*Article*

# Supervised Machine Learning Algorithms to Discriminate Two Similar Marble Varieties, a Case Study

**Lluís Casas** [1,*] **, Anna Anglisano** [1] **, Roberta Di Febo** [1] **, Berta Pedreño** [1] **and Ignasi Queralt** [2]

1 Department of Geology, Campus de la UAB, Autonomous University of Barcelona, 08193 Barcelona, Catalonia, Spain; anna.ar.93@gmail.com (A.A.); roberta.difebo@uab.cat (R.D.F.); berta.pedreno@uab.cat (B.P.)

2 Department of Geosciences, Institut of Environmental Assessment and Water Research, Consejo Superior de Investigaciones Científicas (IDAEA-CSIC), Jordi Girona 18-26, 08034 Barcelona, Catalonia, Spain; ignasi.queralt@idaea.csic.es

* Correspondence: lluis.casas@uab.cat

**Abstract:** A multi-analytical approach is usually applied in provenance studies of archaeological marbles. However, for very similar marble varieties, additional techniques and approaches are required. This paper uses a case study to illustrate this with two Catalan marble districts (Gualba and Ceret) and three sets of archaeological marbles. The common multi-method approach is unable to discriminate between the two districts, and such distinction is only partially glimpsed using unsupervised multivariate data analyses on a transformed geochemical dataset of reference samples. In contrast, several supervised classification models have been successfully trained to discriminate between the quarries without any special data transformation. All the trained models agree to assign the three sets of archaeological samples to the Gualba quarry district. Additional outcomes of the paper are a comprehensive archaeometric characterization of the little-known marbles of Gualba and Ceret and the first archaeometrically supported evidence of the use of Gualba marble during Roman and Medieval times.

**Keywords:** marble; provenance studies; supervised methods; machine learning; clustering; XRF; heritage science

## 1. Introduction

Provenance studies of stone materials are one of the main applications among the wide range of disciplines covered by archaeometry. In the case of white marbles, the widespread occurrence of the material and its rather homogenous color and composition complicates the identification of their origin. Usually, marble provenance cannot be designated by macroscopic criteria alone. However, a multi-technique approach to ascertain the provenance of such materials is well established; it usually consists of a combination of petrographic and cathodoluminescence characterization and the determination of the stable carbon ($\delta^{13}C$) and oxygen ($\delta^{18}O$) isotopic ratios [1–3]. The relatively few high-quality sources of marble quarried in Antiquity enables the distinction of the main so-called classical marbles using the common multi-technique approach. However, the number of documented minor quarry sites and the study of their marble is continuously expanding. This results in an increased chance of overlapping properties between marbles from different quarries. To overcome this possibility, additional analyses and characterization techniques are coming into play, such as elemental ratios [4], Sr isotopes [5], electronic paramagnetic resonance [6] or nuclear magnetic resonance [7], among many others [8]. Some methodologies have even been developed to specifically discriminate between two similar lithotypes. That is the case of Carrara and Göktepe fine-grained marbles, which can be distinguished by measuring the calcite unit cell, which is larger for Sr-bearing calcite in Göktepe marbles [9].

The potential of geochemical analyses as a tool for marble provenance was already devised in the pioneering work of the German geologist Lepsius [10]. However, trace elements and chemical analyses, in general, are only rarely used to infer marble provenance. This is due to the existing inherent and unpredictable inhomogeneities in the quarry sites along with the difficulty of comparing analyses taken using different sample preparations, techniques or even different setups of a given technique. Only in recent years, and in some measure due to advances in the analytical methods [11], was the trace element composition of the marbles again tested to contribute to provenance studies of marbles, often as part of multi-method strategies [12–14]. Nevertheless, discrepancies in results from different techniques and setups continue to be an issue. Moreover, to deal with large compositional datasets, the common approach echoes the usual practices applied to geochemical analyses of archaeological pottery. Therefore, common descriptive statistical tools are used, such as boxplots to display the variation of single variables, biplots of chemical elements, principal component analysis (PCA) and the biplots of the corresponding main components [11,15].

Despite the widespread use of PCA, this unsupervised multivariate data analysis (MDA) method is not strictly a classification method [16]. PCA is a great tool to reduce the dimensionality of a dataset and to visualize the differences between the objects of the dataset, but it is not designed to discriminate different classes within the dataset. Furthermore, there is no agreement on the best suitable approach to transform and standardize the dataset before applying the MDA [17]. In contrast with unsupervised methods, supervised machine learning methods (using data previously labeled with their corresponding class) can be trained to classify unlabeled data into a class. However, predictive modeling using supervised methods is still largely an underexploited field within archaeology [18,19]. The supervised classification approach has been successfully applied to classify archaeological soils [20], obsidians [21], clays [22] and pottery [23]. Recently, the approach has been used to distinguish groups made of both clays and baked clays from similar modern production centers [24], and the same approach has been applied to infer the provenance of archaeological pottery using a freeware R code [25]. In this paper, the supervised approach is applied for the first time, to the best of our knowledge, to determine the provenance of archaeological marble samples.

A case study consisting of three sets of local marble samples from three archaeological sites in Catalonia (NE Spain) illustrates the potential of the supervised approach in solving a specific provenance dilemma. The archaeological samples have been characterized following the common multi-technique approach, and all of them appear to be compatible with two Catalan marble districts in Gualba (Barcelona province, Spain) and Ceret (Roussillon province, France); unsupervised MDA methods are also generally unable to discriminate between the two possible origins and only by using carefully transformed data will some hints of effective discrimination appear. In contrast, several supervised classification models have been successfully applied to produce probability values belonging to either Gualba or Ceret for the three sets of unlabeled samples using the "Supervised Provenance Analysis" code from [25], and all of them clearly assign the samples to the Gualba quarry district.

## 2. Materials and Methods

### 2.1. Archaeological Samples

Three sets of marble samples were collected within three archaeological sites in Catalonia (Figure 1), corresponding to different chronologies. The samples were carefully retrieved, when possible, from fracture surfaces, in the form of tiny chips using a small hammer and a chisel, avoiding weathered material and patinas and minimizing the damage on polished or worked parts.
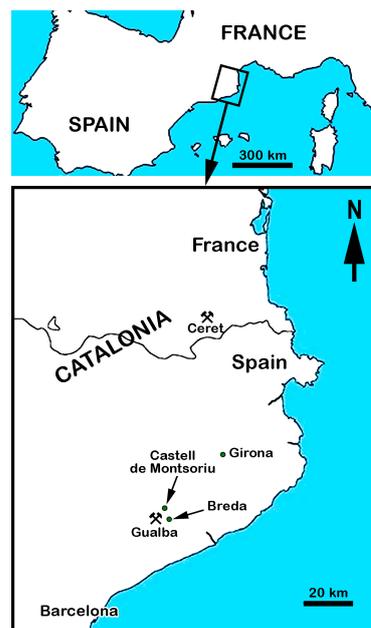
**Figure 1.** Location of the archaeological sites (green dots) where the archaeological marbles were retrieved, and that of the two marble quarry districts (Gualba and Ceret), indicated with the mining symbol and considered here as probable provenances of the archaeological samples.

### 2.1.1. The City Walls of Gerunda (Girona)

Gerunda, presently Girona, was an oppidum at the confluence of the Ter and Onyar rivers. It is located ~85 km northeast of Barcelona. The city was built by the Romans in the early 1st century CE over a pre-roman settlement. The original Republican city walls were rebuilt during the Roman Empire period (3rd century CE) following the same limits and including quadrangular towers [26]. In the Middle Ages (early 9th century), the city walls were rebuilt and reinforced, again within the same limits, and new frustoconical towers were erected. In the 14th–15th centuries, new city walls were built, enclosing a larger surface; these were reinforced in the 17th century. Finally, from the 19th century onwards, several segments of the different walls were demolished, allowing the city to expand.

The Roman foundational city walls were mainly made of a local nummulitic limestone using the opus siliceum technique; in contrast, for the Republican walls, a local sandstone was used, applying the opus quadratum technique [27]. In 1988, excavations in one of the eastern Medieval towers (known as Torre del Telègraf Òptic) revealed the presence of 15 marble ashlars bearing architectonic decorations along with other common sandstone ashlars clearly reused from Roman buildings [28]. The marble blocks have been interpreted to be part of an honorific arch-gate [29]. All the marble ashlars share similar macroscopic features, and some of them were sampled in a previous study; a possible provenance from either Ceret or Gualba had already been hypothesized by macroscopic inspection [30]. Four samples (labeled Gi-1 to Gi-4) obtained from the previous sampling were made available to us for analysis.

### 2.1.2. The Cloister of Saint Saviour's Monastery (Breda)

Breda is a small rural town situated ~50 km northeast of Barcelona that developed around a Benedictine monastery that was built in the 11th century. The monastery construction was promoted by the ruling viscounts of Girona (or Cabrera for the family name of the lineage). In the 14th–16th centuries, a new Gothic church was added to the ensemble. From the 19th century onwards, because of lootings linked to the Peninsular War, the monastery entered a decline period and ceased its religious activities because of a disentailment by the Spanish government.

In 2018, excavations works were undertaken in the cloister, in an area formerly occupied by modern buildings that were demolished in 2014. In the southeastern corner of the cloister, the works revealed a modern pavement made of marble slabs that were dismantled. However, the marble slabs have been interpreted to be reused materials from the old cloister galleries [31]. Six of these marble slabs were sampled (samples were labeled Br-1 to Br-6). Considering the proximity of the marble outcrops in Gualba (only ~5 km west of Breda), it makes sense to hypothesize that the slabs come from these outcrops, but no archaeometric study has been conducted so far to verify this possibility.

### 2.1.3. The Inner Bailey of Montsoriu Castle

The Montsoriu Castle is an important Gothic castle (10th–14th centuries) that is located on top of the homonymous hill, situated ~50 km northeast of Barcelona, between the municipalities of Breda and Arbúcies. In the 11th–12th centuries, the castle became the residence of the viscounts of Girona. In the 15th century, it was no longer a stately home and, in the next two centuries, was only occasionally used for military purposes and then abandoned indefinitely. By the end of the 20th century, the castle was a ruin and is only now undergoing the process of restoration.

During the 13th–14th centuries, the viscounts of Girona made some modifications to the castle, including the construction of an inner bailey [32]. The pavement of this upper bailey was made with big marble slabs. The pavement was dismantled in 1997 as part of excavation works. Their slabs are still presently stored temporarily in the remains of the castle's chapel, awaiting the restoration of an underlying cistern and the pavement itself. Ten marble slabs were sampled, covering the different grain sizes and macroscopic variations shown in the stored material (samples were labeled Mo-1 to Mo-10). Again, taking into account the proximity of the marble outcrops in Gualba (only ~5 km southwest of the castle), it makes sense, and it is generally accepted that the slabs and other marble elements of the castle were cut in Gualba marble [33], but, to date, this has not been archaeometrically proven.

### 2.2. Quarry Samples

Taking into account the archaeological hypotheses, two Catalan marble quarry districts have been considered as possible provenance of the archaeological marbles. These are located in Gualba (Barcelona province, Spain) and Ceret (Roussillon province, France) (Figure 1).

### 2.2.1. Gualba Marbles

The Gualba marbles outcrop on both sides of the valley carved by the Gualba stream, around 2–3 km upstream of a small town also named Gualba, situated ~50 km north-east of Barcelona, only ~5 km west of Breda. Currently, there is active quarrying only on the right side of the valley, basically underground mining. However, there are several abandoned opencast quarries on the left side. The marbles are part of Cambrian–Ordovician sediments affected by the Hercynian orogeny. The central part of the stratigraphic sequence is formed by micaschists with some interlayered quartzite beds and with marbles and gneiss lenses in the lower part [34], and the marbles can reach a maximum thickness of 20 m. As a consequence of late-Hercynian intrusions, contact metamorphism also took place, and in areas directly in contact with the magma (presently granitoid-type rock), the marbles often developed a skarn-type mineralogy. Apart from the areas affected by metasomatism, the marbles appear relatively pure (either calcitic or dolomitic), usually coarse-grained, white or gray/green veined, slightly foliated and occasionally affected by mylonitization processes.

It is well documented that the Gualba marbles have been quarried from the early 20th century onwards. However, it is also generally accepted that some elements of the nearby local monumental heritage are made of Gualba marble, and therefore, the marble extraction would have certainly started earlier. Examples of presumed uses of this marble are the portal of the baroque Saint Martin's Church in Sant Celoni, the ashlars corners of

El Ribot (a 16th-century lodge) in la Batllòria or a staircase and the sampled pavement in the Montsoriu Castle—all these examples lie within a 5 km radius from Gualba. However, the adscription to Gualba marble is based mainly on the proximity criterion and macroscopic similarity. To our knowledge, no previous archaeometric studies exist for this marble.

Nine different outcrops (labeled from G1 to G9) were selected for sampling (Figure 2), comprising seven sampling points within the large and active Xauxa quarry, including two underground sampling spots (G6 and G7). Two abandoned quarries were also sampled (G8 and G9). Several samples were retrieved from every sampling spot to apply different characterization techniques. For instance, 18 samples were prepared as petrographic thin sections, and a total of 26 were subjected to chemical analyses. The samples have been labeled with a unique cardinal number followed by the corresponding sampling spot.
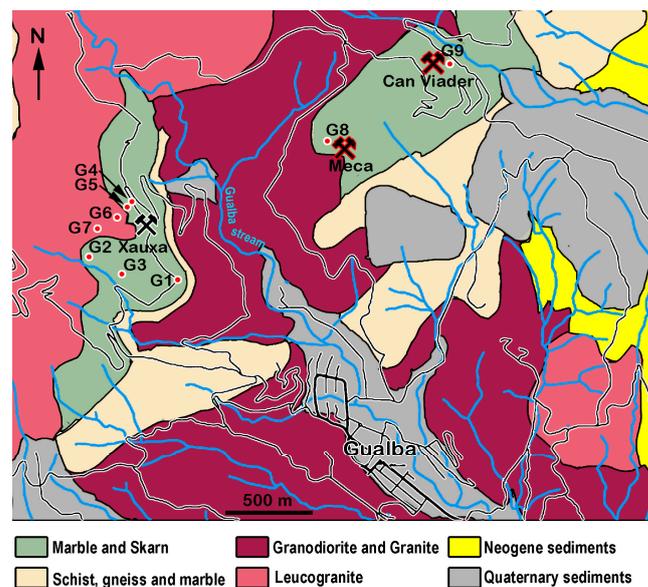


**Figure 2.** Geological map of the Gualba marble district with indication of the active mines (Xauxa) and the abandoned quarries (Can Viader and Meca) as well as the labeled sampling points (red dots).

### 2.2.2. Ceret Marbles

The so-called Ceret marbles outcrop 2–3 km south of the town of Ceret, a French town ~130 km northeast of Barcelona and ~20 km southwest of Perpignan. The marbles are part of the metasediments interlayered between the gneisses of the Roc de Frausa massif (Eastern Pyrenees). The area has been affected both by the Hercynian and Alpine orogenies. The metasediments are mainly metapelites to metagreywackes, but marble layers can also occasionally appear with thicknesses from various centimeters to about 30 m [35]. The marbles appear either white or gray-veined. The metasedimentary sequence is attributed to the upper Proterozoic to lower Cambrian ages [36]. Several small and abandoned marble quarries are known along the Vallera Valley [37]. The quality of the marble is rather irregular, sometimes shows visible gray and white bands, even including schist inclusions while other times appear homogeneously white and fine-grained [38]. The Ceret marble could have been quarried during Antiquity [39,40], although some authors [41] have argued that the use of Ceret marble would have only started in the 12th century. It seems certain their use during the Middle Ages in various Northern Catalan Romanesque artworks [42,43], although the identification is basically based on macroscopic inspection. Moreover, the distribution would be rather limited; some elements in the cloister of Fontfroide [44] (~75 km north of Ceret) and the portal of Santa Maria Basilica in Castelló d'Empúries [38] (~38 km southeast of Ceret) are the most distant hypothesized uses.

In fact, this marble had only been archaeometrically characterized to a limited extent, and therefore, the actual use of it is not well-known. A few isotopic measurements have been published so far [45,46], showing that this marble exhibits particularly negative $\delta^{18}O$

values. Its cathodoluminescence is also known to be distinctly intense and homogeneously orange [44], and it has even been used as a reference material to normalize cathodoluminescence measurements [47]. Six different outcrops (labeled from C1 to C6) were selected for sampling (Figure 3), comprising three abandoned quarries and three sites where the marble formation is cut by a hiking track. Several samples were retrieved from every outcrop to apply different characterization techniques. For example, 9 were prepared as petrographic thin sections, and 28 were subjected to chemical analyses. The samples have been labeled with a unique cardinal number followed by the corresponding sampling spot.
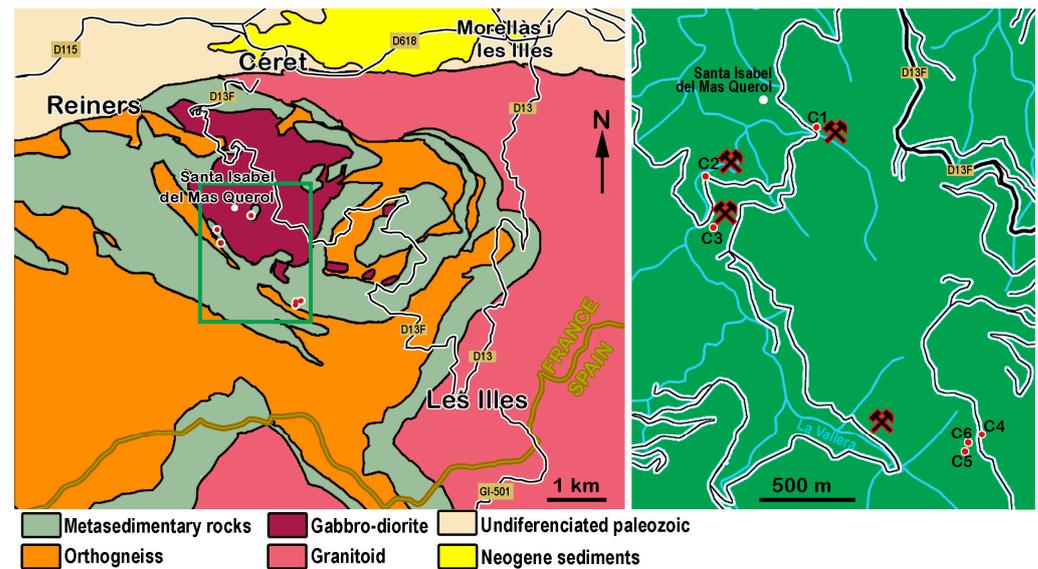


**Figure 3.** Geological map of the Ceret marble district with indication of the sampling points (red dots). The highlighted green square has been zoomed on the left and reproduced as a geographical map with indication of the abandoned quarries and the sampling points, now labeled.

### 2.3. Methods

#### 2.3.1. Experimental Methods

The samples prepared as thin sections were described petrographically using a polarizing optical microscope (Nikon Eclipse E600, Tokyo, Japan) at the Geology department of Universitat Autònoma de Barcelona, using transmitted light (TL) in both the plane-polarized light (PPL) and the cross-polarized light (XPL) modes. Petrographic images were retrieved using an attached camera (Nikon DS-Fi3). Selected thin sections were half stained with alizarin red S dye to detect the presence of $CaCO_3$. Cathodoluminescence (CL) measurements were also performed on the thin sections using a CL8200 Mk5-1 (Cambridge Image Technology Ltd., Welwyn Garden City, UK), available at the Earth Science Faculty, Universitat de Barcelona, operating at 6–7 kV with a gun current of ~300 µA. The CL images were recorded using 1 s exposure time.

Small amounts of powdered samples were used to perform the oxygen and carbon isotope analyses. About 0.08 mg of powder was reacted at 70 °C with 100% phosphoric acid in a Kiel Device III, coupled online with a mass spectrometer MAT-252 (both devices from Thermo Finnigan LLC, San José, CA, USA). The measurements were performed in the Centres Científics i Tecnològics (Universitat de Barcelona). The results were calibrated with the NBS-18 and NBS-19 international reference materials. Results of carbon and oxygen isotopic composition are reported in ‰ deviation relative to Vienna Pee Dee Belemnite (V-PDB). The reproducibility, determined by replicate analysis of standards, was better than ±0.02‰ for carbon and ±0.06‰ for oxygen.

Additional amounts (~5 g) of powdered sample were ground using a laboratory mill (Pulverissette™, Fritsch GmbH, Idar-Oberstein, Germany) to pass a 125 µm mesh. The powders were then prepared in the form of pressed powder pellets using a methyl

methacrylate resin as a binding agent (Elvacite™ commercial resin) under a pressure of 10 T. The chemical composition of the pellets was measured by Energy Dispersive X-ray Fluorescence (EDXRF) in an S2 Ranger system (Bruker/AXS, GmbH, Karlsruhe, Germany). The raw data were fitted using the SPECTRA.EDX package (Bruker AXS, GmbH, Karlsruhe, Germany). Quantification was made by the assisted fundamental parameters method. Analyses were made in a vacuum atmosphere for better detection of low Z elements and using different conditions of voltage to properly excite low, medium and high atomic number elements existing in the samples; the measuring time was set at 400 s. In all the analyzed marbles, Ca prevailed along with minor quantities of other elements (Si, Al, Mg, Mn, Fe, Zn, Sr and Y) that occasionally were present below the corresponding detection limits.

2.3.2. Statistical Methods

Several statistical methods have been applied to the geochemical values obtained from the analyzed samples. Their chemical composition constituted a set of features associated with a given object (i.e., the sample). Following the approach presented in previous works [24,25], these chemical data were used in two different stages: unsupervised and supervised.

In the first stage, the objects were analyzed, disregarding their provenance using two unsupervised methods: principal component analysis (PCA) and hierarchical clustering analysis (HCA). These methods were tested both with non-transformed chemical data and transformed and/or standardized data. Standardization to zero mean and unit variance was preferred. As Ca (or expressed as carbonate) is obviously a highly correlated variable to almost all the other geochemical variables, tests were performed, removing it before PCA and HCA computation to avoid a redundant variable. Occasional non-detected (n.d.) values were considered zeros. Different distance and clustering methods were tested to produce the HCA dendrograms. Euclidian distance and Ward's linkage methods were preferred. The unsupervised methods were used to explore the internal variability of the samples from every quarry and to assess the similarities between the marbles from the two quarries and the archaeological sites.

In the second step, a number of supervised methods were used following a train–test split procedure to estimate their performance to classify and distinguish samples from the two marble districts. Again, occasional non-detected (n.d.) values were considered zeros, and a correlation matrix was computed to display correlation coefficients for the different variables. Negative correlation of $CaCO_3$ values with practically all the other chemical variables was confirmed, and therefore, this feature was removed from the input data to deal with classification models. Apart from this, no other data transformation was required due to the adaptability of the models during their training. Namely, the used supervised classification models were the following: weighted k-nearest neighbors (kkNN), random forest (RF), artificial neural network (ANN), linear discriminant analysis (LDA) and generalized linear models (GLM). Additionally, a stacking classifier was also used following a random forest approach. The stacking classifier uses the output from the other mentioned methods as an input to a meta-classifier designed to further improve the class-prediction performance. Different splits were used to check the robustness of the models.

Once trained, the supervised methods were used to assign classes to the unlabeled samples (i.e., the archaeological marbles). Probability values belonging to either Gualba or Ceret were computed for each of the three sets of archaeological marbles combining the results obtained using different splits and different samples. Both the test-train step using labeled samples and the final step of class prediction for unlabeled samples was performed running the "Supervised Provenance Analysis" code from [25]. This code is freely downloadable as a project folder containing R Markdown files that integrate text, code and, after running, also the results. The code uses different R packages for every model: class (for kkNN), randomForest (for RF and the stacking classifier), nnet (for ANN), mass

(for LDA) and glmnet (for GLM). The packages are units of reproducible code containing reusable functions. Free documentation is available to describe the use and fundamentals of the package contents [48].

## 3. Results and Discussion

### 3.1. Petrography

The petrographic characterization under thin-section microscope observation reveals that the analyzed quarry samples from Gualba and Ceret are not homogenous (Figure 4). Furthermore, many of their common features overlap to a great extent and also match with those observed in the archaeological samples.
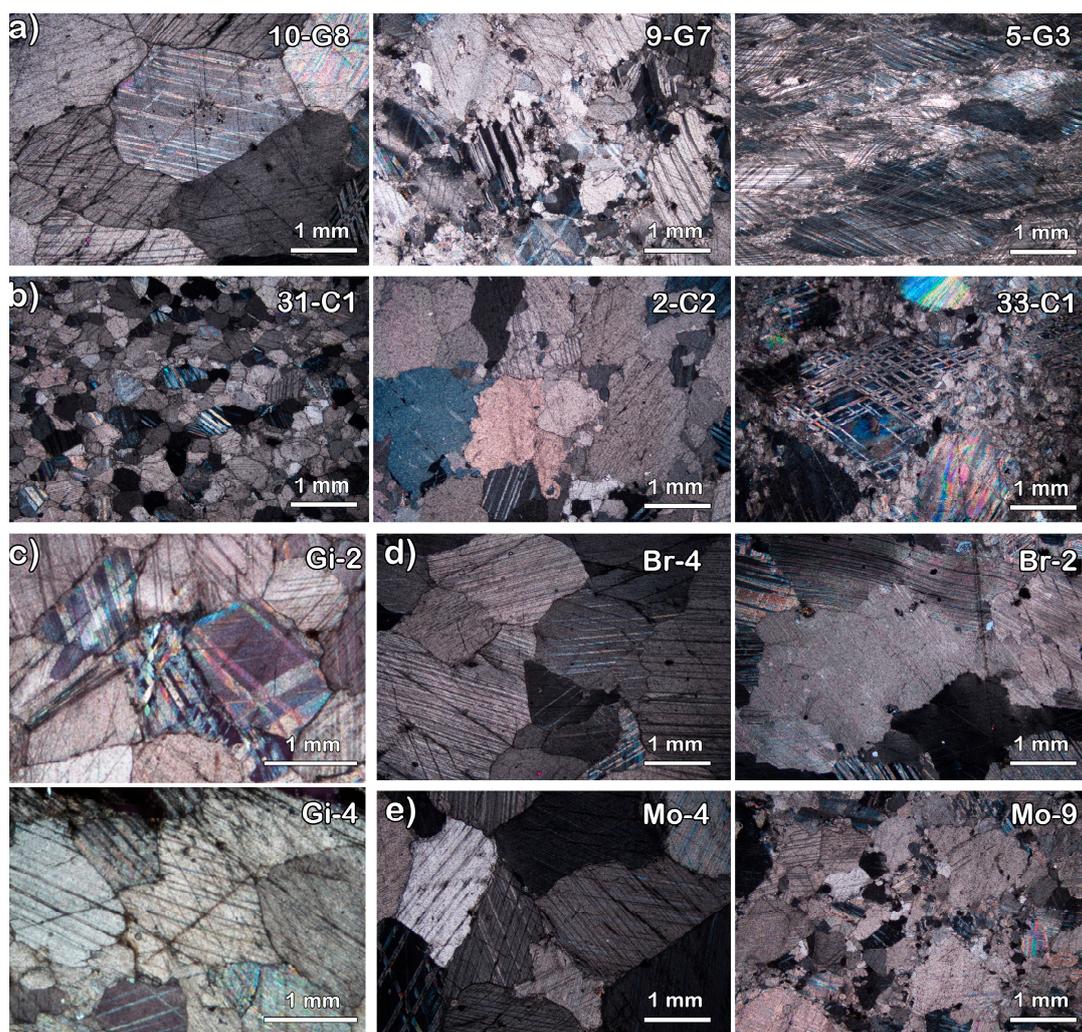


**Figure 4.** Petrographic images of thin sections of some of the analyzed marbles viewed under XPL. (**a**) Three examples of textures found in samples from the Gualba quarries ranging from polygonal to mylonitic. (**b**) Three examples of textures exhibited by samples from the Ceret quarries, including granoblastic (polygonal and interlobate) and mortar. (**c**) Examples (top and down) of granoblastic coarse-grained textures of the archaeological samples from Girona. (**d**) Examples (left and right) of granoblastic coarse-grained textures of the archaeological samples from Breda with variable amounts of plastic deformation. (**e**) An example (left) of the common granoblastic coarse-grained texture of the archaeological samples from Montsoriu and the only one (right) exhibiting a mortar texture.

### 3.1.1. Quarry Marbles

The sampled marbles from Gualba show the widest range of petrographic features. Most of the samples consist of very coarse-grained marbles with maximum grain size

(MGS) values often around 5 mm, but in some cases, MGS is only 0.3 mm. The marbles are basically calcitic with slightly to highly anisotropic fabrics; occasionally, inclusions of accessory minerals appear, including olivine, epidote and serpentine. The microstructures are granoblastic, generally heteroblastic. The amount of plastic deformation can vary from almost nonexistent (Figure 4a, left), with calcite grains that often display twin sets both thin and thick (types I and II in [49]), to highly deformed (Figure 4a, right), with profusion of bent twins (type III in [49]) and undulose extinction. The contacts between the grains usually look curved but occasionally can appear embayed or straight. In many samples, the boundary between large grains is formed by smaller crystals defining mortar-like microstructures (Figure 4a, middle).

The sampled marbles from Ceret are generally also calcitic and coarse-grained. However, the occurrence of fine-grained (Figure 4b, left) marbles (MGS < 2 mm) is higher compared with Gualba marbles; the main accessory mineral inclusions consist of mica. Their fabric varies from isotropic to moderately anisotropic. Their most common microstructure is granoblastic with a homogeneous texture, but it can vary from polygonal (straight grain boundaries and frequent triple junctions) to interlobate (highly irregular grain boundaries, Figure 4b, middle). Type I and II mechanical twins are also common. Micrograins defining mortar-like textures (Figure 4b, right) and plastic deformation features are less frequent compared with the Gualba samples, but they also appear occasionally. It happens the same with plastic deformation features; slightly bent twins and undulose extinction appear only sporadically.

### 3.1.2. Archaeological Marbles

The four archaeological samples retrieved from the city walls of Gerunda (Girona) share common petrographic features. They are calcitic coarse-grained marbles (MGS ~3.5 mm) with isotropic (or slightly anisotropic) fabric (Figure 4c). The calcite grains display a seriate size distribution with curved grain boundaries that occasionally appear straight. The grains often show thin or thick twins (types I and II).

The six samples retrieved from the Cloister of Saint Saviour's monastery in Breda are not so homogeneous; all of them are very coarse-grained (MGS 4–5 mm) calcitic marbles markedly heterogranular. Three samples show hints of plastic deformation (bent twins and undulose extinction) and predominant sutured grain boundaries (Figure 4d, right). In contrast, three other samples exhibit a granoblastic microstructure with a homogeneous polygonal texture with mainly straight grain boundaries (Figure 4d, left) and occasional type I and II twins.

The marbles from the inner bailey of the Montsoriu Castle are also calcitic. Nine out of the ten sampled marbles exhibit coarse-grained polygonal textures (MGS 3–5 mm) with curved or straight grain boundaries (Figure 4e, left). Only one sample deviates from this description and presents micrograins that concentrate in the grain boundaries defining a mortar-like texture (Figure 4e, right). Type I and II twins are common in all ten samples.

### 3.2. Cathodoluminescence

All the measured calcitic samples (both quarry and archaeological samples) display particularly intense orange cathodoluminescence (Figure 5). The vast majority of samples exhibit a very homogeneous response with slightly higher luminescence concentrated in exfoliation lines and grain boundaries.

Most of the accessory minerals appear with no or lower luminescence compared with the calcite crystals, but these only occasionally constitute an important fraction within the sample (e.g., Gualba sample 8-G6 in Figure 5a). A few samples from Ceret exhibit a patchy luminescence with small crystals of lower luminescence embedded in a mosaic of highly luminescent crystals (e.g., sample 11-C4), but most of the Ceret samples show a homogeneously intense orange luminescence similar to the Gualba samples as well as the archaeological samples from Girona, Breda and Montsoriu.
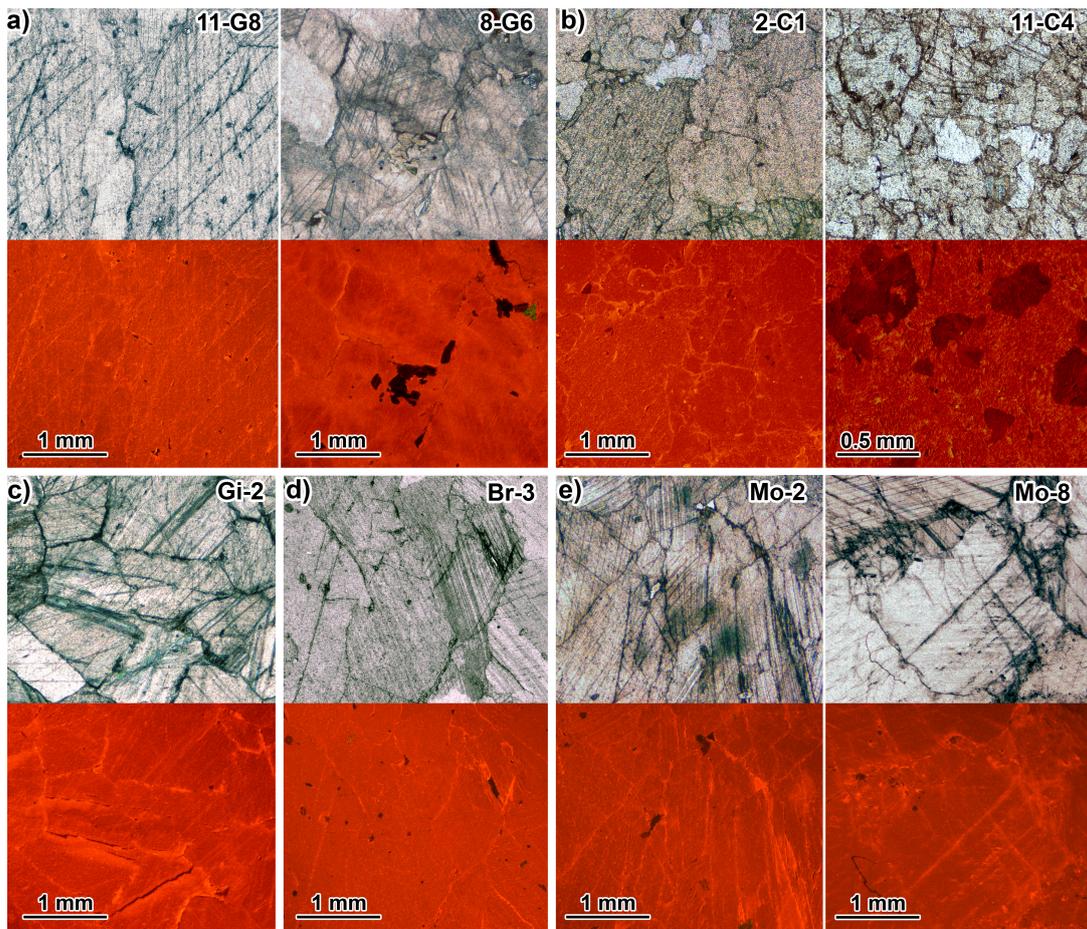
**Figure 5.** Petrographic images of thin sections of some of the analyzed marbles, including the same areas viewed under PPL (top) and CL (bottom), all of them exhibiting intense orange cathodoluminescence (exposure time 1 s). (**a**) Two samples from the Gualba quarries. (**b**) Two samples from the Ceret quarries. (**c**) Archaeological sample Gi-2 from Girona. (**d**) Archaeological sample Br-3 from Breda. (**e**) Archaeological samples Mo-2 and Mo-8 from Montsoriu.

*3.3. Stable Isotopes*

3.3.1. Quarry Marbles

The main factors controlling the isotopic ratios in marbles relate to the nature of the protolith and the temperature and interaction with water during metamorphism and later weathering. Oxygen isotope ratios of carbonate rocks are especially sensitive to hydrothermal fluids; during metamorphism, oxygen (but also carbon) isotope ratios tend to decrease compared with the unmetamorphosed limestone [50]. To our knowledge, no previous isotope ratios have been published for Gualba marbles and only a few for Ceret marbles. Both marbles share similar isotope ratios, although distinctly different from classical marbles.

16 out of the 18 measured isotope ratios on Gualba marbles group defining a compact cloud of points. The other two appear as outliers with lower $\delta^{13}$C and $\delta^{18}$O values (Figure 6). The outliers correspond to samples from the G2 sampling site, a quarry front characterized by the presence of impure, gray/green veined marbles. Their particularly low $\delta^{13}$C ratios could indicate that these samples come from silica-bearing (or calc-silicate) marbles [51]. The rest of Gualba isotopic ratios correspond to white or slightly gray-veined marbles; their values group in an area with particularly negative $\delta^{18}$O ratios with values ranging from −21.23 to −14.60‰ and less scattered $\delta^{13}$C values ranging from −1.91 to 0.82‰.
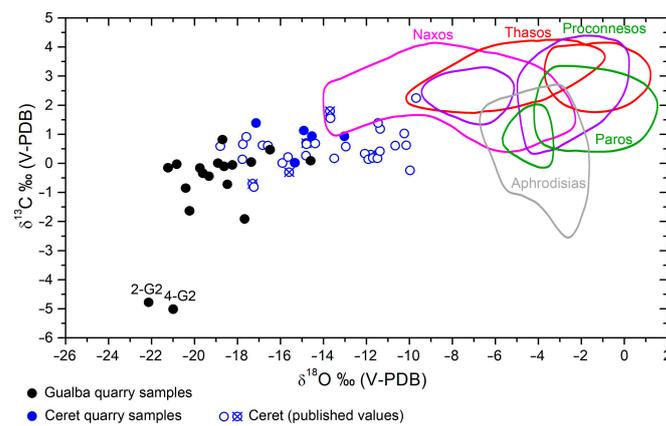
**Figure 6.** Isotopic results from the analyzed quarry samples and additional published values from Ceret [46,52] plot with the isotopic fields for medium/coarse-grained classical marbles from [53].

Only five samples from the different sampled sites at Ceret were subjected to isotopic measurements. However, the corresponding isotope ratios were joined to the already published values. These are, on the one hand, 27 isotope ratios published in [46], all but one are also published in [45], and on the other hand, 5 more ratios from architectural marbles from the nearby Perpignan [52] that were assigned with confidence to the Ceret quarries. The 37 values group together along a range of negative $\delta^{18}O$ ratios ($-18.80$ to $-9.69‰$) and again with rather constant $\delta^{13}C$ values (ranging from $-0.81$ to $2.25‰$).

Both sets of isotope ratios for Gualba and Ceret marbles are distinctly different from the known reference isotope diagrams [53] of classical coarse-grained marbles (Figure 6); in particular, they exhibit lower $\delta^{18}O$ ratios. Only three previously reported ratios for Ceret marbles lie within a classical-marble reference diagram (Naxian marble). Using the sets of experimental isotopic ratios, two different approaches have been used to outline the limits of the reference diagrams for Gualba and Ceret marbles. On the one hand, a tight envelope has been computed using a concave hull algorithm [54] based on a k-nearest neighbors approach (Figure 7a), and a less compact envelope has been defined by computing the 95% confidence ellipses for each set (Figure 7b). The position of the newly defined reference envelopes of Gualba and Ceret are relatively similar, the tight envelopes have an overlap percentage of about 17% and the elliptic envelopes have a roughly 27% overlap. Taking into account the limited number of available data, the looser envelopes should be preferred as reference diagrams for the marbles quarried at Gualba and Ceret.
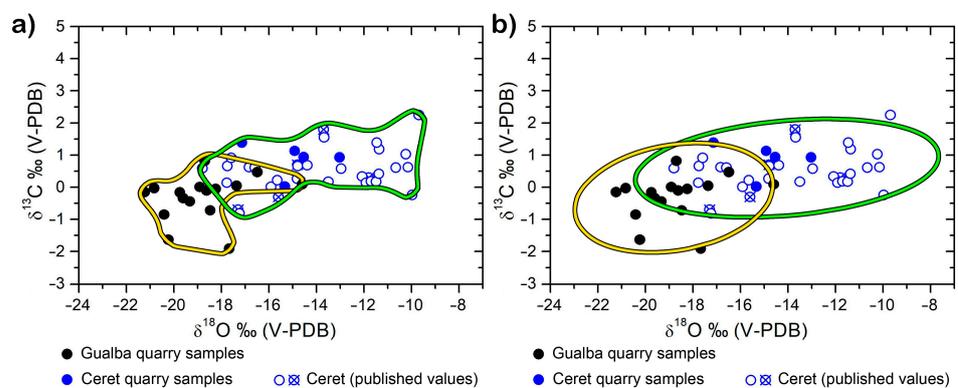


**Figure 7.** Isotopic results from the analyzed quarry samples along with additional published values from Ceret [46,52], including envelopes for Gualba (yellow) and Ceret (green), computed using (**a**) a concave hull algorithm or (**b**) 95% confidence ellipses.

### 3.3.2. Archaeological Marbles

Despite the definition of isotope reference ellipses, the isotope ratios measured on the archaeological samples do not allow for discrimination between the two presumed provenances (Figure 8), as most of the measured isotope ratios lie within the area of intersection between the reference ellipses computed for the two marble districts.
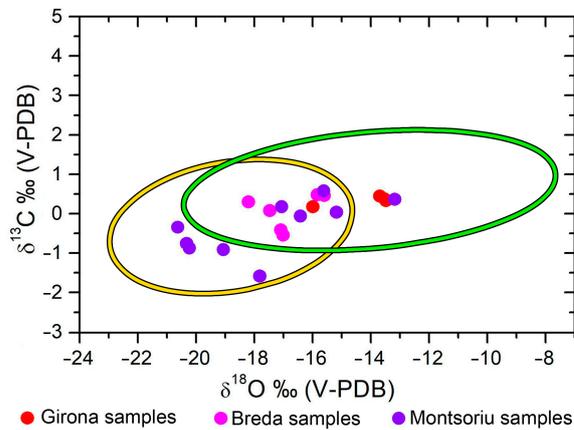


**Figure 8.** Isotopic results from the analyzed archaeological samples, including envelopes for Gualba (yellow) and Ceret (green).

Three out of the four measured isotope ratios for samples retrieved from the marble ashlars in the city walls of Gerunda have almost identical isotope ratios that lie near the center of the Ceret reference diagram and outside the Gualba ellipse. This would suggest that Ceret is the source of the ashlars. However, another measured ratio (corresponding to sample Gi-3) lies within the area of intersection between the reference ellipses. The whole six isotope ratios measured on marble slabs of the Sant Saviour's cloister (Breda) also lie within the area on the intersection, and therefore, both marble districts would be equally probable sources. Finally, the ten isotope ratios measured on the slabs that constituted the inner-bailey pavement of Montsoriu Castle scatter along the areas of both reference ellipses, five ratios lie exclusively within the Gualba ellipse, four other ratios lie within the area of intersection, and only one appears near the center of the Ceret reference diagram.

Only a limited number of measurements were performed, but it seems reasonable to assume that with an increased number of isotopic measurements, the isotope ratios of the three archaeological sites would spread indistinctly within the two reference ellipses, in fact, as it already happens with the Montsoriu marbles. The use of tighter reference envelopes would have resulted in similar conclusions.

### 3.4. Energy Dispersive X-ray Fluorescence (EDXRF) and Unsupervised Multivariate Data Analysis (MDA)

Tables 1 and 2 display the full set of chemical analyses obtained by EDXRF measurements for samples from the Gualba and Ceret quarry districts, respectively. With only a few exceptions, most of the samples can be classified as pure calcite marbles [55].

Among the 26 analyzed samples from Gualba, one of them (4-G2) has a very high Si content, and this is not actually a marble but a carbonate-silicate rock, and another one (3-G2) would be a pure dolomite marble [55] with about 30 wt. % of $MgCO_3$ (over the total carbonate content). Both samples are from outcrop G2, a quarry front particularly rich in impure gray/green veined marbles. Regarding the 28 analyzed samples from the Ceret outcrops, four of them resulted in being pure dolomite marbles (4-C2, 7-C3, 10-C4, 28-C5), also with around 30 wt. % of $MgCO_3$. In this case, the dolomitic samples do not concentrate in any particular outcrop. The higher Mg contents of the dolomitic samples (both from Gualba and Ceret) correlate with higher content in Fe and Mn.

**Table 1.** Chemical composition of the analyzed quarry samples from Gualba.

| | CaCO$_3$ (%) | MgO (%) | SiO$_2$ (%) | Al$_2$O$_3$ (%) | Mn (ppm) | Fe (ppm) | Sr (ppm) | Zn (ppm) | Y (ppm) |
|---|---|---|---|---|---|---|---|---|---|
| 1-G1 | 93.4 | 1.53 | 3.16 | 0.34 | 1365 | 4791 | 183 | 26 | 14 |
| 2-G2 | 93.4 | 1.57 | 3.13 | 0.30 | 1310 | 4966 | 186 | 25 | 14 |
| 3-G2 | 81.0 | 16.70 | 0.86 | 0.14 | 618 | 3987 | 92 | 30 | 2 |
| 4-G2 | 32.0 | 17.80 | 44.90 | 3.08 | 1492 | 6050 | 29 | 1719 | 1 |
| 5-G3 | 97.9 | 0.10 | 1.32 | 0.38 | 375 | 906 | 144 | 14 | 11 |
| 6-G4 | 99.0 | n.d. | 0.44 | 0.31 | 402 | 601 | 121 | 12 | 9 |
| 7-G5 | 95.5 | n.d. | 3.81 | 0.38 | 644 | 734 | 150 | 23 | 18 |
| 8-G6 | 95.7 | 1.45 | 1.86 | 0.54 | 473 | 1434 | 151 | 11 | 9 |
| 9-G7 | 97.6 | 0.59 | 1.02 | 0.38 | 400 | 1259 | 108 | 8 | 0 |
| 10-G8 | 98.8 | 0.02 | 0.64 | 0.35 | 339 | 598 | 110 | 17 | 9 |
| 11-G8 | 98.6 | n.d. | 0.75 | 0.31 | 448 | 801 | 118 | 21 | 9 |
| 12-G8 | 98.4 | n.d. | 0.79 | 0.35 | 564 | 1206 | 138 | 13 | 11 |
| 13-G9 | 99.2 | 0.04 | 0.38 | 0.24 | 277 | 395 | 101 | 9 | 2 |
| 14-G9 | 98.1 | 0.08 | 0.89 | 0.50 | 691 | 1140 | 173 | 14 | 10 |
| 15-G8 | 98.6 | 0.01 | 0.74 | 0.31 | 497 | 1000 | 130 | 10 | 9 |
| 16-G8 | 98.7 | 0.05 | 0.68 | 0.31 | 401 | 787 | 118 | 19 | 8 |
| 17-G5 | 98.7 | 0.02 | 0.70 | 0.29 | 418 | 608 | 129 | 7 | 9 |
| 18-G5 | 96.9 | 1.25 | 1.28 | 0.31 | 406 | 790 | 113 | 16 | 3 |
| 19-G5 | 99.0 | n.d. | 0.58 | 0.23 | 383 | 416 | 104 | 10 | 9 |
| 20-G9 | 98.6 | 0.15 | 0.69 | 0.37 | 233 | 430 | 116 | 5 | 3 |
| 21-G6 | 98.9 | 0.04 | 0.51 | 0.22 | 342 | 741 | 110 | 9 | 9 |
| 22-G6 | 99.0 | 0.07 | 0.49 | 0.22 | 347 | 706 | 112 | 9 | 10 |
| 23-G3 | 98.9 | 0.07 | 0.57 | 0.22 | 310 | 472 | 139 | 10 | 10 |
| 24-G6 | 99.2 | 0.03 | 0.38 | 0.16 | 328 | 591 | 113 | 6 | 4 |
| 25-G7 | 97.3 | 0.67 | 1.46 | 0.22 | 457 | 1039 | 96 | 108 | 4 |
| 26-G5 | 98.5 | n.d. | 0.96 | 0.31 | 287 | 465 | 139 | 15 | 10 |

**Table 2.** Chemical composition of the analyzed quarry samples from Ceret.

| | CaCO$_3$ (%) | MgO (%) | SiO$_2$ (%) | Al$_2$O$_3$ (%) | Mn (ppm) | Fe (ppm) | Sr (ppm) | Zn (ppm) | Y (ppm) |
|---|---|---|---|---|---|---|---|---|---|
| 1-C1 | 98.9 | 0.37 | 0.46 | 0.20 | 148 | 138 | 109 | 4 | 2 |
| 2-C1 | 98.5 | 0.57 | 0.52 | 0.22 | 275 | 546 | 156 | 9 | 8 |
| 3-C2 | 98.5 | 0.48 | 0.66 | 0.23 | 186 | 455 | 95 | 4 | 5 |
| 4-C2 | 79.6 | 16.80 | 1.38 | 0.24 | 1350 | 5700 | 96 | 1012 | n.d. |
| 5-C2 | 99.0 | 0.36 | 0.33 | 0.18 | 166 | 271 | 151 | 24 | 4 |
| 6-C3 | 98.7 | 0.54 | 0.33 | 0.19 | 260 | 525 | 131 | 10 | 9 |
| 7-C3 | 81.6 | 16.00 | 0.46 | 0.22 | 1476 | 5141 | 134 | 23 | n.d. |
| 8-C3 | 98.9 | 0.28 | 0.46 | 0.22 | 138 | 256 | 130 | 6 | 3 |
| 9-C4 | 94.9 | 3.50 | 0.97 | 0.16 | 530 | 1350 | 151 | 45 | n.d. |

**Table 2.** *Cont.*

| | CaCO₃ (%) | MgO (%) | SiO₂ (%) | Al₂O₃ (%) | Mn (ppm) | Fe (ppm) | Sr (ppm) | Zn (ppm) | Y (ppm) |
|---|---|---|---|---|---|---|---|---|---|
| 10-C4 | 74.9 | 16.20 | 6.63 | 0.34 | 1539 | 6050 | 83 | 47 | n.d. |
| 11-C4 | 99.0 | 0.21 | 0.51 | 0.19 | 144 | 106 | 132 | 10 | n.d. |
| 12-C5 | 99.3 | n.d. | 0.36 | 0.13 | 209 | 296 | 458 | 10 | 7 |
| 13-C6 | 98.6 | 0.05 | 1.01 | 0.24 | 185 | 281 | 126 | 11 | 8 |
| 14-C2 | 91.7 | 0.65 | 6.57 | 0.79 | 210 | 678 | 129 | 10 | 6 |
| 15-C6 | 98.3 | 0.21 | 1.13 | 0.27 | 158 | 272 | 126 | 10 | 9 |
| 16-C6 | 91.6 | 0.61 | 6.91 | 0.65 | 218 | 650 | 136 | 11 | 9 |
| 17-C1 | 99.2 | 0.19 | 0.37 | 0.16 | 147 | 198 | 112 | 7 | 3 |
| 18-C2 | 96.9 | 1.50 | 0.94 | 0.41 | 302 | 825 | 196 | 7 | 4 |
| 20-C5 | 97.3 | 0.57 | 1.43 | 0.49 | 281 | 514 | 156 | 11 | 11 |
| 22-C3 | 99.2 | 0.09 | 0.36 | 0.21 | 174 | 227 | 129 | 8 | 6 |
| 23-C5 | 98.9 | 0.19 | 0.66 | 0.19 | 124 | 127 | 129 | 10 | n.d. |
| 24-C6 | 97.0 | 0.36 | 2.28 | 0.24 | 211 | 346 | 135 | 3 | 4 |
| 25-C6 | 99.0 | 0.12 | 0.52 | 0.19 | 215 | 315 | 134 | 8 | 8 |
| 26-C6 | 97.9 | 0.38 | 1.19 | 0.27 | 296 | 483 | 188 | 10 | 10 |
| 27-C6 | 94.6 | 2.35 | 1.80 | 0.70 | 318 | 1112 | 196 | 8 | 4 |
| 28-C5 | 80.1 | 16.00 | 1.76 | 0.70 | 884 | 4546 | 218 | 24 | n.d. |
| 29-C2 | 97.8 | 1.20 | 0.47 | 0.20 | 500 | 1032 | 114 | 6 | 7 |
| 30-C2 | 98.6 | 0.43 | 0.60 | 0.18 | 275 | 462 | 100 | 6 | 4 |

n.d.: not detected.

In Table 3, it can be verified that the analyzed archaeological samples from the three sites are monotonically calcitic with very low Mg contents and a chemical composition that is comparable to the prevalent calcitic marbles from Gualba and Ceret.

**Table 3.** Chemical composition of the analyzed archaeological samples from Girona (Gi), Breda (Br) and Montsoriu (Mo).

| | CaCO₃ (%) | MgO (%) | SiO₂ (%) | Al₂O₃ (%) | Mn (ppm) | Fe (ppm) | Sr (ppm) | Zn (ppm) | Y (ppm) |
|---|---|---|---|---|---|---|---|---|---|
| Gi-1 | 98.7 | n.d. | 0.68 | 0.45 | 302 | 371 | 101 | 9 | 1 |
| Gi-2 | 99.3 | 0.01 | 0.39 | 0.18 | 266 | 378 | 87 | 7 | 2 |
| Gi-3 | 98.3 | 0.32 | 0.69 | 0.27 | 399 | 1262 | 107 | 11 | 8 |
| Gi-4 | 98.9 | 0.08 | 0.59 | 0.29 | 320 | 413 | 94 | 10 | 8 |
| | | | | | | | | | |
| Br-1 | 98.1 | 0.19 | 0.55 | 0.32 | 574 | 2630 | 341 | 11 | 11 |
| Br-2 | 98.3 | 0.00 | 1.14 | 0.35 | 357 | 532 | 111 | 11 | 5 |
| Br-3 | 98.7 | 0.02 | 0.71 | 0.34 | 321 | 511 | 120 | 10 | 11 |
| Br-4 | 98.7 | 0.05 | 0.68 | 0.28 | 398 | 920 | 123 | 14 | 11 |
| Br-5 | 98.6 | 0.03 | 0.83 | 0.35 | 367 | 588 | 128 | 12 | 10 |
| Br-6 | 97.1 | 0.85 | 1.14 | 0.59 | 227 | 874 | 121 | 11 | 1 |

**Table 3.** *Cont.*

| | CaCO₃ (%) | MgO (%) | SiO₂ (%) | Al₂O₃ (%) | Mn (ppm) | Fe (ppm) | Sr (ppm) | Zn (ppm) | Y (ppm) |
|---|---|---|---|---|---|---|---|---|---|
| Mo-1 | 98.0 | 0.34 | 1.11 | 0.37 | 202 | 458 | 116 | 10 | 1 |
| Mo-2 | 98.5 | 0.07 | 0.73 | 0.42 | 298 | 724 | 129 | 9 | 6 |
| Mo-3 | 98.2 | 0.16 | 0.90 | 0.39 | 369 | 1007 | 129 | 12 | 9 |
| Mo-4 | 98.8 | 0.02 | 0.65 | 0.38 | 267 | 371 | 102 | 14 | 11 |
| Mo-5 | 98.8 | n.d. | 0.53 | 0.46 | 376 | 588 | 114 | 10 | n.d. |
| Mo-7 | 99.1 | n.d. | 0.50 | 0.25 | 291 | 507 | 121 | 7 | 8 |
| Mo-8 | 98.7 | n.d. | 0.75 | 0.43 | 266 | 416 | 105 | 18 | 10 |
| Mo-9 | 95.5 | 1.18 | 2.88 | 0.19 | 366 | 717 | 123 | 7 | 4 |
| Mo-10 | 98.1 | 0.49 | 0.74 | 0.32 | 461 | 1021 | 120 | 12 | 11 |

Note: n.d.: not detected.

Using unsupervised MDA, like principal component analysis (PCA), most of the samples (including the archaeological ones) appear to form a very compact cluster. Using non-standardized compositional values, only the dolomitic samples (3-G2, 4-C2, 7-C3, 10-C4 and 28-C5) appear outside the compact cluster, forming an additional distinct cluster, and the carbonate-silicate rock (sample 4-G2) appears in a very different position compared to the rest of the samples. Using standardized values, the role of non-major elements is highlighted (Figure 9), the Mg-rich samples (correlated to the Mn and Fe values) also appear as outliers, and sample 4-G2 appears again isolated. Some quarry samples (non-Mg-rich) present anomalous high PC2 values, and they also appear out of the compact cluster. However, an archaeological sample (Br-1) also exhibits anomalous PC2 values (basically related to high Sr values). A close look at the compact cluster (Figure 9) reveals that the rest of the archaeological samples distribute intermingled with most of the Gualba and Ceret samples without any clear pattern.
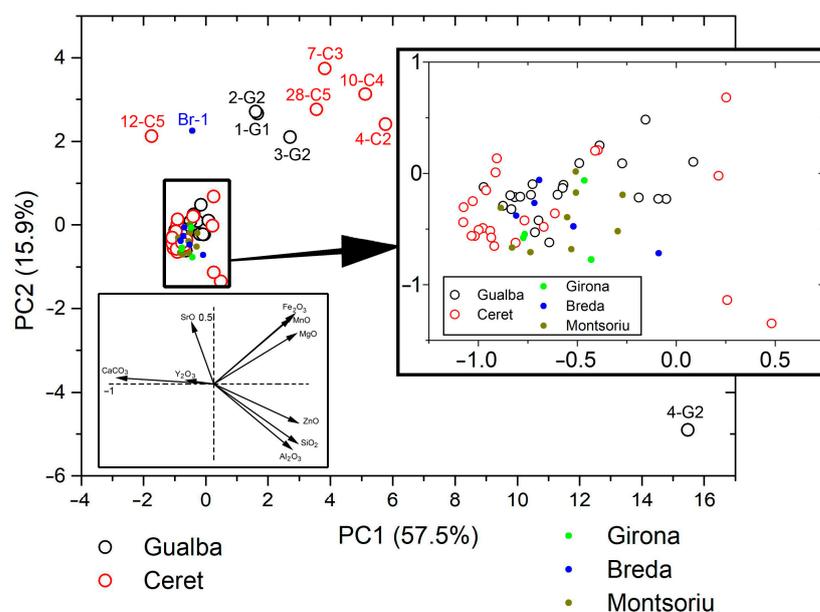


**Figure 9.** Standardized PCA biplot of factor scores for the first two principal components for all the chemically analyzed samples. Labels of the outliers have been specified, and the small area that concentrates most of the samples has been zoomed in. Inset: PCA biplot of the most relevant variables.

Considering that all the archaeological samples are pure calcite marbles, it was considered appropriate to remove the Mg-rich quarry samples (along with the carbonate-silicate rock, 4-G2). Besides that, the $CaCO_3$ content appears heavily correlated to almost all the other compositional variables, and therefore, it could be removed from the PCA computation to avoid redundancy. Among the different tested PCA, one was particularly effective in decompressing the data. A standardized PCA computed after removing the $CaCO_3$, and normalizing the remaining variables to sum 100, results in a biplot (Figure 10) that distributes most of the Ceret samples in the area with negative PC1 values and most of the Gualba samples in the area with positive PC1 values. This suggests a possible criterion to distinguish both quarry districts. Looking at the biplot of the original variables (Figure 10, inset), this classification would imply that the $SiO_2$ content is irrelevant to distinguish the two quarry districts, and in contrast, the marbles from Ceret are statistically richer in Mg and poorer in all the other minor elements.
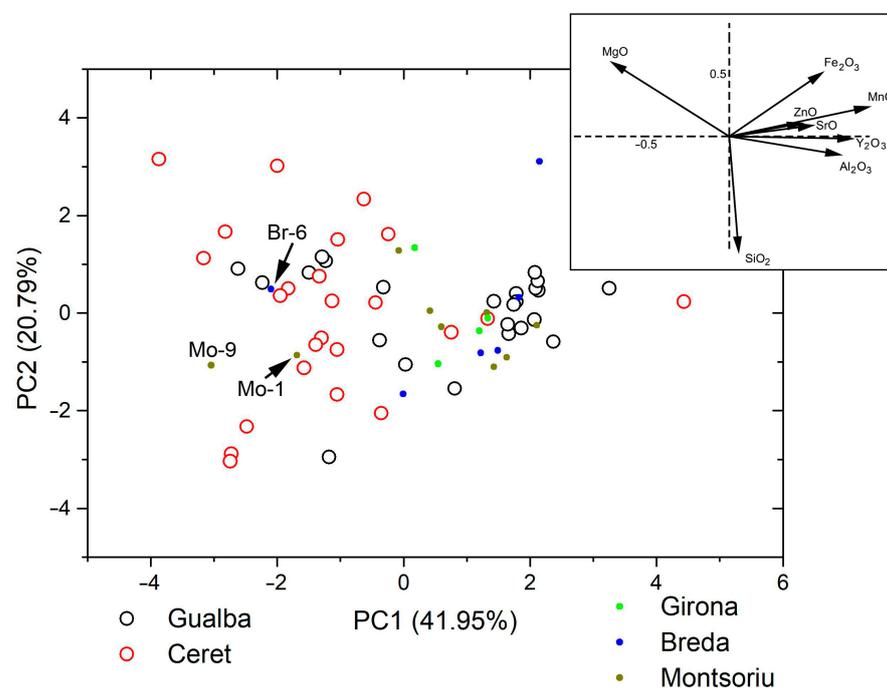


**Figure 10.** PCA biplot of factor scores for the first two principal components for all the analyzed calcitic marbles with the following data transformation: removal of $CaCO_3$ as variable, normalization of the remaining variables to sum 100 and standardization. Relevant archaeological samples have been labeled. Inset: PCA biplot of the original variables.

Within the mentioned biplot, the archaeological samples tend to lie in the area with zero or slightly positive values. According to the suggested criterion based on PC1 values, this would imply that most of the archaeological samples have a higher affinity with the samples from Gualba. However, these are only general trends that are not systematically kept and therefore appear speculative to draw provenance conclusions from them.

Another tested unsupervised MDA method was the hierarchical cluster analysis (HCA). The results are presented as dendrograms showing the hierarchical relationships and dissimilarity between samples (Figure 11).

Every dendrogram (computed using different transformed data) suggests different relationships between the samples. However, for all of them, Gualba and Ceret samples generally appear inextricably mixed together, and the archaeological samples never appear closely grouped but disseminated in different parts of the dendrogram and, therefore, do not show the expected affinity (e.g., Figure 11a,b). However, like the PCA shown in Figure 10, the HCA computed after removing the $CaCO_3$ and normalizing the remaining variables to sum 100 (Figure 11c) produces a dendrogram that separates most of Gualba

from Ceret samples. Indeed, a cut of this dendrogram at a dissimilarity threshold of 18.5 produces two groups (Table 4); group 1 contains most of the Gualba samples, and group 2 most of the Ceret samples.
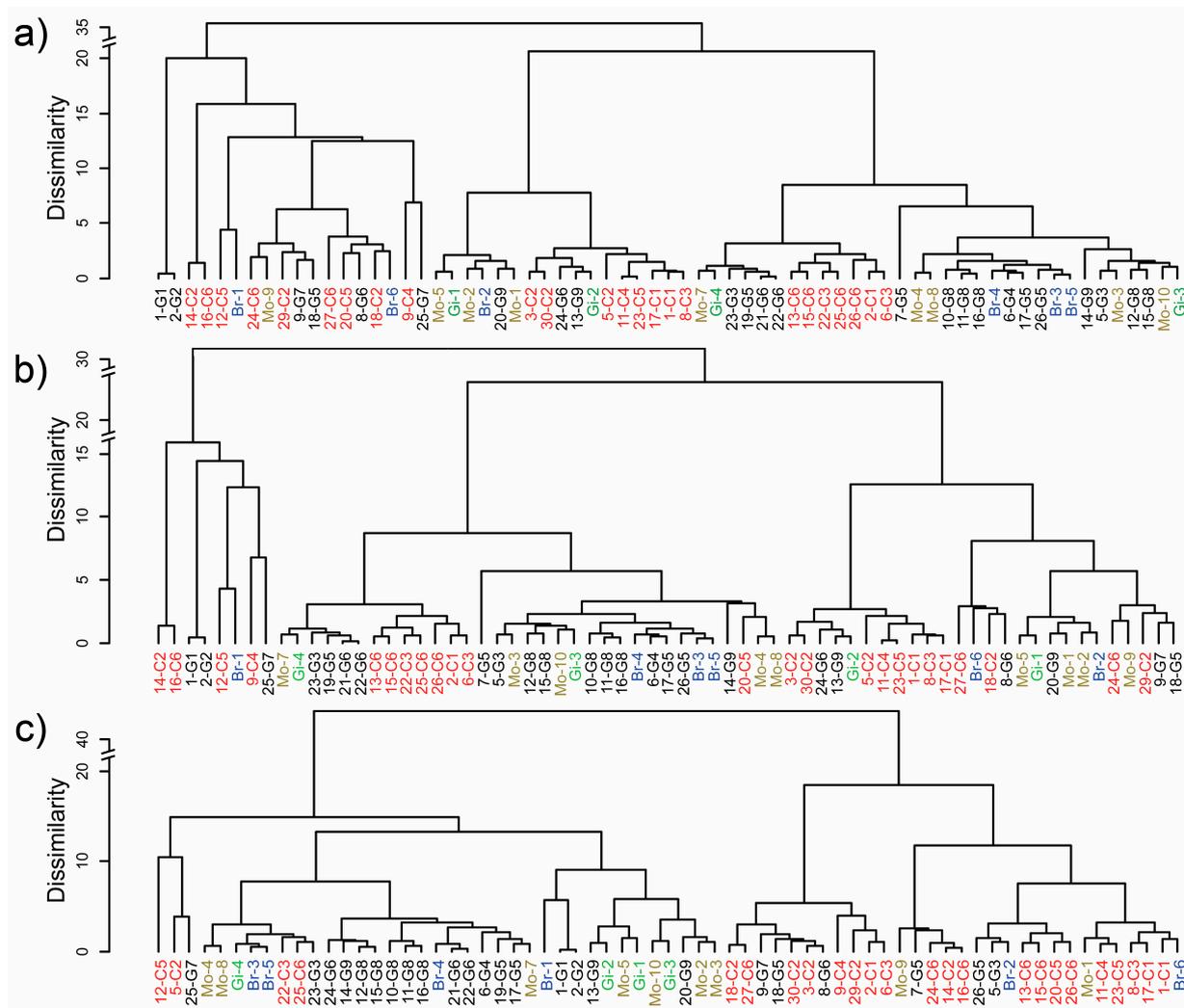


**Figure 11.** HCA dendrograms for all the analyzed calcitic marbles computed using the Euclidean distance and Ward's linkage method: (**a**) standardized; (**b**) disregarding CaCO$_3$ and standardizing; (**c**) disregarding CaCO$_3$, normalizing the remaining variables to sum 100 and then standardizing. The labels of all the samples have been specified and colored according to the corresponding quarry or archaeological site.

**Table 4.** Summary of the distribution of the quarry samples from Gualba and Ceret and the archaeological samples (Girona, Breda and Montsoriu) within the HCA dendrogram shown in Figure 11c.

|  | Group 1 | Group 2 |
| --- | --- | --- |
| Gualba | 19 | 6 |
| Ceret | 4 | 20 |
| Girona | 4 | 0 |
| Breda | 4 | 2 |
| Montsoriu | 7 | 2 |

The archaeological samples from the three sites were placed preferentially within group 1, suggesting again that Gualba could be the actual provenance of these samples. The archaeological samples that appear classified within the Ceret-dominated group (group 2) are Mo-1, Mo-9, Br-6 and Br-2. The first three appear to be the same divergent samples detected in the PCA outlined in Figure 10. However, it is worth mentioning that four Gualba samples were also classified within group 2, two of them closely related to the Br-2 sample. Additionally, six Ceret samples were placed within group 1, therefore in the absence of a dendrogram that could neatly separate the Gualba samples from the Ceret ones, the deductions from the position of the archaeological samples within the dendrogram should be relativized.

### 3.5. Supervised Machine Learning Classification Models

The supervised approach involves two steps. In the first step, the models are trained, and their performance is tested with additional labeled data (i.e., the quarry samples). Once assessed a successful training, the models were used to assign classes (i.e., origin) to the unlabeled samples (i.e., the archaeological marbles).

#### 3.5.1. Model Train and Test

After removing the Mg-rich samples, the reference database contains 24 calcite marbles from Gualba and 24 from Ceret. Every sample is described using eight features (as $CaCO_3$ data was disregarded). The 80-20 train–test split uses 40 samples to train the models, and only 8 samples (4 per class) are left to test them. Ten random seeds were used to produce ten different splits to run the models. After the test phase, for every split, a confusion matrix was obtained as an indication of the performance of the trained model, and for every split, six models were trained (including the stacker classifier). The 60 resulting confusion matrices can be found within the Supplementary Materials Section (Table S1), including the corresponding accuracies. The corresponding average values of accuracy (true positives divided by the total predictions) computed after the model training were therefore produced using 80 different model predictions for every tested model.

The undertaken train–test approach was close to the leave-one-out-cross-validation (LOOCV), which would be suitable for dealing with small datasets. However, LOOCV could have only produced 48 predictions resulting in a higher accuracy variance. The results from the 80 produced predictions per model indicate that all the models are capable of correctly assigning the origin of quarry samples with a very high accuracy (see the computed values in Table S1). Using some of the seeds, the eight samples are correctly classified (see, for instance, the GLM model using splits 1, 3, 6, 7 and 8 in Table S1).

Statistically, the average accuracy values (Figure 12) are around 0.9 for all the models except kkNN (with an average accuracy of 0.84). For all the other models, even the worst accuracies have values as high as 0.75 (and only for one or two seeds out of the ten). The kkNN model (with an accuracy of only 0.625 for one of the tested seeds) is confirmed as the model with the lowest performance, and apart from it, all the other models are highly proficient at distinguishing the Gualba and Ceret marbles. Considering the high average accuracies and their moderate variances, none of the tested classification models can be reported as a nonproficient classifier, and all of them will be taken into account to predict the class of the archaeological samples.
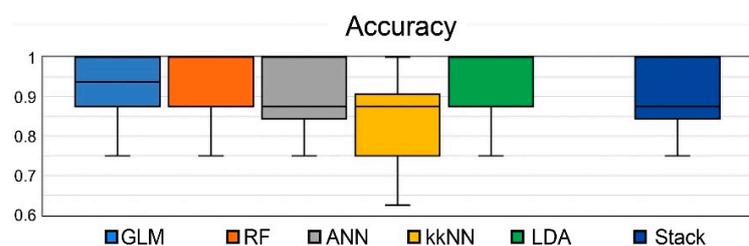


**Figure 12.** Bloxplot with the accuracy variation for each model using different splits.

3.5.2. Class Prediction

The ten random seeds result in ten different splits and hence ten trained setups for each model. Any of these setups can be used to classify the unlabeled groups of samples (i.e., the archaeological samples of unknown provenience). The classification results for any of the ten setups (or splits) can be consulted in Table S2. This table is an Excel spreadsheet organized into ten tabs and a final tab containing the combined results from all the splits). In Table S2, for every sample, the split tabs contain the assigned provenance produced by the five models (and also that from the stacking classifier) along with the corresponding numerical probabilities to belong to the Gualba quarries.

From Table S2, it can be noticed that the archaeological samples are mostly classified as members of the class "Gualba", with few exceptions. The most remarkable exception is sample Mo-9, which is systematically classified as "Ceret" by the GLM, NNET and LDA models regardless of the considered setup and also by RF and the stacking classifier in nine out of the ten trained setups. The only model that does not classify this sample so systematically within the "Ceret" class is KNN (only in four out of the ten setups). Another two samples (Mo-1 and Br-6) are also classified occasionally within the "Ceret" class by all the trained setups of the GLM model and most of the RF setups as well. However, the other models tend to classify these two samples as "Gualba". It is worth noting that the three mentioned samples (Mo-9, Mo-1 and Br-6) were also spotted as exhibiting a divergent behavior compared to the rest of the archaeological samples using unsupervised MDA. Sample Br-2 (classified within the Ceret-dominated group in the HCA dendrogram, Figure 11c) is unanimously classified in the "Gualba" class by all the models regardless of the considered trained setup. In contrast, sample Br-1 is preferentially classified as "Ceret" by the LDA model (and also, but not so clearly, by the KNN model). However, Br-1 is labeled as "Gualba" by most of the trained setups from the other models. Some other samples are also occasionally classified as "Ceret", but only for a few trained setups and not systematically for any of the classification models.

At any rate, the interpretation of the results from the supervised classification models has to be made statistically, considering the results from every group of archaeological samples (i.e., from every archaeological site) and combining all the trained splits for every classification model. The group probabilities of provenance for every set of archaeological samples are also available in Table S2. For any split and the resulting setup, the group probabilities always point to a Gualba provenance for the three archaeological sets regardless of the used classification algorithm. Table 5 (and also the CombinedSplits tab within Table S2) shows the overall probability values of belonging to either Gualba or Ceret for each of the three sets of archaeological marbles. These were produced by combining the results obtained using the ten different trained setups and the different samples from each archaeological site.

Using the statistical probability values, it is apparent that all the models point unambiguously towards Gualba as the class (i.e., origin) corresponding to the archaeological samples from the three investigated sites. In particular, the stacking classifier assigns the Gualba origin to the archaeological samples of Girona, Breda and Montsoriu with probabilities of 99%, 91% and 85%, respectively. The convergence of results produced using different trained models should be taken as proof of their reliability [25]. Of all the tested approaches, the supervised approach is the one that more solidly demonstrates competence to distinguish the marbles from the two quarry districts enabling an effective classification of the marbles from the three archaeological sites.

**Table 5.** Classification results, using the different trained models, in form of p-values (and associated uncertainties) describing the probability of provenance of the archaeological samples from Girona, Breda and Montsoriu to either the Gualba or Ceret quarries.

| Model | Quarry | Archaeological Site | | |
|---|---|---|---|---|
| | | Gerunda walls (Girona) | Breda's monastery | Montsoriu Castle |
| GLM | Gualba | $0.75 \pm 0.15$ | $0.75 \pm 0.23$ | $0.66 \pm 0.27$ |
| | Ceret | $0.25 \pm 0.15$ | $0.25 \pm 0.23$ | $0.34 \pm 0.27$ |
| RF | Gualba | $0.71 \pm 0.10$ | $0.80 \pm 0.21$ | $0.69 \pm 0.19$ |
| | Ceret | $0.29 \pm 0.10$ | $0.20 \pm 0.21$ | $0.31 \pm 0.19$ |
| ANN | Gualba | $0.94 \pm 0.06$ | $0.95 \pm 0.05$ | $0.81 \pm 0.31$ |
| | Ceret | $0.06 \pm 0.06$ | $0.05 \pm 0.05$ | $0.19 \pm 0.31$ |
| kkNN | Gualba | $0.96 \pm 0.12$ | $0.91 \pm 0.26$ | $0.89 \pm 0.29$ |
| | Ceret | $0.04 \pm 0.12$ | $0.09 \pm 0.26$ | $0.11 \pm 0.29$ |
| LDA | Gualba | $0.91 \pm 0.12$ | $0.84 \pm 0.30$ | $0.79 \pm 0.32$ |
| | Ceret | $0.09 \pm 0.12$ | $0.16 \pm 0.30$ | $0.21 \pm 0.32$ |
| Stacking classifier | Gualba | $0.99 \pm 0.07$ | $0.91 \pm 0.16$ | $0.84 \pm 0.29$ |
| | Ceret | $0.01 \pm 0.07$ | $0.09 \pm 0.16$ | $0.16 \pm 0.29$ |

## 4. Conclusions

The marbles from Gualba and Ceret districts share very similar characteristics. Petrographically, their marbles are generally calcitic and coarse-grained, and both exhibit a particularly intense orange cathodoluminescence, mostly homogeneous. Their measured isotopic ratios are also rather similar, although distinctly different compared with the known classical coarse-grained marbles due to their lower $\delta^{18}O$ ratios. The conventional multi-technique approach allows for the discrimination of these marbles from other coarse-grained marbles, but it is unable to discriminate between the marbles from the two districts.

Considering a binary classification problem using archaeological samples suspected to have been quarried from one of the two districts, the conventional approach also fails to assign a provenance to the archaeological samples.

Geochemically, the composition of the marbles from the two districts seems to be also rather similar. Unsupervised MDA methods like PCA and HCA can easily identify outliers with significant compositional differences (i.e., the few non-calcitic samples), but in general, these methods are unable to differentiate the Gualba samples from the Ceret samples. Only by applying a particular transformation to the geochemical dataset (namely removing the main component ($CaCO_3$), normalizing the remaining variables to sum 100 and standardizing to zero mean and unit variance) has it been possible to produce a PCA biplot and an HCA dendrogram with hints of effective discrimination between the two studied quarry districts. The position of the archaeological samples in these particular PCA and HCA diagrams shows a higher statistical affinity of most of these samples with those from the Gualba quarries. However, a few archaeological samples have divergent behavior.

The supervised approach applied to the geochemical data enables the discrimination between the two marble districts without any special transformation of the geochemical dataset (apart from removing $CaCO_3$, a heavily correlated feature). Different models, using different train–test splits, were successfully trained to discriminate between marbles from the two quarry districts, reaching accuracies generally above 0.9. The class of the archaeological samples predicted by the trained models is generally "Gualba", with few exceptions. The more persistent exceptions correlate with the samples with divergent behavior in the PCA and HCA with hints of effective discrimination. Considering the

results statistically, the few exceptions vanish, and the three sets of archaeological samples can clearly be assigned to Gualba.

From the archaeological point of view, the classification of the marble from the cloister in Breda and the Montsoriu Castle as marbles quarried in Gualba was already hypothesized, but it had not been previously proved archaeometrically. This origin appears almost obvious, taking into account the proximity of Gualba to both Breda and the Montsoriu (~5 km), and it could be taken as an additional verification of the consistency of the classification results.

In contrast, the provenance of the marble from the Gerunda city walls was not so obvious. Chronologically, this conclusion has a high significance as it would demonstrate that the Gualba marbles were already quarried during Antiquity. The possibility is not unlikely as the Via Augusta, the major road built by the Romans in Hispania, is known to pass through the present-day location of Gualba and, to the north, headed to Gerunda. Even the location of one of the mansiones (state post offices) of this Via, named Seterrae, had been occasionally identified with Gualba [56].

# References

1. Ricca, M.; Paladini, G.; Rovella, N.; Ruffolo, S.A.; Randazzo, L.; Crupi, V.; Fazio, B.; Majolino, D.; Venuti, V.; Galli, G.; et al. Archaeometric Characterisation of Decorated Pottery from the Archaeological Site of Villa Dei Quintili (Rome, Italy): Preliminary Study. *Geosciences* **2019**, *9*, 172. [CrossRef]
2. Herz, N. Provenance Determination of Neolithic to Classical Mediterranean Marbles Isotopes. *Archaeometry* **1992**, *34*, 185–194. [CrossRef]
3. Brilli, M.; Savin, M.-C. Provenance Study of the White Marbles of the "Baths of Elagabalus" at the Palatine Hill in Rome. *Archaeol. Anthropol. Sci.* **2019**, *11*, 5539–5551. [CrossRef]

4. Wielgosz-Rondolino, D.; Antonelli, F.; Bojanowski, M.J.; Gładki, M.; Göncüoğlu, M.C.; Lazzarini, L. Improved Methodology for Identification of Göktepe White Marble and the Understanding of Its Use: A Comparison with Carrara Marble. *J. Archaeol. Sci.* **2020**, *113*, 105059. [CrossRef]

5. Brilli, M.; Cavazzini, G.; Turi, B. New Data of 87Sr/86Sr Ratio in Classical Marble: An Initial Database for Marble Provenance Determination. *J. Archaeol. Sci.* **2005**, *32*, 1543–1551. [CrossRef]

6. Attanasio, D.; Platania, R.; Rocchi, P. White Marbles in Roman Architecture: Electron Paramagnetic Resonance Identification and Bootstrap Assessment of the Results. *J. Archaeol. Sci.* **2005**, *32*, 311–319. [CrossRef]

7. Gutiérrez Garcia-M, A.; Savin, M.C.; Cantin, N.; Boudoumi, S.; Lapuente, P.; Chapoulie, R.; Pianet, I. NMR as a New Tool for Cultural Heritage Application: The Provenance of Ancient White Marbles. *Archaeometry* **2019**, *61*, 795–808. [CrossRef]

8. Zöldföldi, J.; Hegedüs, P.; Székely, B. Interdisciplinary Data Base of Marble for Archaeometric, Art Historian and Restoration Use. In *Ancient Mining in Turkey and the Eastern Mediterranean*; Yalcin, Ü., Özbal, H., Pasamehmetoglu, G., Eds.; Atilim University: Ankara, Turkey, 2008; pp. 225–251.

9. Antonelli, F.; Nestola, F. An Innovative Approach for Provenancing Ancient White Marbles: The Contribution of X-ray Diffraction to Disentangling the Origins of Göktepe and Carrara Marbles. *Sci. Rep.* **2021**, *11*, 22312. [CrossRef]

10. Lepsius, G.R. *Griechische Marmorstudien*; Verlag der Königlichen Akademie der Wissenschaften: Berlin, Germany, 1890.

11. Poretti, G.; Brilli, M.; De Vito, C.; Conte, A.M.; Borghi, A.; Günther, D.; Zanetti, A. New Considerations on Trace Elements for Quarry Provenance Investigation of Ancient White Marbles. *J. Cult. Herit.* **2017**, *28*, 16–26. [CrossRef]

12. Prochaska, W. The Challenge of a Successful Discrimination of Ancient Marbles (Part II): A Databank for the Alpine Marbles. *J. Archaeol. Sci. Rep.* **2021**, *38*, 102958. [CrossRef]

13. Prochaska, W.; Attanasio, D. The Challenge of a Successful Discrimination of Ancient Marbles (Part III): A Databank for Aphrodisias, Carrara, Dokimeion, Göktepe, Hymettos, Parian Lychnites and Pentelikon. *J. Archaeol. Sci. Rep.* **2022**, *45*, 103582. [CrossRef]

14. Vasiliki, A.; Sabine, L.; Walter, P. Meter Reliefs from Ephesos and the Obizzi Collection (KHM Wien). Archaeometric Considerations on the Production and Trade of Small-Format Marble Objects in the Greek World. *J. Archaeol. Sci. Rep.* **2023**, *47*, 103742. [CrossRef]

15. Vettor, T.; Sautter, V.; Pont, S.; Harivel, C.; Jolivet, L.; Moretti, I.; Moretti, J.C. Delos Archaeological Marbles: A Preliminary Geochemistry-Based Quarry Provenance Study. *Archaeometry* **2021**, *63*, 907–922. [CrossRef]

16. Efenberger-Szmechtyk, M.; Nowak, A.; Kregiel, D. Implementation of Chemometrics in Quality Evaluation of Food and Beverages. *Crit. Rev. Food Sci. Nutr.* **2018**, *58*, 1747–1766. [CrossRef]

17. Baxter, M.J. Standardization and Transformation in Principal Component Analysis, with Applications to Archaeometry. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1995**, *44*, 513–527. [CrossRef]

18. McKeague, P.; van't Veer, R.; Huvila, I.; Moreau, A.; Verhagen, P.; Bernard, L.; Cooper, A.; Green, C.; van Manen, N. Mapping Our Heritage: Towards a Sustainable Future for Digital Spatial Information and Technologies in European Archaeological Heritage Management. *J. Comput. Appl. Archaeol.* **2019**, *2*, 89–104. [CrossRef]

19. Fiorucci, M.; Khoroshiltseva, M.; Pontil, M.; Traviglia, A.; Del Bue, A.; James, S. Machine Learning for Cultural Heritage: A Survey. *Pattern Recognit. Lett.* **2020**, *133*, 102–108. [CrossRef]

20. Oonk, S.; Spijker, J. A Supervised Machine-Learning Approach towards Geochemical Predictive Modelling in Archaeology. *J. Archaeol. Sci.* **2015**, *59*, 80–88. [CrossRef]

21. Lopez-Garcia, P.A.; Argote, D.L.; Thrun, M.C. Projection-Based Classification of Chemical Groups for Provenance Analysis of Archaeological Materials. *IEEE Access.* **2020**, *8*, 152439–152451. [CrossRef]

22. Barone, G.; Mazzoleni, P.; Spagnolo, G.V.; Raneri, S. Artificial Neural Network for the Provenance Study of Archaeological Ceramics Using Clay Sediment Database. *J. Cult. Herit.* **2019**, *38*, 147–157. [CrossRef]

23. Ma, Q.; Yan, A.; Hu, Z.; Li, Z.; Fan, B. Principal Component Analysis and Artificial Neural Networks Applied to the Classification of Chinese Pottery of Neolithic Age. *Anal. Chim. Acta* **2000**, *406*, 247–256. [CrossRef]

24. Anglisano, A.; Casas, L.; Anglisano, M.; Queralt, I. Application of Supervised Machine-Learning Methods for Attesting Provenance in Catalan Traditional Pottery Industry. *Minerals* **2020**, *10*, 8. [CrossRef]

25. Anglisano, A.; Casas, L.; Queralt, I.; Di Febo, R. Supervised Machine Learning Algorithms to Predict Provenance of Archaeological Pottery Fragments. *Sustainability* **2022**, *14*, 11214. [CrossRef]

26. Nolla i Brufau, J.M.; Amich i Raurich, N.M.; Castanyer i Masoliver, P. El Baix Imperi. La Consolidació de La Ciutat. In *Història de Girona*; Costa, L., Maroto, J., Eds.; CCG Edicions: Girona, Spain, 2006; pp. 93–100.

27. Nolla, J.M. *Girona Romana. de la Fundació a la Fi del Món Antic*; Ajuntament de Girona. Diputació de Girona: Girona, Spain, 1987.

28. Nolla, J.M.; Sagrera, J. Noves Exploracions Arqueològiques a La Caserna d'Alemanys (Girona). Campanya d'excavacions de 1988. *Cypsela* **1991**, *IX*, 177–195.

29. Oliver Vert, J. The Commemorative Roman Arch-Gate of Gerunda (Girona). *Arqueol. De La Arquit.* **2019**, *16*, e082.

30. Oliver, J.; Gutiérrez, A. Torre Del Telègraf Òptic o Del Llamp–Recollida de Mostres per Anàlisi Arqueomètrica Del Material Lapidi (Girona, Gironès). In *Proceedings of the Quinzenes Jornades d'Arqueologia de les Comarques de Girona*; Burch, J., Buxó, R., Frigola, J., Fuertes, M., Manzano, S., Mataró, M., Eds.; Generalitat de Catalunya, Museu d'Arqueologia de Catalunya, Universitat de Girona, Documenta Universitaria: Castelló d'Empúries, Spain, 2020; pp. 255–258.

31. Pou, J.; Frigola, J.; Òdena, J.L. Intervenció Arqueològica al Claustre Del Monestir de Sant Salvador (Breda, La Selva). In *Proceedings of the Quinzenes Jornades d'Arqueologia de les Comarques de Girona*; Burch, J., Buxó, R., Frigola, J., Fuertes, M., Manzano, S., Mataró, M., Eds.; Generalitat de Catalunya, Museu d'Arqueologia de Catalunya, Universitat de Girona, Documenta Universitaria: Castelló d'Empúries, Spain, 2020; pp. 451–455.

32. Tura i Masnou, J.; Pujadas i Mitjà, S.; Mateu i Gasquet, J.; Llorens i Rams, J.M.; Font i Valentí, G. Excavacions Arqueològiques al Castell de Montsoriu. Campanyes 2016–2017 (Arbúcies-Sant Feliu de Buixalleu, La Selva). In *Proceedings of the Catorzenes Jornades d'Arqueologia de les Comarques de Girona, Caldes de Malavella, 2018*; Buxó, R., Manzano, S., Mataró, M., Nolla, J.M., Eds.; Generalitat de Catalunya, Museu d'Arqueologia de Catalunya, Universitat de Girona, Ajuntament de Caldes de Malavella: Caldes de Malavella, Spain, 2018; pp. 485–495.

33. Font, G.; Llorens, J.M.; Mateu, J.; Pujadas, S.; Rueda, J.M.; Tura, J. El Castell de Montsoriu. Configuració Arquitectònica d'un Gran Palau Gòtic, a Partir de Les Tasques Arqueològiques Desenvolupades Entre Els Anys 2001 i 2004. In *Proceedings of the VI Trobada d'Estudiosos del Montseny*; Hernández, J., Santiago, S., Melero, J., Eds.; Diputació de Barcelona: Barcelona, Spain, 2005; pp. 185–189.

34. Huerta, J. El Paleozoico Inferior Del Montseny. *Acta Geologica Hispanica* **1990**, *25*, 105–111.

35. Liesa, M.; Carreras, J. On the Structure and Metamorphism of the Roc de Frausa Massif (Eastern Pyrenees). *Geodin. Acta* **1989**, *3*, 149–161. [CrossRef]

36. Aguilar, C.; Liesa, M.; Štípská, P.; Schulmann, K.; Muñoz, J.A.; Casas, J.M. P–T–t–d Evolution of Orogenic Middle Crust of the Roc de Frausa Massif (Eastern Pyrenees): A Result of Horizontal Crustal Flow and Carboniferous Doming? *J. Metamorph. Geol.* **2015**, *33*, 273–294. [CrossRef]

37. Pagniez, L. Le Marbre de Céret: Un Matériau Complexe et Méconnu de La Production Artistique Roussillonnaise (XIe-XVe Siècle). *Les Cahiers De Saint-Michel-De-Cuxa* **2002**, *33*, 159–171.

38. Mallet, G. De l'usage Des Marbres En Roussillon Entre Le XIe et Le XIVe Siècle: La Sculpture Monumentale. *Parimoines du Sud.* **2016**, *4*, 29–51. [CrossRef]

39. Álvarez, A.; Rodà, I.; Mayer, M. Mármoles y Calizas de Los Pirineos Centrales y Su Utilización En Época Romana. In *Proceedings of the Les Ressources Naturelles des Pyrénées. Leur Exploitation durant l'Antiquité (Toulouse 1999)*; Sablayrolles, R., Ed.; Musée archéologique départemental: Saint-Bertrand-de-Comminges, France, 2001; pp. 47–67.

40. Mallet, G. De Catalogne En Languedoc Méditerranéen Au Moyen Âge (IVe-XIIe Siècles): Questions Sur Les Remplois En Marbre Blanc à Travers Les Exemples Roussillonnais. *Hortus Artium Medievalium* **2011**, *17*, 77–84. [CrossRef]

41. Ponsich, P. Les plus Anciennes Sculptures Médiévales Du Roussillon (Ve-XIe Siècle). *Les cahiers de Saint-Michel-de-Cuxa* **1980**, *11*, 293–331.

42. Pagniez, L. *La Production Artistique Roussillonnaise En Marbre de Céret (XIe-XVe Siècle)*; Mémoire de Maîtrise, Université Paul-Valéry: Montpellier, France, 1999.

43. Gély, J.-P. Le Marbre de Céret (Pyrénées-Orientales): Neuf Siècles d'extraction et d'emploi Dans La Décoration de l'art Roussillonnais. In *Proceedings of the Carrières et Constructions en France et dans les Pays Limitrophes, III, Actes du 119e Congrès National des Sociétés Historiques et Scientifiques (Amiens, 1994)*; Lorenz, J., Ed.; Éditions du CTHS: Amiens, France, 1996; pp. 385–397.

44. Blanc, P.; Gély, J.-P. L'analyse de La Provenance Des Marbres Blancs Par Cathodoluminescence et Son Application Au Marbre de Céret. In *Pierre et Archéologie*; Lorenz, J., Miskovsky, J.-C., Eds.; Presses Universitaires de Perpignan: Perpignan, France, 2002; pp. 81–113.

45. Giresse, P.; de Barrau, C.; Bromblet, P. Les Marbres et Les Calcaires de La Façade de l'abbatiale Romane de Saint-André-de-Sorède (Sud Du Roussillon). Sources, Emplois et Réemplois. In *El coste de la construcción medieval: Materiales, recursos y sistemas constructivos para la petrificación del paisaje entre los siglos XI y XIII.*; Maira Vidal, R., Rodríguez, A., Eds.; Instituto de Historia (CSIC)-Instituto Juan de Herrera: Madrid, Spain, 2021; pp. 187–205.

46. Giresse, P.; Bromblet, P.; Barrau, C. de Le Marbre de Carrare Dans Les Oeuvres Romanes Du Roussillon Pendant Le Xie Siècle (Linteau de Saint-Genis-Desfontaines, Linteau et Table d'autel de Saint-André-Desorède). Analyses Texturales et Isotopiques. *ArcheoSciences* **2020**, *44*, 71–79. [CrossRef]

47. Lapuente Mercadal, M.P.; Cuchí Oterino, J.A.; Royo Plumed, H.; Preite Martínez, M.; Blanc, P.; Garcés Manau, C. Study of Provenance of the Roman Sarcophagus Known Today as the Tomb of King Ramiro II of Aragon. In *Proceedings of the Interdisciplinary Studies on Ancient Stone: Proceedings of the IX Association for the Study of Marbles and Other Stones in Antiquity (ASMOSIA) Conference (Tarragona 2009)*; Gutiérrez García-Moreno, A., Lapuente Mercadal, M.P., Rodà de Llanza, I., Eds.; Institut Català d'Arqueologia: Tarragona, Spain, 2012; pp. 419–425.

48. Wickham, H.; Bryan, J. *R Packages*, 2nd ed.; O'Reilly Media Inc.: Sebastopol, CA, USA, 2023; ISBN 9781098134945.

49. Burkhard, M. Calcite Twins, Their Geometry, Appearance and Significance as Stress-Strain Markers and Indicators of Tectonic Regime: A Review. *J. Struct. Geol.* **1993**, *15*, 351–368. [CrossRef]

50. Valley, J.W. Stable Isotope Geochemistry of Metamorphic Rocks. *Rev. Mineral. Geochem.* **1986**, *16*, 445–489.

51. Sharp, Z.D. Metamorphic Petrology. In *Principles of Stable Isotope Geochemistry*; Pearson/Prentice Hall: Upper Saddle River, NJ, USA, 2007; pp. 1–36. ISBN 0130091391.

52. Dessandier, D.; Bromblet, P.; Leroux, L. *Étude Des. Pierres de Monuments Emblématiques Du. Bâti Historique de Perpignan. (66)*; BRGM: Perpignan, France, 2011.

53. Antonelli, F.; Lazzarini, L. An Updated Petrographic and Isotopic Reference Database for White Marbles Used in Antiquity. *Rendiconti Lincei* **2015**, *26*, 399–413. [CrossRef]

54. Moreira, A.J.C.; Santos, M.Y. Concave Hull: A k-Nearest Neighbours Approach for the Computation of the Region Occupied by a Set of Points. In *Proceedings of the International Conference on Computer Graphics Theory and Applications*; SciTePress: Setúbal, Portugal, 2007.
55. Carr, D.D.; Rooney, L.F. Limestone and Dolomite. In *Industrial Minerals and Rocks*; Lefond, S.Y., Ed.; American Institute of Mining, Metallurgical, and Petroleum Engineers, Inc.: New York, NY, USA, 1983; pp. 833–868.
56. De Soto Cañamares, P. Anàlisi de La Xarxa de Comunicacions i Del Transport a La Catalunya Romana: Estudis de Distribució i Mobilitat. Doctoral Thesis, Institut Català d'Arqueologia Clàssica-Universitat Autònoma de Barcelona, Tarragona, Spain, 2010.