



# Principal Components Analysis (PCA) of Monument Stone Decay by Rainwater: a case study of “Basílica da Estrela” church, Portugal<sup>†</sup>

Carlos Figueiredo<sup>1\*</sup>, Carlos Alves<sup>2</sup>

<sup>1</sup> CERENA - Centro de Recursos Naturais e Ambiente, UID/ECI/04028/2013, DEcivil, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisbon, Portugal; carlos.m.figueiredo@ist.utl.pt

<sup>2</sup> LandS/Lab2PT (Landscape, Heritage and Territory Laboratory), UID/AUR/04509/2013; POCI-01-0145-FEDER-007528 and Earth Sciences Department, School of Sciences, University of Minho, Portugal; casaix@dct.uminho.pt

\* Correspondence: carlos.m.figueiredo@ist.utl.pt; Tel.: +351-21-841-7023

† Presented at the title, place, and date.

Received: date; Accepted: date; Published: date

**Abstract:** An extended version of Principal Components Analysis (PCA) of monument stone decay phenomena occurring at “Basílica da Estrela” church, Lisbon, Portugal, is now presented. The PCA rationale and general methodological procedure is presented, as a first step of a stepwise approach to the eigenvector methods of data analysis. PCA, as others “Eigenvector Methods”, seeks to reveal the underlying structure that might exist within a set of multivariate observations. Temperature, pH, electrical conductivity and main ionic species were measured on several seepage samples over three years inside the monument. PCA results are discussed in the perspective of a nondestructive tool.

**Keywords:** Built environment; water-induced alteration processes; monument stone decay; multivariate data; Principal Components Analysis

---

## 1. Introduction

Based on previous studies [1-3], an extended version of Principal Components Analysis (PCA) of monument stone decay phenomena occurring at “Basílica da Estrela”, Portugal, is now presented [1-3]. The “Basílica da Estrela” church is the most relevant 18<sup>th</sup> century monument in the city of Lisbon, Portugal. Built with limestone, it is located in a moderately polluted area about 15 km away from the sea. Physical weathering forms, such as granular disintegration, flakes, scales and spalling, prevail inside. However, chemical weathering forms are also present and are largely dominated by calcite re-precipitation forming large white zones. Soluble salts were, in contrast, practically non-existent or hardly found. Regarding monument stone decay and conservation issues, infiltration of rainwater through the terrace of the church was pointed out as being the main problem of the monument [3]. In [2] there was a preliminary presentation of PCA (plot of variables in the plane of the two first components) and here the structure of the relations between variables and samples will be developed.

To study the water-rock interaction and the stone decay phenomena occurring at “Basílica da Estrela”, chemical and physical analyses were performed on seepage water samples collected over three years inside (high choir) the church. Temperature, pH, electrical conductivity and main ionic species were measured on each sample. The data gathered over the sampling period have been worked out and will be analysed here using Principal Components Analysis (PCA).

PCA approach will be used, in this paper, to help data interpretation and also as a first step of a stepwise approach to the eigenvector methods of data analysis. Only the rationale and the general

methodological procedures used in PCA will now be presented. This means that, the mathematics (theoretical and practical manipulation) and the computational essentials underlying PCA implementation, are beyond the scope of our paper. Those interested in should be referred to [4], for instance. As a first approach, our description of PCA will rely heavily on Davis [4]. In this paper we have used the computer program “STAT-ITCF (1987, 3<sup>rd</sup> version). ITCF stands for “Institut Technique des Céréales et des Fourrages”.

There are some examples of the use of PCA in studies to address monument stone decay problems. Moropoulou et al. [5] tried to show how PCA could be a powerful tool in conservation and restoration problems. Bacci al. [6] used PCA for discriminating areas resulting from degradation process of calcareous stones due to interaction with atmospheric pollutants in urban and industrialized areas. Benavente et al. [7] applied PCA to analyse the relationship between petrophysical properties and durability when they were studying salt weathering in dual-porosity building dolostones. Maurício et al. [8] applied Benzécri Multiple Correspondence Factor Analysis (MCFA) to chemical data of salt efflorescences collected at Sta Marija Ta’ Cwerra Church in Malta. Torfs et al. [9] applied also PCA, combined with other methods, in the analysis of the data they had collected in their study on the environmental effects on the deterioration of the church of Sta Marija Ta’ Cwerra, Malta. Zezza et al. [10] included also Principal Components Analysis in their study of the environmental effects on the deterioration of the Cathedral of Bari, Italy.

## 2. Methods

### 2.1. Data Sampling

To assess the rainwater-induced alteration processes at “Basílica da Estrela”, physical and chemical analyses were performed on seventeen seepage water samples collected over three years inside the church at the elevated choir. Seepage water samples resulted from rainwater that was able to penetrate the building through the roof and percolated the monument changing its composition through water-rock interactions. This approach can be considered a NDT tool for the characterization of alteration of geologic materials in the built environment as it does not involve the extraction of samples from those materials.

For each sample, the main chemical species determined were respectively  $\text{Cl}^-$ ,  $\text{NO}_3^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{HCO}_3^-$ ,  $\text{CO}_3^{2-}$  and  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ . Electrical conductivity, pH and temperature were also measured. The type of data gathered all over the sampling period form the raw data set that was worked out and analysed using Principal Components Analysis (PCA), in this paper. PCA approach was used to help data interpretation. Table 1 gives the basic statistical parameters of the raw data set. This consists of 12 selected variables (physical and chemical properties, table columns) measured on seventeen samples.

**Table 1.** Raw data set: basic statistical parameters of physical and chemical properties measured on seventeen seepage water samples. The chemical analyses are in weight per cent. MAX: maximum value; AV: average (mean); MIN: minimum value; STD: standard deviation.

Variables - Physical and chemical properties												
	T (°C)	pH	$\sigma$ ( $\mu\text{S}/\text{cm}$ )	$\text{HCO}_3^-$ (%)	$\text{CO}_3^{2-}$ (%)	$\text{Cl}^-$ (%)	$\text{NO}_3^-$ (%)	$\text{SO}_4^{2-}$ (%)	$\text{Na}^+$ (%)	$\text{K}^+$ (%)	$\text{Ca}^{2+}$ (%)	$\text{Mg}^{2+}$ (%)
MAX	25.0	11.7	1424.0	310.6	163.7	67.4	20.5	29.9	122.5	330.0	46.5	0.20
AV	19.5	10.3	722.4	111.0	98.5	34.9	9.4	9.1	67.6	177.9	4.4	0.06
MIN	16.7	8.2	100.4	4.0	1.2	23.7	0.00	0.00	41.0	115.0	0.7	0.00
STD	2.3	0.8	305.9	93.2	36.1	11.8	6.0	7.1	22.5	58.2	10.9	0.06

## 2.2. Principal Components Analysis (PCA)

### 2.2.1. The Rational

According to Davis [4], PCA is a factor analysis technique designed for interval or ratio data that are measurements made on a continuous numerical scale. If the raw multivariate data matrix has  $n$  rows that represent observations/samples and  $m$  columns of variables, the  $n$  samples or objects may be regarded as being points located in the  $m$ -dimensional space defined by the  $m$  variables. In general, PCA as any other eigenvalue and eigenvector methods was originally devised to explain the interrelationships in a large numbers of variables by the presence of a few factors or principal components or axes. It has as its main purpose to decompose the larger  $m$ -dimensional space (a multivariate set of observations) into a smaller  $p$ -dimensional one, by computing new, uncorrelated orthogonal components that are linear combination of the original variables and losing as less as possible of the variance in the original data set. The new components are called *principal components* of the multivariate data matrix. The linear transformation of  $m$  original variables to  $p$  new variables is performed in a fashion that requires that each new variable accounts for, successively, as much of the total variance as possible. But the question, now, is: how many factors should be retained? A general pragmatic approach may consist of extracting only two or three factors, because this is the maximum number that can conveniently be displayed as scatter diagrams and any number larger than this increases the dimensionality of the problem to the point where it again becomes difficult to grasp [4]. The factor axes/or principal components are, then, plotted two at a time as 2D flat diagrams that are more easily manageable and perceptible dimensions at just one glance. We should need only  $p$  factor axes to explain our data and the usual assumption is that  $p < m$ .

Finally, it should be stressed out that the principal components have to be interpreted in terms of original variables. However, sometimes this may not be made as easier as we wish. Then a full circle approach from variables to principal components, for reduction in the size of the problem, back to variables for interpretation of the principal components, is usually used. 2D diagrams of the principal component's loadings show the correlation among the original variables themselves and also between these and the principal components analysis axes. On the other hand, projecting the samples scores (samples co-ordinates) onto the first two principal axes (interpreted in terms of the original variables according to their loadings in the principal components) some significant insight into the inter-samples relationships in the data set could also be obtained. This way, this may help the analyst to explore the inter-variable relationships, the inter-object (similitude) relationships in a given data set, as well as the interrelations between the variables and objects with each of the respective principal components.

As the principal components are linear transformations of the  $m$  original variables, we are able to plot PCA scores simply by projecting our original observations onto the principal axes [4].

### 2.2.1. General Methodological Algorithm: a simple layout

Here, we will try to sketch out only some of the general or basic computational steps usually involved in principal components analysis, as it is, supposedly, implemented in several commonly available libraries of computer programs. As other factor analysis techniques, PCA tries to explore some of the mathematical and computational relationships that exist between a data matrix, its matrices of cross-products, and their eigenvalues and eigenvectors.

Principal components are nothing more than the eigenvectors of a variance-covariance or a correlations matrix. PCA is, then, concerned with finding these axes and measuring their magnitude. It starts by extracting the eigenvalues and eigenvectors of a variance-covariance or correlations matrix, and then discarding the less important of these. The eigenvectors are the coordinates of the principal components axes of the data set. They may provide significant insights into the structure underlying the data set, yielding the orientations of the principal axes of the cloud of points. On the other hand, the eigenvalues represent the lengths of the successive principal axes or principal components. That is, they represent the amount or proportion of the total variance transported or accounted for by the eigenvectors.

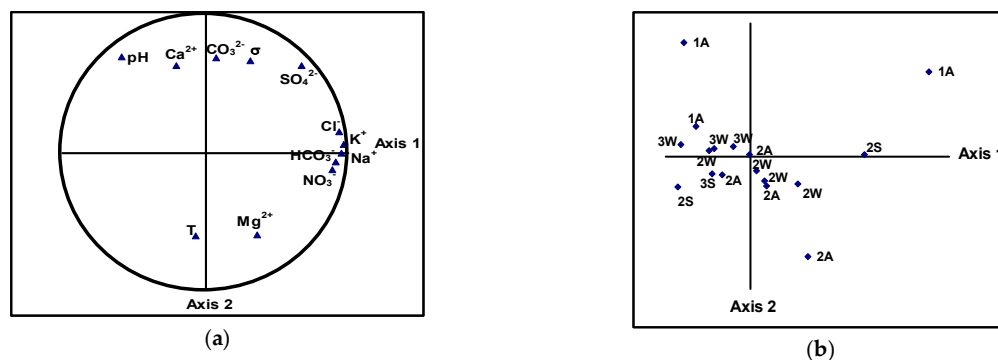
So, in general, a PCA implementation may involve only a few steps, starting by computing the matrix of variance-covariance or correlations. Then it proceeds by extracting the eigenvalues and their associated eigenvectors from the matrix of the cross-products of an original raw or transformed data set. The matrix of the cross-products may be obtained from an original raw or transformed data set. The variance-covariance matrix will contain elements of correlations when all the initial raw variables in the data set are standardized so they have means of 0.0 and variances of 1.0. Standardization may be unavoidable if the original variables are expressed in different, incompatible units. In a third *step*, we compute what is called principal component scores by projecting onto the principal components each sample or original observation. Principal components loadings are the elements of the eigenvectors that are used to compute the scores of observations and they are simply the coefficients of the linear equation which the eigenvector defines [4]. A final step in PCA implementation may involve the plotting and interpretation of 2D scatter diagram defined by each pair of the principal components. By cross-plotting, the variables and the samples are shown at positions representing, respectively, their loadings and scores on the principal components. The arc on the diagram is part of a circle representing a communality of 1.00. The communalities are the amount of variance of each variable retained in the principal components. If a variable falls on the circle, the two components account for all of its variability. Variables that plot inside the circle are characterized by variability that is not represented by the two principal components.

### 3. Results and Discussion

The matrix of pairwise correlations and the eigenvalues for the first eigenvectors are given in Table 2. The loadings of the eleven original variables on axis I are plotted against the loadings on axis II, in Figure 1. The samples are plotted on the score space defined also by the two first principal components (Figure 2). The samples are shown at positions corresponding to their scores on the first two axes. The first principal axis contains about 45.2 % of the total variance, whereas the second principal component represents an additional 24.7% (both correspond to almost 70.0 % of the total variance of the results from seepage water samples).

**Table 2.** Matrix of pairwise correlations and the eigenvalues for the first five eigenvectors.

	T	pH	$\sigma$	HCO <sub>3</sub> <sup>-</sup>	CO <sub>3</sub> <sup>2-</sup>	Cl <sup>-</sup>	NO <sub>3</sub> <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	Na <sup>+</sup>	K <sup>+</sup>	Ca <sup>2+</sup>	Mg <sup>2+</sup>
T	1.00											
pH	-0.35	1.00										
$\sigma$	-0.23	0.20	1.00									
HCO <sub>3</sub> <sup>-</sup>	-0.08	-0.62	0.07	1.00								
CO <sub>3</sub> <sup>2-</sup>	-0.26	0.58	0.24	0.05	1.00							
Cl <sup>-</sup>	-0.15	-0.40	0.32	0.90	0.22	1.00						
NO <sub>3</sub> <sup>-</sup>	0.16	-0.55	0.29	0.72	0.04	0.78	1.00					
SO <sub>4</sub> <sup>2-</sup>	-0.39	-0.01	0.65	0.56	0.41	0.67	0.48	1.00				
Na <sup>+</sup>	-0.14	-0.54	0.27	0.87	0.05	0.86	0.81	0.63	1.00			
K <sup>+</sup>	-0.10	-0.49	0.31	0.86	0.18	0.94	0.86	0.66	0.90	1.00		
Ca <sup>2+</sup>	-0.28	0.45	0.58	-0.25	0.07	-0.19	-0.22	0.28	-0.13	-0.23	1.00	
Mg <sup>2+</sup>	0.22	-0.56	-0.09	0.190	-0.47	0.12	0.42	-0.12	0.46	0.30	-0.24	1.00
Eigenvectors		I		II		III		IV		V		
Eigenvalues		5.422		2.960		1.284		0.837		0.599		
Percentage of total variance contributed by each eigenvalue												
		45.2 %		24.7 %		10.7 %		7.0 %		5.0 %		
Cumulative (%) value of total variance												
		45.2 %		69.90%		80.60%		87.60%		92.60%		



**Figure 1.** Seepage waters: plot of the first two principal components or principal axes. Horizontal axis is principal component I, vertical axis is principal component II: **(a)** Variables are shown at positions corresponding to their loadings on the first two axes (adapted from [2]); **(b)** Samples are shown at positions corresponding to their scores on the first two axes. In this plot, the numbers and the letters in the dot labels stand, respectively, for the year of sampling and the seasons: W, for winter; A, for autumn and S, for springtime e (rain is very rare in the summer in Lisbon).

$K^+$ ,  $Na^+$ ,  $Cl^-$ ,  $SO_4^{2-}$ ,  $HCO_3^-$ ,  $NO_3^-$  and pH are all well represented on the plan defined by the first two principal eigenvectors, with communality values varying, respectively, between 0.95 and 0.80 (Figure 1). On the other hand,  $Ca^{2+}$ ,  $CO_3^{2-}$ ,  $Mg^{2+}$ , T and specific conductivity ( $\sigma$ ) are not well represented on this plan as they plot far from the circle representing a communality of 1.00. These last variables involving also the third or fourth principal components seem thus to be less important to explain the overall variability in the seepage water samples.

The first principal eigenvector is strongly and positively correlated with  $K^+$ ,  $Na^+$ ,  $Cl^-$ ,  $HCO_3^-$  and  $NO_3^-$  and negatively correlated with pH. Samples projected on to the right side of the scores plot have values for these chemical species higher than their mean values. On the left side we find the samples with the highest values of pH. The second principal eigenvector is positively correlated with  $\sigma$ , pH,  $CO_3^{2-}$ ,  $Ca^{2+}$ ,  $SO_4^{2-}$  and negatively with T and  $Mg^{2+}$ . However, this axis does not clearly contribute for the analysis of samples position onto this scores plan of the two principal components.

Only  $K^+$ ,  $Na^+$ ,  $Cl^-$ ,  $HCO_3^-$  and  $NO_3^-$  form one cluster on this factor plan that is, in general, positively correlated with  $SO_4^{2-}$ . This seems to suggest the same source or process involving the strongly and positively correlated variables forming the cluster as well as a not very different source and alteration process or possibly a slight combination of other ones involving  $SO_4^{2-}$ . Together with pH these variables seem to play a significant role in the characterisation of seepage waters. They explain most of the variation observed in the chemical composition of the samples, while the other variables, including  $Ca^{2+}$  as well, do not. This surprisingly secondary role played by  $Ca^{2+}$  is possibly associated with stalactite formation observed inside the church. However, all the variables analysed do not provide enough discrimination of seepage waters to allow sub-classifications other than richer/poor samples in the content of these variables. Taking into account the sampling period, no time- or seasonal-dependent control of seepage water composition has appeared clearly from the analysis of the data. This could reflect a significant uniformity contribution of ion sources and stone alteration processes.

#### 4. Conclusions

The water-rock interaction and environmentally-induced processes, at “Basílica da Estrela”, seem to promote essentially the enrichment of seepage waters in  $K^+$ ,  $Na^+$ ,  $Cl^-$  and  $HCO_3^-$ . In this case, PCA has produced a result in general agreement with the one obtained by the analysis of the same data set performed by Figueiredo et al. [1-3] using a, perhaps, more classical geochemical approach [1-3]. However, it should be stressed that this kind of multivariate analysis may provide a basis for the management of environmental and stone decay data which may, when needed, be combined with other geochemical and petrophysical studies. For instance, PCA may help in the planning of future searching campaign for related studies. Hence PCA can help in non-destructive studies of

stone decay by the study of these samples that represent the product of interaction between pollutants and stones without sampling the cultural materials.

**Author Contributions:** Both authors contribute to the preparation of this publication.

**Funding:** The authors gratefully acknowledge the support of the CERENA strategic project FCT-UID/ECI/04028/2013) and the LAMPIST of the Instituto Superior Técnico, University of Lisbon as well the Lab2PT, Landscape, Heritage and Territory Laboratory(UID/ECI/04028/2013) of the University of Minho supported by the FCT – Fundação para a Ciência e Tecnologia (Portugal), with Portuguese funds and funds from the European Union (FEDER, Programa Operacional Factores de Competitividade – COMPETE 2020 – Programa Operacional Competitividade e Internacionalização, POCI, POCI-01-0145-FEDER-007528).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Figueiredo, C.A.M.; Marques, J.M.; Maurício, A.M.; Aires-Barros, L. Water-rock Interaction and Monuments Stone Decay: the case of Basílica da Estrela, Portugal. Proceedings of the 9th International Congresso on Deterioration and Conservation of Stone, Volume 1, Venice, Italy, June 2000; Edited by Vasco Fassina; Elsevier Science B.V., Amsterdam, The Netherlands. pp. 79-87.
2. Figueiredo, C.A.M.; Maurício, A.M.; Aires-Barros, L. A. Basílica da Estrela Stone Decay: the role of rain-water. Proceedings of the 10<sup>th</sup> International Symposium on Water-Rock Interaction, WRI-10, Volume 1, Villasimius, Italy, July 2001; Edited by Rosa Cidu, Swets & Zeitlinger, A.A.Balkema Publishers, Lisse, 2001. ISBN 90 2651 824 2. pp. 387-390.
3. Figueiredo, C.A.M.; Aires-Barros, L.; Basto, M.J.; Graça, R.C.; Maurício, A.M. The Weathering and Weatherability of Basílica da Estrela Stones, Lisbon, Portugal. In *Building Stone Decay: From Diagnosis to Conservation*. Přikryl, R & Smith, B.J. Eds; Geological Society, London, Special Publications, 2007, Volume 271, pp. 99-107.
4. Davis J.C. *Statistics and Data Analysis in Geology*. 2d ed.; John Wiley & Sons, Inc. New York, New York 10158, United States of America and Canada, 1986; pp. 646. ISBN 0-471-83743-1.
5. Moropoulou, A.; Polikreti, K. Principal Component Analysis in monument conservation: Three application examples. *Journal of Cultural Heritage*, 2009, Volume 10, pp.73-81. A.B.
6. Bacci, M.; Porcinai, S.; Radicati, B. Principal Component Analysis of Near-Infrared Spectra of Alteration Products in Calcareous Samples. *Applied Spectroscopy*, 1997, Volume 51, Issue 5, pp. 700-706.
7. Benavente, D.; Martínez-Martínez, J.; Cueto, N.; García-del-Cura, M.A. Salt weathering in dual-porosity building dolostones, *Engineering Geology*, 2007, Volume 94, pp. 215–226. doi:10.1016/j.enggeo.2007.08.003.
8. Maurício, A.M.; Aires-Barros, L.; Fassina, V.; Cassar, J.; Torpiano, A. Multivariate Data Analysis Applied to Salt Efflorescences Occuring at Sta. Marija Tà Cwerra Church (Malta). Proceedings of the European Commission Research Workshop on Origin, Mechanisms and Effects of Salts on Degradation of Monuments in Marine and Continental Environments, Protection and Conservation of the European Cultural Heritage, Research Report n° 4, Bari, Italy, March 1996; Edited by Fulvio Zezza. pp. 395-402.
9. Torfs, K.; Van Grieken, R.; Cassar, J. Environmental effects on deterioration of monuments: case study of the Church of Sta. Marija Tà Cwerra, Malta. Environmental effects on deterioration of monuments: case study of the Church of Sta. Marija Tà Cwerra, Malta. Proceedings of the European Commission Research Workshop on Origin, Mechanisms and Effects of Salts on Degradation of Monuments in Marine and Continental Environments, Protection and Conservation of the European Cultural Heritage, Research Report n° 4, Bari, Italy, March 1996; Edited by Fulvio Zezza. pp. 443-451.
10. Zezza, F.; Torfs, K.; Van Grieken, R.; García Pascua, N.; Macrì, F. Study of environmental effects on deterioration of monuments: case study of the Cathedral of Bari, Italy. Proceedings of the European Commission Research Workshop on Origin, Mechanisms and Effects of Salts on Degradation of Monuments in Marine and Continental Environments, Protection and Conservation of the European Cultural Heritage, Research Report n° 4, Bari, Italy, March 1996; Edited by Fulvio Zezza. pp. 461-471.

