errors were made; many ashlars were labelled as being joint when actually they were not.

In the second trial, using normalized data, an improvement in the overall reliability regarding training was observed (Table 6). Again, the values of the kappa index were affected by the inequality of the categories, albeit not as much as with non-standard data. They fall to values of 0.49, which can be considered a moderate classification.

Here the previously trend changes completely. While a little over 40% of the joints in the elevations of San Vicentejo (RU) barely classify well, nevertheless not too many incorrect values are inputted—only 13—which is well below the previous example of 266. Logically, the statistics provided around the minority block were just as affected as in the previous round of trials. To sum up, there are fewer correctly identified joints and also few errors.

As we have already mentioned, the presence of joint ashlars in an elevation may be an indication of some kind of action carried out on it. As a result, the appearance of many incorrectly classified ashlar as such provides very little useful information for the researcher, since we would be pointing out possible structural actions that never really existed. Therefore, to keep the first of the options would mean, in our view, including a lot of “noise” in the stratigraphic analysis process.

Choosing the second option provides us with a certain security, as we know that the vast majority of ashlar classified as being joint correspond to this label, and that the number of errors entered in the reading process is lower.

Regarding the location of the errors that occur when identifying the joint ashlar, the prediction model works quite well with the joints that actually exist in the elevations and the errors are fundamentally in the windowsills. This is where it is necessary to lower the stone to be able to make the shape that they really have, which means these blocks bear a certain resemblance to the joints.

3.3 Discussion

J. A. Barceló said in his publication *Automatic archaeology. Artificial Intelligence in Archaeology* [68], “At the present time, automatic archaeology is more of a future promise than a reality; it is an unexplored field that will provide extraordinary and unpredictable results in the coming years.”

Such a statement—made more than 20 years ago—is today closer to becoming a reality, thanks to the technical progress of the past few decades that is allowing aspects such as artificial intelligence, expert systems, and machine learning to become increasingly present in many diverse areas of society, making it look increasing less like science fiction.

When—even before the present work—we were beginning with this line of research, we were firmly convinced that artificial intelligence was possible in archaeology, and more specifically in built heritage [47]. Today, not only have we not given up on this idea, but we are continuing to study it more in depth. The aim of this work is for it to be a first step towards the promising future that Barceló predicted, although there are still lots of things to do. Our objective of facilitating automation of some archaeological processes is not intended to replace either the archaeologist or the architect, but simply to provide them with an auxiliary tool to help them in the stratigraphic analysis of a building—a task not without its complexity.

Throughout our research, we have confirmed how archaeologists are often excessively “impressionistic” when describing the reasoning that leads to their stratigraphic conclusions about a building. This was one of the great obstacles we inevitably had to overcome, namely, the difficulty of clearly visualizing the inference mechanisms used by the researcher. This is because any attempt to automate a technique requires—first—having a clear “flow chart” of the internal process of the technique to be systematised.

Be that as it may, if we had to highlight a single key aspect as the fundamental driving force behind this internal process involving the stratigraphic analysis technique, that would be the typological aspect. Logically, we cannot expect to reduce all the complexity involved in the stratigraphic reading of elevations to a single matter, but we do so because, just like Barceló, we are convinced that the success of the application of expert systems in archaeology depends on the implementation of automatic typologies [68]. This will make it easier for archaeologists to input mathematical-statistical descriptions of their discoveries into databases, so it may be the expert system that deconstructs the object of study stratigraphically.

Some examples of automatic typologies have existed for some time in archaeology; they relate to ceramics, the lithic industry, fibulas, and the orientation of tombs, among many other subjects of interest. Even so, we hardly find anything related to built heritage. There are, of course, advances made in the field of mensiochronology—both in brick and stone—that proved fundamental in planning this research. However, they do not contemplate the use of multivariate techniques, but rather, only basic statistical procedures, as the study population was minimal—just like the variables that defined it, which can be summarised in the dimensions of the pieces analysed.

If the typological key is fundamental for automation of the stratigraphic reading of elevations, it is also because the typology is, as it were, the active principle that allows us to analyse thoroughly, to deconstruct the building—in an organised and systematic manner—to transform all its components into magnitudes, into figures that may then be subject to mathematical-statistical treatment.

4 CONCLUSION

Every building of a minimum age is considered by archaeological science to be a multi-layered site, because its structure is the result of successive building operations involving contribution and degradation that have been superimposed, leaving an indelible mark on its walls. Reading of elevations is the most common technical procedure used to determine this stratification, which has so far been carried out by visual analysis. This analysis introduces a high degree of subjectivity into the process since, depending on the researcher’s expertise or experience, the result may vary. In this article, we introduce a framework that seeks to use machine learning to

semi-automate the process of reading elevations. Based on the most common evidence that specialists in the archaeology of architecture were looking for in the elevations of buildings, we tried to make the machine responsible for this by identifying and pointing out the different types of stones, the existence of joint ashlar and interfaces, voussoirs of arches, and even clusters that simulate construction phases. In this way, we eliminate the subjectivity with which the method started, and we incorporate far more objective procedures that allow for the stratigraphic analysis of the building in an identical way for all researchers. It will be later when the specialists with all this evidence then go on to undertake their historical interpretation of the constructive evolution of the historical building. All the buildings that have been used in this research shared their lithology, the same quarry of origin, and so as long as more buildings with similar characteristics can be incorporated (which is relatively easy to do, as it is a widely used lithology in Alava), the system will be able to continue learning and differentiating between building signs with greater precision. This does not mean that the method proposed only works with Palaeocene limestone, as the method itself can be extrapolated to any building, albeit not the absolute values obtained from it. In other words, the most discriminatory variables will remain the same, but not the value they acquire.

In this article, we have shown how to train the system to identify and point out the masonry and ashlar, the joint ashlar, and the interfaces on the elevations, obtaining good results in all cases, by starting from the premise that this issue had never been tackled before. It is important to see how the passive role that, until now, the geometric documentation of the building had acquired in a project involving reading elevations, changes radically and becomes an active part of it, since it provides the data required for that autonomous learning. We experimented with several algorithms, but, ultimately, we opted for the random forest, since it provided the best reliability in both the training and validation phases.

Regarding predictive models, these worked slightly better in the training phase than in the validation phase. However, there is an obvious explanation for this: To a large extent, this is due to the imbalance between the categories defined in the San Vicentejo hermitage database, because the elements to be classified (rough stones and joint ashlar) always represented the minority group within the sample. This imbalance was reflected in the precision obtained when classifying the different categories. Thus, while for the majority group—the ashlar—user reliability exceeds 95%, in terms of discrimination of the rest of the elements this value varies between 40% and 90%. In the classification of rough stones, the resulting overall reliability provides values around of 95%. Even so—and this is the important thing—the predictive model created correctly classifies 85% of the building’s rough stones and 99% of the ashlar.

The case of the joint ashlar is somewhat different, as we had two predictive models—one that was obtained using standardised data and the other that was obtained using non-standardised data. If we only pay attention to the numerical values, it seems logical to opt for the second model, where precision, when it came to classifying these type of blocks, was 92% with a 7% error in the classification of normal ashlar, compared to the 40% obtained in the first case for the ashlar, albeit with only 0.4% error in terms of their categorisation. However, we have chosen to keep the model that, *a priori*, offers the least precision, given the importance of these types of pieces in dating possible alterations carried out on the elevation. When interpreting the structural evolution of a building, it is preferable to have few joint ashlar—that are well classified and that provide information—rather than many erroneous ones that may cause confusion.

4.1 Future Work

As we have commented in the article, this is the first time that machine learning has been used to study a historical building from the archaeological standpoint, so logically there is still room for improvement. We will focus future works around two different lines: the variables and the buildings.

For our research, we generated the different predictive models from a set of 33 variables, some of them qualitative, but mostly quantitative. One line of improvement involves the incorporation of new variables related to

radiometric aspects, since, if both visible and thermal image information are available, the digital levels of the images may help in the classification process. For example, if we work with buildings made of bricks, the colour may provide us with information about the amount of clay they have, the extraction point of the raw material, or even the firing temperature. The different reflective and emitting behaviour of the walls allows us to know the composition of the materials used in their construction, as well as the moisture or the existence of materials “hidden” from human sight.

Another aspect to point out and which we have already advanced in the conclusions, is the possibility of testing this learning on more buildings. On one hand, it would be interesting to continue working on buildings whose material comes from the same quarry in Ajarte to ascertain whether the increase in the number of individuals analysed improves the classification obtained so far or otherwise. This would involve mainly trying to obtain data for those minority categories, such as the joint ashlar or the voussoirs of the arches. And, on the other hand, there is the possibility of applying the methodology to buildings whose construction material is totally different. In this sense, we have gained a little experience in buildings made entirely of brick in the city of Buenos Aires, and we will continue to develop this, since the initial results obtained are quite encouraging.

REFERENCES

- [1] Agustín Azkarate, Luis Caballero, and Juan Antonio Quirós. 2002. Arqueología de la arquitectura: definición disciplinar y nuevas perspectivas. *Arqueol. Architect.* 1 (2002), 7–10.
- [2] Kate Giles. 2014. Buildings archaeology. In *Encyclopedia of Global Archaeology*, C. Smith (Ed.).– Springer, New York, 1033–1041.
- [3] Agustín Azkarate. 2020. Archaeology of architecture. In *Handbook of Global Archaeology*, Ch. E. Orser, P. Funari, S. Lawrence, A. Zarankin (Eds). Routledge, Taylor & Francis.
- [4] Eduard C. Harris. 2017. Harris matrices and the stratigraphic record. In *Encyclopedia of Geoarchaeology. Encyclopedia of Earth Sciences Series*, A. S. Gilbert (Ed.). Springer, Dordrecht New York, London, 404–410.
- [5] Juan Antonio Quirós. 1996. Indicadores cronológicos de ámbito local: Cronotipología y mensiocronología. In *Arqueología de La Arquitectura. el Método Arqueológico Aplicado al Proceso de Estudio E Intervención en Edificios Históricos*, L. Caballero and C. Escribano (Eds.). –Junta de Castilla y León, Burgos, 1–13.
- [6] Roberto Parenti. 1995. Historia, importancia y aplicaciones del método de lectura de paramentos (History, importance and applications of the stratigraphic method). *Informes de la Construcción* 46 (1995), 19–29.
- [7] Daniela Pittaluga. 2009. La mensiocronologia dei Mattoni. *Per datafire, Per Conoscere e Per Comprendere le Strutture Storiche*. ECIG.
- [8] Nancy M. Mangialardi. 2015. The dimensional analysis of the bricks. *The Case Study of Lucera between the Thirteenth and Fourteenth Century*. Benevento, Italia.
- [9] Lorenzo Cantini, Mattia Previtali, Rossella Moioli, and Stefano Della Torre. 2019. The mensiochronology analysis supported by the new advanced survey techniques: Field tests in Milanese area. *Arch. Photogramm. Remote Sens. Spatial Inf. Sci* XLII(2/W11), 359–365.
- [10] José Ángel Asensio. 1995a. Arquitectura de tierra y madera en la protohistoria del valle medio del ebro y su relación con la del mediterráneo. *Caesaraugusta* 71 (1995), 23–56.
- [11] José Ángel Asensio. 1995b. La ciudad en el mundo prerromano en aragón. *Caesaraugusta* 70 (1995), 1–461.
- [12] David Andrews. 1988. La muratura medievale. Descrizione, analisi e storia economica. *Castrum* 2. Structures de l’habitat et occupation du sol dans les pays Méditerranéens: Les methodes et l’apport de l’archeologie extensive. Actes de la rencontre organisée par l’ Ecole Française de Rome, Ghislaine Noyé (Ed.). Casa de Velázquez editores, 309–317.
- [13] Tiziano Mannoni and Ennio Poleggi. 1974. Fonti scritte e strutture medievali del “castello” di Genova. *Archeologia Medievale* 1 (1974), 171–194.
- [14] Tiziano Mannoni, Aurora Cagnana, Stefano Falsini, Paola Ghislanzoni, and Daniela Pittaluga. 1991. Archeologia ed archeometria dei muri in Pietra: Superfici e strutture in liguria. In *Le Pietre Nell’architettura: Struttura e Superfici: Atti del Convegno Di Studi, bressanone*. Libreria Progetto Editore, 151–162.
- [15] Aurora Cagnana and Tiziano Mannoni. 1995. Materiali e tecniche nelle strutture murarie di Luni. *Risultati preliminari. Quaderni di Studi Lunensi* 1(1995), 137–164.
- [16] Stefano Della Torre. 1996. Storia delle tecniche murarie e tutela del costruito. *Esperienze e Questioni di Metodo*. Guerina e Associati, Brescia, Milano.
- [17] Nicoletta Benacchio and Paolo Castellani. 1997. Indagine sulle tipologie murarie di Assisi tra XI e XIV secolo: primi risultati. *Archeol. Archit.* 2 (1997), 121–130.
- [18] Yves Esquieu. 1997. L’archéologie du bâti en France. *Archeol. Archit.* 2 (1997), 133–140.
- [19] Michele Nucciotti. 2000. Le murature medievali di Santa Fiora (Monte Amiata-Toscana): Mensiocronologia dell murature in Pietra, un caso di studio. *Archeol. Archit.* 5 (2000), 65–85.

- [20] Daniel Prigent and Christian H. Sapin. 2004. La construction en pierre au moyen âge. In *La Construction. Les Matériaux Durs: Pierre Et Terre Cuite*. Errante ediciones, 117–148.
- [21] Giuseppe Fichera. 2005. Archeologia dell'architettura di un edificio ecclesiastico: La chiesa di Sant'Antimo sopra i canali a piombino. *Archeol. Archit.* 10 (2005), 153–171.
- [22] Gian Pietro Brogiolo and Aurora Cagnana. 2012. *Archeologia dell'Architettura—Metodi e Interpretazioni*. All'Insegna del Giglio.
- [23] Agustín Azkarate, Ismael García-Gómez, and Amaia Mesanza-Moraza. 2018. Análisis clúster: Un primer paso sobre técnicas cuantitativas en arqueología de la arquitectura. *Arqueol. Arquít.* 15 (2018), 1–18.
- [24] Agustín Azkarate, Leandro Cámara, Juan Ignacio Lasagabaster, and Pablo Latorre. 2001. *Plan Director de Restauración de la Catedral de Santa María de Vitoria-Gasteiz*. Diputación Foral de Álava, Vitoria-Gasteiz.
- [25] Agustín Azkarate and José Luis Solaun. 2013. *Arqueología e Historia de Una Ciudad. Los Orígenes De Vitoria-Gasteiz*. Vitoria-Gasteiz. Servicio editorial de la Universidad del País Vasco/Euskal Herriko Unibertsitatea.
- [26] Leandro Sánchez-Zufiaurre. 2007. Técnicas constructivas medievales. *Nuevos Documentos Arqueológicos Para el Estudio de la Alta Edad Media en Álava*. Servicio Central de Publicaciones del Gobierno Vasco, Vitoria-Gasteiz.
- [27] Alberto Plata. 2008. Génesis de una villa medieval. *Arqueología, Paisaje y Arquitectura del Valle Salado de Añana (Álava)*. Servicio Central de Publicaciones del Gobierno Vasco, Vitoria-Gasteiz.
- [28] Alberto Plata and Andoni Erkiaga. 2018. *El Sistema de Producción de Sal de Añana: Valle Salado (Araba-Álava. País Vasco)*. Colección Patrimonio, Territorio y Paisaje. Vol. 3. Servicio editorial de la Universidad del País Vasco/Euskal Herriko Unibertsitatea, Vitoria-Gasteiz.
- [29] Luis Caballero. 2009. Edificio histórico y arqueología: Un compromiso entre exigencias, responsabilidad y formación. *Arqueol. Arquít.* 6 (2009), 11–19.
- [30] Bryan F. J. Manly. 2005. *Multivariate Statistical Methods. A Primer* (3rd ed.). CRC Press.
- [31] Guillermo Rolón. 2013. *La Vivienda Popular Riojana del Ambito Rural: Patrones Arquitectónicos y Contexto Social en los Valles durante el Periodo Republicano*. PhD Dissertation. Universidad de Buenos Aires, Argentina.
- [32] Francisco Contreras. 1984. Clasificación y tipología en arqueología. In *El Camino Hacia La Cuantificación. Cuadernos de Prehistoria y Arqueología de la Universidad de Granada* 9 (1984), 327–385.
- [33] Guillermo Rolón. 2014. Patrones arquitectónicos, clústeres constructivos homogéneos y variabilidad en el estudio de edificios históricos. Aspectos técnico-formales de la vivienda rural en la provincia de La Rioja (Argentina) durante el periodo republicano. *Arqueol. Arquít.* 11 (2014), 1–24.
- [34] Gayane Shalunts, Yll Haxhimusa, and Robert Sablatnig. 2011. Architectural style classification of building facade windows. In *Proceedings of the 7th International Symposium on Visual Computing*. 280–289.
- [35] Gayane Shalunts, Yll Haxhimusa, and Robert Sablatnig. 2012. Architectural style classification of domes. In *Proceedings of the 8th International Symposium on Visual Computing*. 420–429.
- [36] Maarten Bassier, Maarten Vergauwen, and Bjorn Van Genechten. 2017. Automated classification of heritage buildings for as-built BIM using machine learning techniques. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume IV(2/W2), 26th International CIPA Symposium 2017, Ottawa, 25–30.
- [37] Kun Guo and Ning Li. 2017. Research on classification of architectural style image based on convolution neural network. In *Proceedings of the IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC'17)*. 1062–1066.
- [38] Matthias Zeppezauer, Miroslav Despotovic, and Muntaha Sakeena. 2018. Automatic prediction of building age from photographs. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR'18)*. 126–134.
- [39] Luis Miguel Martínez-Torres. 2014. Lithologic combinations in Romanesque churches of Alava, Northern Spain. *Mater. Construcc.* 64, 313 (2014), 1–6.
- [40] Manuel Castiñeiras. 2012. San Vicentejo de Treviño, un edificio excepcional en la encrucijada del tardorrománico hispánico. In *Viaje a íbita, ed. Ayuntamiento del Condado de Treviño* (1st ed.). Condado de Treviño, España, 230–328.
- [41] Juan Ignacio Lasagabaster, Agustín Azkarate, and Ander de la Fuente. 2006. Plan director de intervención para la restitución de los valores históricos y simbólicos de la basílica de san prudencio de armentia y su entorno. Grupo de Investigación en Patrimonio Construido, UPV/EHU, 1–176.
- [42] Peter Swallow, Ross Dallas, Sophie Jackson, and David Wall. 2004. *Measurement and Recording of Historic Buildings* (2nd ed.). Donhead, London.
- [43] David Andrews, Jon Bedford, Bill Blake, Paul Bryan, Tom Cromwell, and Richard Lea. 2010. *Measured and Drawn Techniques and Practice for the Metric Survey of Historic Buildings*. English Heritage, Swindon.
- [44] Luis Caballero. 2010. Experiencia metodológica en arqueología de la arquitectura de un grupo de investigación. In *Arqueología Aplicada Al Estudio e Interpretación de Edificios Históricos. Últimas Tendencias Metodológicas Ministerio de Cultura* (Ed.). Secretaría General Técnica. Subdirección General de Publicaciones, Información y Documentación, 103–120.
- [45] Luis Caballero. 1995. Método para el análisis estratigráfico de construcciones históricas o “lectura de paramentos.” *Inf. Construcc.* 46, 435 (1995), 37–46.

- [46] Irantzu Álvarez-González, Ane Lopetegui, Amaia Mesanza-Moraza, José Manuel Valle, and Iratxe Vicente. 2003. Diferentes propuestas para la representación geométrica de edificios históricos. *Arqueol. Arquit.* 2 (2003), 9–12.
- [47] Ismael García-Gómez, Miren Fernández de Gorostiza, and Amaia Mesanza-Moraza. 2011. Láser escáner y nubes de puntos. Un horizonte aplicado al análisis arqueológico de edificios. *Arqueol. Arquit.* 8 (2011), 25–44.
- [48] Restituto Sierra. 1994. *Técnicas de Investigación Social. Teoría y Ejercicios*. Paraninfo, Madrid.
- [49] Agustín Azkarate. 2010b. Estudio histórico-arqueológico del casco histórico de Vitoria-Gasteiz. *Protocolos de intervención, modelos de gestión y sistema de información patrimonial (SWIP)*. Grupo de Investigación en Patrimonio Construido. UPV/EHU.
- [50] Agustín Azkarate. 2010a. El análisis estratigráfico en la restauración del patrimonio construido. Consideraciones conceptuales e instrumentales. In *Arqueología Aplicada al Estudio e Interpretación de Edificios Históricos. Últimas Tendencias Metodológicas*. Ministerio de Cultura (Ed.). Secretaría General Técnica. Subdirección General de Publicaciones, Información y Documentación, 51–64.
- [51] Gian Pietro Brogiolo. 2010. Introduzione. In *Archeologia dell' Architettura* 15 (2010), 11–14.
- [52] Noelia Oses, Fadi Dornaika, and Abdelmalik Moujahid. 2014. Image-based delineation and classification of built heritage masonry. *Remote Sensing* 3 (2014), 1863–1889.
- [53] David S. Moore. 2004. *Estadística Aplicada Básica*. (2nd ed), Antoni Bosch (Ed.). Barcelona.
- [54] Pablo Rodríguez-González, Diego González-Aguilera, David Hernández-López, Hector González-Jorge. 2014. Accuracy assessment of airborne laser scanner dataset by means of parametric and non-parametric statistical methods. *IET Sci., Meas. Technol.* 9, 4 (2014), 505–513.
- [55] Charlotte H. Mason and Willian D. Perreault. 1991. Collinearity, power and interpretation of multiple regression analysis. *J. Market. Res.* 28 (1991), 220–268.
- [56] Felipe Oliva and Patricia Verdú. 2008. Revisión de la utilización de distintos índices espectrales y técnicas para la cartografía de área quemada en imágenes modis. *Uned. Espacio, Tiempo y forma. Serie VI, Nueva Época. Geografía* 1:129–152.
- [57] Michael R. Berthold, Christian Borgelt, Frank Höppner, and Frank Klawonn. 2010. *Guide to Intelligent Data Analysis. How to Intelligently Make Sense of Real Data*. Springer, London.
- [58] Jorge García-Gutiérrez. 2012. *Intelligent Techniques on LiDAR for Environmental Applications*. PhD Dissertation. Universidad de Sevilla.
- [59] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2014. *Introduction to Data Mining*. Pearson.
- [60] Lior Rokach and Oded Maimon. 2015. *Data Mining with Decision Trees. Theory and Applications* (2nd ed.). World Scientific Publishing Co. Pte. Ltd.
- [61] Leo Breiman. 2001. Random Forests. In *Machine Learning* 45 (2001), 5–32.
- [62] Leo Breiman. 1996. *Out-of-Bag Estimation*. University of California, Berkeley.
- [63] John James. 1989. The template-makers of the Paris basin. *Toichological Techniques for Identifying the Pioneers of the Gothic Movement with an Examination of Art-historical Methodology*. West Grinstead Publishing, Australia.
- [64] J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33 (1977), 159–174.
- [65] Ismael García-Gómez. 2017. *Vitoria-Gasteiz y su Hinterland. Evolución de un Sistema Urbano entre los Siglos XI y XV*. Colección Patrimonio, Territorio y Paisaje. Servicio editorial de la Universidad del País Vasco/Euskal Herriko Unibertsitatea.
- [66] Luis Miguel Martínez-Torres. 2007. Lithological maps of churches in the Diocese of Vitoria (Spain): Space-time distribution of building stones and ancient quarries. *Build. Environ.* 45 (2007), 860–865.
- [67] Charles A. Lantz and Elliott Nebenzahl. 1996. Behaviour and interpretation of the k statistic: Resolution of the two paradoxes. *J. Clin. Epidemiol.* 49 (1996), 431–434.
- [68] Juan Antonio Barceló. 1996. Arqueología automática, Inteligencia artificial en arqueología. *Cuadern. Arqueol. Mediterr.* 2 (1996), 13–149.

Received April 2020; revised August 2020; accepted September 2020