Annals of Silvicultural Research
49(1), 2024: 13-23
https://creajournals.crea.gov.it/asr/

Research paper

Checking the stability of correlation of chronologies over time: an example on *Pinus pinea* L. rings widths

Sergio Camiz[1]*, Francesco Spada[2]

Received: 15/01/2023 Accepted: 2/10/2023 Published online: 16/10/2023

ABSTRACT An exploratory study concerning the variation over time of multidimensional time series allows to check to what extent factors and dendrograms, issued by ordination and classification methods, keep stable over time or change – even dramatically. In this paper, using five chronologies of *Pinus pinea* L. growth rings from literature, principal component analysis and hierarchical factor classification are applied on a ten years window, moving along time-series. These may be resumed through graphics showing the variation of the eigenvalues issued by the principal component analyses and of the correlations between time-series and principal components, through their corresponding time series, as well as through an animation and a compact representation of the time series of dendrograms. The results show that the studied period could be partitioned in seven intervals different in both correlations and groups structure, some of them highly stable: this suggests a second study, where two time intervals, identified as more homogeneous, showed really different structures.

KEYWORDS: Evolutionary Principal Component Analysis, Evolutionary Hierarchical Factor Classification, *Pinus pinea* L., chronologies.

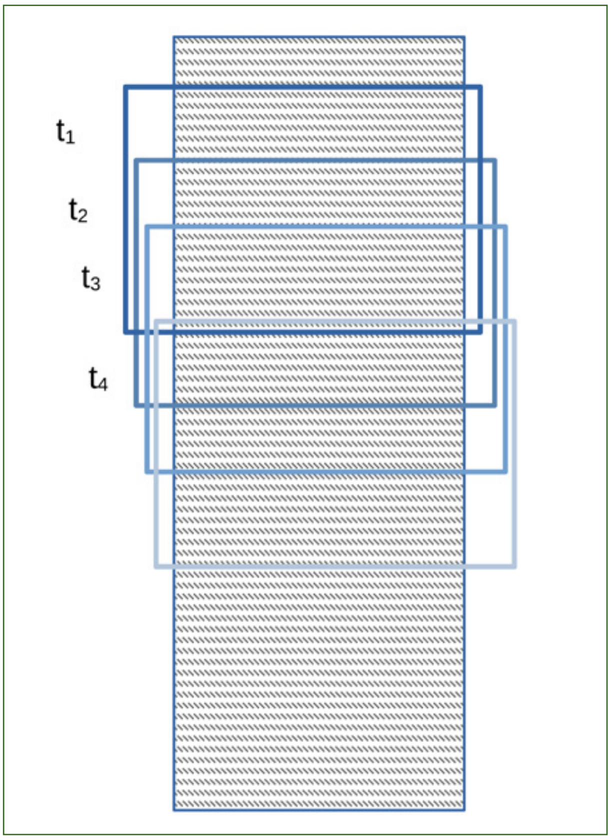
Introduction

The aim of this paper is to propose a method able to deepen the study of a multidimensional time series, taking into account the time itself, overcoming the limitations of static methods, such as Principal Component Analysis (*PCA*, Jolliffe 2002) and Hierarchical Factor Classification (*HFC*, Denimal 2001 and 2007) applied to the whole data table. In fact, both methods are based on the correlations between variables measured *statically*, because the units are supposed observed independently to each other at the same time during the data collection. On the opposite, while dealing with time series, observations are not simultaneous but taken at pre-definite time intervals, so that the computed correlation is a kind of average measure of total association that may vary over time, specifically when the observed variables depend on non-constant exogenous factors. Should this be the case, there is no reason why mutual relations - correlation in particular - could remain the same in different time intervals, hence the need to study their variation over time too. Indeed, both methods may be run on a mobile window (Fig. 1) – that is a small sub-table involving the same time series but limited to a fixed small number of consecutive dates only - shifted one date at a time, throughout the whole studied period.

Running on a mobile window, the quoted methods produce a time series of results, whose study may inform on the variation over time of the relations revealed between the time series at local level, something impossible to be detected by an overall static study, where time has no role and the local variation remains undetected.

A joint static use of both methods was proposed and discussed by Camiz et al. (2020) in view of a better understanding of the structure of five synchronic *Pinus pinea L.* chronologies, built and studied by Piraino et al. (2013).

Figure 1 - The principle of mobile window: given a data set with dates by rows and synchronous time series by column, a time interval, subset of a given number of adjacent dates, is chosen – the window – and it is shifted one date at a time through the entire table.



1 - CNR – ISPC – Italy
2 - University of Rome, La Sapienza - Italy
*Corresponding author: sergio@camiz.it

http://dx.doi.org/10.12899/asr-2455

While these authors were assessing the response to climatic variability of the radial growth of *P. pinea* in Central Italy, the first quoted paper focused on the utility of these exploratory methods, to ascertain homogeneity in the chronologies i.e., the opportunity to extract a chronology common to all. This way, to what extent such results could be considered stable over time remained unknown, hence the need of a dynamic study that may enlighten about the relations changes, probably issued by the local variation over time of exogenous factors influencing the trees growth. For this reason, in this paper we study how both *PCA* and *HFC* may be used dynamically, i.e., providing time-dependent results.

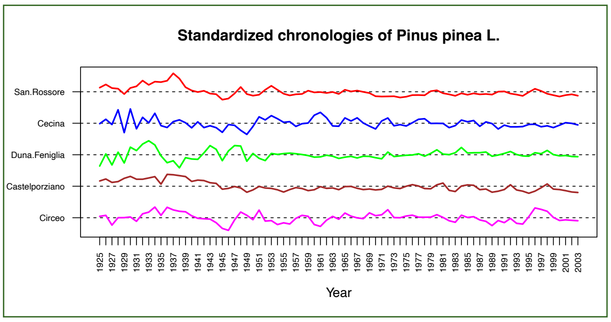
This is what has been proposed for *PCA* in both dendrochronological (Camiz et al. 2010, Camiz and Roig 2011) and economic studies (Camiz and Diblasi 2013) under the name of Evolutionary Principal Component Analysis (*EPCA*). Note that the terms "evolution" and "evolutionary" introduced there only refer to study of the change over time of time series, without any reference with their use in natural sciences: we shall use them in the following with this meaning. In this paper, we parallel *EPCA* with the Evolutionary *HFC* (*EHFC*), both applied to mobile windows along the said *P. pinea* chronologies. A mayor synthesis of the – usually most cumbersome – results is proposed, yet able to suggest interpretations and directions of further deepening of the investigation. In particular, a method for partitioning the time series in more homogeneous time intervals is also proposed, where the same static methods may provide different, while more homogeneous, results.

The chronologies

This work logically continues the study carried out from Camiz et al. (2020), thus, we deal with the same 79-years long multidimensional time series, composed by five synchronous chronologies of tree-ring widths of *P. pinea*, ranging from 1925 to 2003. They were built by Piraino (Piraino et al. 2013) selecting sites along the Tyrrhenian coast of the Italian peninsula: *San Rossore, Cecina*, and *Duna Feniglia* in Tuscany, and *Castelporziano* and *Circeo* in Latium. These sites are aligned NW–SE and are located close to the coastline of the Tyrrhenian Sea, between 43° 43' and 41° 18' North, their farthest distance being 350km approximately. All stands have been planted on sandy dunes: the populations of *San Rossore, Cecina*, and *Circeo* originate from plantations carried out during the first half of the 20th century, while that of *Duna Feniglia* is some decades older and at *Castelporziano* the pine stands date back to 18-19th century (see Piraino et al. 2013, Camiz et al. 2020, for further details). These forests grow under Mediterranean climatic conditions, locally characterized by summer drought, ranging from one to three months.

In Figure 2, the patterns of the five chronologies along their 79-years long time-span are represented, their vertical order corresponding to their geographical NW-SE alignment. Note that, to keep all chronologies comparable, they have been standardized, and so will be the issued results. Thus, all have zero mean, unit variance, and no physical unit of measure. Indeed, no loss of information occurs since a simple transformation may restore the original values.

Figure 2 - The five standardized tree-ring widths chronologies of *P. pinea* of Central Italy in this study. The top-down order mirrors the geographical NW-SE alignment of the sites.



In Camiz et al. (2020) three distinct groups of chronologies were identified: 1) *Cecina*, 2) *Duna Feniglia*, and 3) a group encompassing the populations of *San Rossore, Castelporziano,* and *Circeo*. Such partition had been revealed mainly by *HFC* results, where too high differences between the highest nodes prevented further aggregations. The distinction between the first two nodes from the third group could be interpreted on the basis of their slightly different environmental scenarios, considering that both stands of *Cecina* and *Duna Feniglia* were established on highly dynamic sites of littoral sand dunes, while the other stands were planted on inland flattened fossil dunes. In addition, the stand of *Duna Feniglia* grows on a narrow sandy strip of dunes separating the sea from a lagoon: thus, it is affected on both sides by salt water and dominating high winds, both apparently inducing higher environmental stress on the trees.

Theoretical background

For the rationale of both *PCA* (Jolliffe 2002) and *HFC* (Denimal 2001 and 2007) we refer to both literature and Camiz et al. (2020): here we only remind their essentials of specific interest for the evolutionary studies.

While *PCA* partitions the total data information into uncorrelated components of decreasing relevance, *HFC* builds a hierarchy of partitions of the characters. Moreover, for each created group, *HFC* builds a representative character, that plays a role analogous to the first principal component of *PCA*, i.e., it gathers most of the information common to all groups' characters. This is quite interesting when dealing with dendrochronologies: just as *PCA* factors, that – when positively correlated with the original series – may be taken as estimates of common principal component chronologies sensu Peters et al. (1981), the *HFC* representative characters of the groups – in the same conditions – may be representative chronologies of the time series gathered in each group.

Inertia

In exploratory data analysis, the *inertia* (Lebart et al. 2006, Aluja-Banet et al. 2018, Camiz 2021) is a key concept to measure the information contained in the data, since partitioning a table's inertia is a task common to multidimensional exploratory techniques. Dealing with a quantitative character, its variance informs about the departure of the observed values from their mean: when observations are represented as points along a line according to their value, the variance is a measure of information about the points scattering, hence the character variability. This may be generalized to a quantitative data table: when observations form a cloud in a multidimensional space, whose dimensions correspond to the considered characters, their scattering may be measured with respect to their centroid – a point whose coordinates correspond to the averages of all observed characters. Thus, their dispersion may be measured by their inertia, the weighed sum of squared distances of the points to the cloud's centroid. Inertia is the key measure of information in exploratory data analysis, since most results are provided associated to its amount, they explain. For its computation, given a quantitative data table X, with n observations by rows and p characters by columns, each observation is provided by a weight w_i, i=1,...,n such that, $\sum_{i=1}^{n} w_i = 1$ indicating its relative importance. When observations are weighted, mean and variance of any character j=1,...,p are given by $\bar{x}_j = \sum_{i=1}^{n} w_i x_{ij}$ and, $\sum_{i=1}^{n} w_i (x_{ij} - \bar{x}_j)^2$ respectively, and the cloud's centroid is the point whose coordinates are the means $G = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)$. Therefore, the cloud's inertia is given by the sum of the characters' variances:

$$Inertia(X) = \sum_{i=1}^{n} w_i d^2(x_i, G) = \sum_{j=1}^{p} \sum_{i=1}^{n} w_i (x_{ij} - \bar{x}_j)^2 = \sum_{j=1}^{p} var(x_j).$$

Now, the considered analyses partition the total inertia according to either uncorrelated components – as in *PCA* – or separate groups – as in *HFC*; therefore, the share of inertia attributed to each of these items is a measure of the amount of information that they are attributed, hence, of their relevance. To avoid that characters' different means and variances could bias the results, in the following they will be standardized, i.e., transformed to have zero mean and unit variance: thus, our table's inertia will be worth p, the number of characters.

Mobile window

The idea of evolutionary analysis is quite simple: define a mobile window of a given number m of dates and re-define it iteratively, each window resulting from the previous one by withdrawing the first date and adding that immediately following the m-th. This way, if n is the number of the considered dates, $n-m+1$ tables may be built, each differing from the two adjacent ones by one date only. Submitted to the same analyses, each window may give slightly different results, this way describing the change over time of the data structure.

In particular, it must be observed that the difference in structure between two adjacent windows would depend on the exchange between two dates only: the first, dropped, and the last, included. Therefore, the choice of the mobile windows length m should be defined considering the following points: *i)* the length of the time series of results is reduced to $n-m+1$ dates, usually labelled by that in the middle of the window; *ii)* the average contribution of each date to any analysis would be worth $1/m$; in particular, this would be the effect of the substitution of one unit with another, hence the difference between two adjacent windows, the other observations keeping the same; *iii)* if the window is short – i.e., m is small – the resulting variation has a heavy weight and it depends on two relatively close units, thus their interpretation may be easy, but local noise may be overweighted; *iv)* if the window is long – i.e., m is large – the difference could be much higher, albeit its weight would be much lower; in this case, since the units would be far away, the meaning of their differences may be more difficult to understand, whereas local noise may become irrelevant; *v)* the number of dates should be much larger than p, the number of characters, to ensure a sufficient inertia to be taken into account; and *vi)* as the time series may have an intrinsic periodicity (daily, weekly, moon phase, yearly), it would be wise to choose a multiple of such period (24 hours, 7 days, etc.), if any, so that all periodic variations would be taken into account at least once in each window.

To reduce the influence of such replacement and to involve all observations in the windows variation, we decided to give each observation a different weight, according to its position within the window. This way, not only the substitution of the two extreme points would contribute to the structure differences, but also the variation of position of the observations within the window. Moreover, giving higher values to the central windows' dates and lower to their extremes would smooth sufficiently well the differences over time. Indeed, any date will be little weighed while entering the mobile window, its weight progressively rising until the window central date, then lowering until exit. As a result, the central variation within the window will be given more importance.

Some experiments carried out by Camiz and Diblasi (2013) showed a progressive smoothing of the results while raising weights differences. In this work, not having an intrinsic periodicity as reference, we chose a window length of m = 10 years, twice the number of the time series. A kind of Gaussian distribution was chosen for the weights: we started from the set of numbers 1, 2, . . . , m, whose mean equals $\mu = \frac{m+1}{2}$ and the standard deviation $\sigma = \frac{m-1}{\sqrt{2}}$ Then, given a parameter k, for i = $1, \ldots, m$, the values $v_i = e^{\frac{1}{2}(\frac{(i-\mu)}{\sigma}k)^2}$ were computed according to the Gaussian function, and re-scaled to get the weights , $w_i = \frac{v_i}{\sum v_i}$ all positive and summing up to 1. After some experimentation, k = 3 resulted good for an acceptable smoothing. As said, weights were kept constant along the $n-m+1$ = 70 windows.

Principal Component Analysis

In an exploratory framework, it is not strictly necessary to get a stopping rule to identify the significant dimensions, because each method's results may be inspected to the end to discover the deepest details: nevertheless, based on Camiz et al. (2020) results, we took into account the first three dimensions issued by the *PCA*s, as all corresponding eigenvectors were larger than 0.7, the lower eigenvalue threshold for relevant dimensions recommended by Jolliffe (2002). Indeed, it could be expected that in a smaller window the relevant information could be largely covered by a smaller number of factors, but we kept this same level to compare the tentative dimension of the windows with respect to the overall analysis.

Hierarchical Factor Classification

Unlike *PCA*, where only the first principal component may be seen as a chronology when all time-series correlations have the same sign, in the case of *HFC*, when concordance of all correlations signs occur, the first representative series of any group may be seen as its representative chronology. In any case, their interpretation may only be based on the chronologies forming the group. On the other side, the representative chronologies of different groups are not necessarily orthogonal, so that their correlation may be an indication of their affinity. It is worthy to observe that they are better situated and interpretable than rotated and oblique principal components, sometimes preferred to classical *PCA* in dendrochronological studies (Frank and Esper 2005, Büntgen et al. 2007, Leland et al. 2013) even for classification purposes.

As for the number of groups to take into account, the hierarchy index, measuring differences between the gathered series, is a good indicator of the uni-dimensionality of a formed group. Indeed, we decided to merge nodes as far as the index keeps lower than 0.7, as Jolliffe (2002) suggests for *PCA*, to prevent gathering with other poorly correlated series.

Evolutionary analyses

The results issued by the evolutionary analyses correspond to those of the static ones, but multiplied by the number of windows: therefore, a synthesis is necessary to get possible their interpretation and we considered to organize them in form of time series, to be understood once graphically represented together with some summary statistics.

Evolutionary *PCA* (*EPCA*) results have been first summarized by the time series of the first three eigenvalues of each window: such graphical representation, including the sums of the eigenvalues too, provides a synthetic information about the variations of the considered dimensions relative importance. In particular, the pattern of the first eigenvalues shows to what extent the first principal component may be representative of the whole time series in the windows over time, or the importance of the following ones in evidencing other relevant components. Then, for each relevant principal component, the time series of the correlations of every time series with them was assembled: these graphics show whether these correlations keep constant over time or not, for their variation would induce a different interpretation of the principal components in different periods.

Note that, since the principal components are issued by the analysis up to the sign, the variation in sign of the correlations may be due to the *PCA* algorithm only. Therefore, to keep a coherence between correlations in two adjacent windows, the sign of all correlations with a component in the following one was chosen to minimise the sum of the squared differences between the corresponding ones in the first window. Nevertheless, such optimization may not always provide the expected optima results.

For evolutionary *HFC* (*EHFC*), the time series of the inertias associated with the chosen nodes and those of their representative characters should be considered together with their correlations with the gathered characters. Moreover, the variation of dendrograms' topology must be taken into account: this is not easy, since their comparison may be done either through an animation where the dendrograms are represented over time, or writing them in Newick format (Felsenstein 2021, Wikipedia contributors 2021), that is, as a string in which each pair of gathering nodes is enclosed within parentheses. Indeed, to ease the comparison between strings, the leaves must be rearranged by hand to maximally show their similarity. Here, as measures of similarity/dissimilarity between dendrograms, among the many available, two Robinson-Foulds distances (Robinson and Foulds 1979 and 1981) seemed the most suitable for the purpose: they differ in that the latter limits attention to the dendrograms topology – i.e. how the branches are connected to each other – whereas the former takes into account the length of the branches too. The weighted distance between two dendrograms is computed by identifying the sub-trees common to both and consequently summing the differences of lengths of the common branches and adding the lengths of the branches of the remaining non-common parts of both trees. The unweighted distance is computed by considering all branches having the same 1-length.

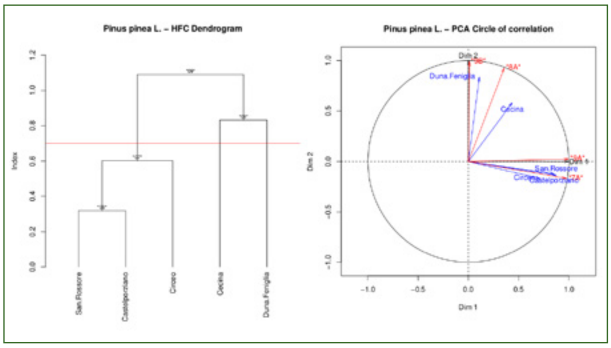
Findings periods

Based on the results issued by the evolutionary study, it may appear reasonable to divide the multidimensional time series at hand into time-slots, whose data structure may appear sufficiently homogeneous. For such task, the time series first eigenvalues issued by either *EPCA* or *EHFC* seemed the best choice. To cut it, the Fisher (1958) algorithm was applied, that identifies cut-points of the series by minimising the inertia within the resulting intervals. Such algorithm was run iteratively raising the number of sought intervals, providing progressively decreasing pooled inertia. To select the best partition, the Caliński and Harabász (1974) F-statistics – with n the series length and k the number of intervals – based on the ratio of mean between (*BSS*) and within (*WSS*) intervals inertia was adopted, searching for local maxima between 2 and 10 intervals.

Results

In Figure 3 the main results of the static analyses applied to the whole data table of *P. pinea* chronologies (Camiz et al. 2020) are reminded: the dendrogram on the left is issued by *HFC* and the circle of correlations on the plane spanned by the first two factorial axes on the right comes from *PCA*.

Figure 3 - Static analyses of the 5 chronologies of *P. pinea*. Left: the dendrogram issued by the hierarchical factor classification. Right: the chronologies (in blue) and the representative variables of the first two nodes and of the group of three (in red) on the circle of correlations on the first factorial plane of principal component analysis.



The dendrogram topology in Newick format is ((1,4),5),(2,3)) and, according to the pattern of fusion levels, a partition in three groups was proposed, with both *Cecina* and *Duna Feniglia* standing alone, and the three remaining sites – *San Rossore, Castelporziano,* and *Circeo* – sharing a common chronology. In the circle of correlation they are close to the first horizontal axis, while the two others are more oriented towards the second, albeit opposed along the third (non-represented) one.

Evolutionary principal component analysis

The evolution of the correlation structure of the five chronologies over time was found by running *EPCA* throughout a ten years mobile window, using symmetric weights obtained considering values up to 3 standard deviations of a standardized normal distribution. Given the length of the window, the time series of the results spans from 1930 to 1999, 70 years.

In Table 1 are reported the main statistics concerning the first three eigenvalues of the 70 *PCA*s performed on the mobile window and of their sums. It is noticeable the higher means of the first two eigenvalues with respect to those issued by the static *PCA*: indeed, this may result from a higher coherence in a short time interval than in the whole period. The first eigenvalue ranges between 1.85 and 3.73 with mean 2.57: therefore, it keeps summarizing at least 37% of the windows inertia, with a maximum of 75%, and averaging over 50%. Compared to it, the following eigenvalues variability is much larger, as shown by their range and coefficients of variation: the second one may lose its interest in some period, given its minimum under 0.7, and the same occurs for the third.

On the opposite, the sums of the first two and three eigenvalues are much more regular, the third in particular, with a very low coefficient of variation and a reduced range. Thus, most of the dishomogeneity of some periods, indicated by low inertias along the first factor, is absorbed by the following ones, given that the inertia cumulated by the three factors is always over 83%. The second eigenvalue ranges between 0.55 and 2.0, averaging 1.31: considering that only in 1940-41 it worths less than 0.7, the threshold considered for a dimension relevance, it keeps nearly always meaningful; as for the third, it is negligible to and fro, indicating that in most years three dimensions might be meaningful.

The Fisher (1958) algorithm was applied to the time series of the first eigenvalues, searching the optimum partition between those into 1 up to 12 groups. A maximum of Caliński and Harabász (1974) statistics resulted for 7 groups: 1) 1930-38, 2) 1939-43, 3) 1944-46, 4) 1947-64, 5) 1965-1981, 6) 1982-87, and 7) 1988-99.

In Figure 4 the evolution of the first three eigenvalues is shown over time, together with the evolution of the sum

Figure 4 - The time series of the first three eigenvalues issued by the *EPCA* of the five chronologies of *P. pinea* (1 = blue, 2 = red, 3 = green) and of the sums of the first two (black) and three (grey) eigenvalues, respectively. The eigenvalues are issued from the *PCA*s performed on a 10-years mobile window. Horizontal dotted lines: series average, vertical: cut-points.

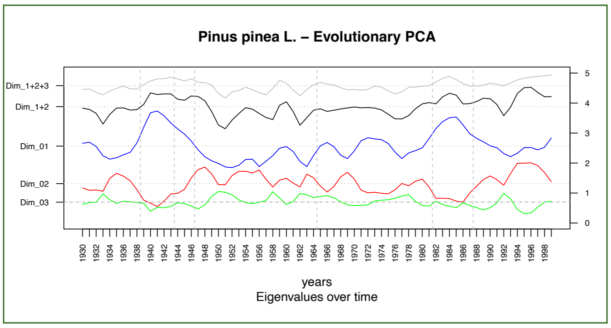


Table 1 - *EPCA* of the five chronologies of *P. pinea*. Statistics of the first three eigenvalues of the *PCA*s performed on a 10-years mobile window. The first row reports the eigenvalues of the static *PCA* performed on the whole period.

	Eigen 1	Eigen 2	Eigen 3	Eig 1+2	Eig 1+2+3
Static eigenvalue	2.191	1.103	0.839	3.294	4.123
Minimum	1.849	0.549	0.315	3.143	4.169
Maximum	3.734	2.007	1.051	4.525	4.938
Mean	2.566	1.314	0.697	3.881	4.578
Standard deviation	0.459	0.349	0.168	0.307	0.180
Coefficient of variation	0.179	0.265	0.241	0.079	0.039

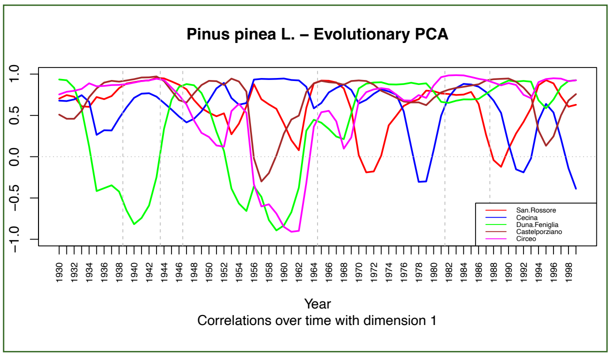
of the first two and three. There, the horizontal lines correspond to each series average and the vertical dotted ones represent the 6 chosen cut-points. Note that the mean of the third eigenvector is nearly 0.7, thus this (dashed) line may be adopted as a threshold for all dimensions relevance.

Looking at the pattern of the first eigenvalue in Figure 4 (in blue), it results that in the periods 2), 3), and 6) this is always larger than its mean (2.57), whereas in 1) and 4) it is always smaller and in the remaining 5) and 7) it is basically around the mean. Thus, in the first group of periods, a stronger agreement between series would be expected, much lower in the others. In particular, the periods 2) and 6) may be taken as nearly uni-dimensional, given that both eigenvalues 2 and 3 are close to 0.7. Looking at the pattern of the second eigenvalue (in red) it is evident a behaviour opposite to that of the first one, confirmed by their strong negative correlation (-.744): indeed in the periods with lower homogeneity, the second eigenvalue is stronger, maybe strengthening its meaning. Unlike periods 2) and 6), in the third one, albeit the first eigenvalue is above the mean, the second one is above 1: this may be interpreted as a period of higher coherence but with a stronger second dimension with respect to the others. A similar effect occurs for the third eigenvalue (in green), and indeed the sums of the first two (in black) or three (in grey) show a much more smoothed pattern, indicating a stability of the three-dimensional solution over time. Therefore, summarizing, it results that the periods 2) and 6) are highly homogeneous, hence uni-dimensional, the 4) and 5) three-dimensional, given the higher third eigenvalues, and the remaining two-dimensional.

For a better understanding of the structure of correlation in the various periods, the exam of the correlations with the principal components over time is necessary: in Table 2 minimum, maximum, mean and standard deviation of the correlations of the time series with the first two principal components over time are reported with the correlations based on the whole time interval.

Concerning the first, the mean correlations are lower than the static ones for the three most correlated sites, whereas they are higher for *Cecina* and *Duna Feniglia*; the maxima are very high for all, and all minima are negative, with extreme values for *Duna Feniglia* and *Circeo*. As for the second component, the high standard deviation and the very large range in all series is a sign of high confusion and consequently difficult interpretation. Even worst seems the third component, which will be not commented here.

Figure 5 - The time series of the correlations of the five chronologies of *P. pinea* with the first principal components issued by their *EPCA*. The principal components are issued by the *PCAs* performed on a 10-years mobile window.



The inspection of Figure 5 helps to better understand what happens: there the time series of the correlations of the five chronologies with the first principal component in all windows are shown, together with the found cut-points. It is easy to observe the different behaviour of the series in the seven periods. With respect to the highest positive correlations, it results that *Duna Feniglia* is negatively correlated in most part of periods 1), 2), and 4), whereas *Circeo* is negatively correlated in the second half of 4) only. Moreover, occasional negative correlations occur for *Castelporziano* in 4) and of *San Rossore* and *Cecina* in 5) and 7), although lower correlations may be found

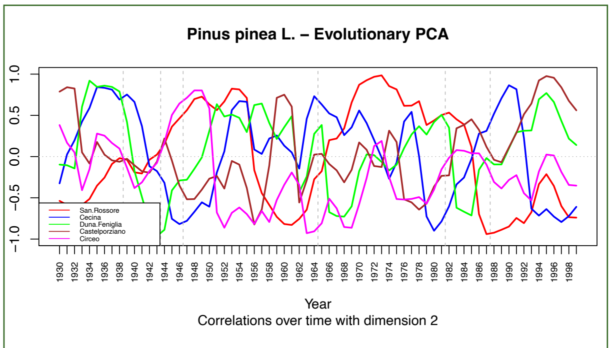
Table 2 - *EPCA* of the five chronologies of *P. pinea*. Statistics of the correlations of the five chronologies with the first two principal components issued by the *PCAs* performed on a 10-years mobile window. The rows Static report the correlations computed on the whole period for comparison.

	San Rossore	Cecina	Duna Feniglia	Castelporziano	Circeo
1st principal component					
Static correlation	0.873	0.436	0.109	0.850	0.710
Minimum	-0.190	-0.388	-0.893	-0.300	-0.909
Maximum	0.954	0.946	0.935	0.972	0.987
Mean	0.605	0.574	0.358	0.716	0.583
Standard deviation	0.294	0.333	0.603	0.281	0.499
2nd principal component					
Static correlation	-0.137	0.581	0.838	-0.123	-0.169
Minimum	-0.940	-0.899	-0.964	-0.823	-0.929
Maximum	0.985	0.865	0.921	0.976	0.802
Mean	-0.003	0.057	0.094	0.047	-0.225
Standard deviation	0.601	0.533	0.478	0.442	0.425

around the negative pitches. From the graphics it is evident that a high homogeneity may be found only in the periods 3) and 6) – which may be deemed to be uni-dimensional – whereas in 2) *Duna Feniglia* sets definitely apart. Following the time flow, some negative correlations are noteworthy: 1935- 43 and 1952-62: *Duna Feniglia*, 1956-63: *Circeo*, 1956-58: *Castelporziano*, 1970- 73: *San Rossore*, 1977-80: *Cecina*, 1988-89: *San Rossore*, 1990-93 and 1998-99 *Cecina*. Therefore, it is possible to argue that *San Rossore* is strongly correlated with the first principal components, but shows anomalies in the period 1952-63 and two more relevant in 1970-73 and 1988-89; *Cecina* keeps medium-high correlations until 1976 and then it highly fluctuates; *Duna Feniglia* shows a pattern opposite to the others until 1970; *Castelporziano* shows a fluctuation during 1956-63; and *Circeo* is unstable between 1948 and 1970 with a large fluctuation in 1955-64. Summarizing, after a coherence of the chronologies, with only *Duna Feniglia* in clear opposition, during three periods, the fourth one is characterized by a heavy instability, followed by a relative stability, with asynchronous fluctuations of *San Rossore* and *Cecina*.

It is worth to observe that, between 1930 and 1955, to the first axis all series were positively correlated, but *Duna Feniglia*: therefore, the first component meaning may be taken as constant in that period. On the opposite, between 1970 and 1994, *Circeo, Duna Feniglia*, and *Castelporziano* showed regular high correlations: thus, the first component meaning would be different from the first one. Note also the irregularity between 1956 and 1973, where only *Cecina* resulted constantly positively correlated, and the final years where *San Rossore* took the place of *Castelporziano* in describing the first component.

Figure 6 - The time series of the correlations of the five chronologies of *P. pinea* with the second principal components issued by their *EPCA*. The principal components are issued by the *PCAs* performed on a 10-years mobile window.



In Figure 6 the pattern of the time-series representing the evolution of the correlations of the chronologies with the second principal component of the *EPCA* is reported: the resulting pattern is really complicated, all series keeping fluctuating correlations with alternate signs. Therefore, a consistent interpretation may be limited to say that a second dimension is nearly always present, but that its meaning varies continuously. As for the third dimension (not shown) the complicated pattern here too prevents any possible interpretation.

Evolutionary hierarchical factor classification

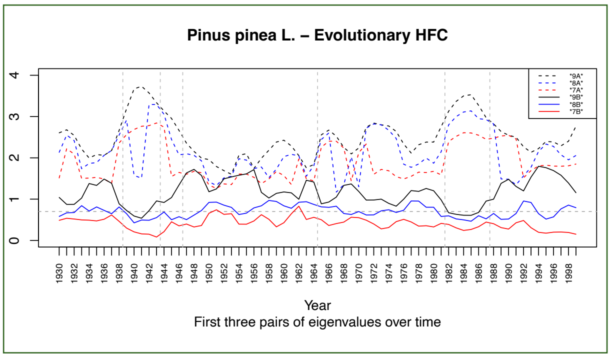
From *EHFC* – run with the same choices of *EPCA* – more important information about the dendrograms' variation results, but more difficult to synthesize, unless sequentially arranged in an animation, which is visible as supplementary file (https://journals-crea.4science.it/index.php/asr/article/view/2455/47) .

Table 3 - *EHFC* of the five chronologies of *P. pinea*. Statistics of the inertias of the representative chronologies of the four nodes of the hierarchies, performed on a 10-years mobile window. The first row reports the corresponding inertias of the static *HFC* performed on the whole period.

	9A	*9B*	*8A*	*8B*
Static eigenvalue	2.157	1.090	1.169	0.831
Minimum	1.602	0.538	1.171	0.426
Maximum	3.733	1.796	3.293	0.968
Mean	2.503	1.166	2.125	0.708
Standard deviation	0.500	0.320	0.541	0.146
Coefficient of variation	0.200	0.275	0.255	0.207
	7A	***7B***	***6A***	***6B***
Static eigenvalue	2.079	0.602	1.681	0.319
Minimum	1.258	0.082	1.576	0.044
Maximum	2.844	0.829	1.956	0.424
Mean	1.901	0.406	1.783	0.217
Standard deviation	0.462	0.151	0.102	0.102
Coefficient of variation	0.243	0.371	0.057	0.471

In Table 3 are reported some statistics concerning the pairs of eigenvalues associated with the *HFC*s nodes, compared to those issued by the static *HFC* during all period. It must be reminded that, over time, they refer to different nodes structures, depending on the local clustering process. Therefore, apart *9*, the other figures should be considered with care. The mean of evolutionary *9A* is higher than the static value, a sign of higher homogeneity in the various dates than in the whole period. The maxima of the second eigenvalues, all above 0.7 but *6B*, reflect the multidimensionality of the corresponding nodes, at least in some periods. On the opposite, given the very low maximum of *6B*, the node *6* always reflects a common chronology, but corresponding to different groups of sites,

Figure 7 - The time series of the first and second eigenvalues of the three upper nodes of the *EHFC* of the five chronologies of *P. pinea* performed on a 10-years mobile window. The vertical dashed lines in grey represent the seven cut-points of the first eigenvalues of the *EPCA*; the horizontal dotted grey line corresponds to 0.7, the threshold to decide the smallest partition.

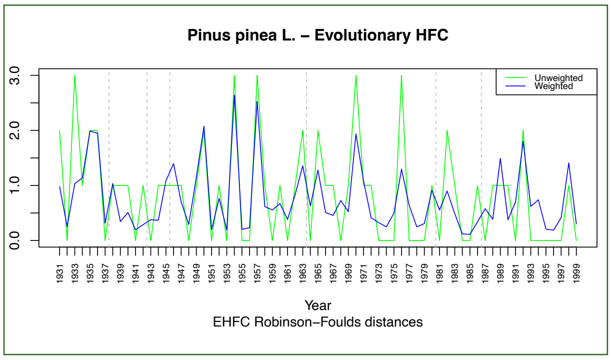


depending on the dates. This may be appreciated by observing the animation.

In Figure 7, the pattern of the pairs of eigenvalues of each hierarchy over time is represented; the horizontal dashed line corresponds to 0.7, the value we adopted as threshold to decide the partitions; in addition, the same cut-points found from *EPCA* are reported as vertical dashed lines. Here, the three upper (dotted) lines represent the first eigenvalues *9A*, *8A*, and *7A* of the first three nodes of the hierarchies, while the three (continuous) lower (*9B*, *8B*, and *7B*) correspond to the second ones. Opposite patterns may be observed between each pair of eigenvalues of the same nodes, reflecting the higher or lower concentration of inertia on the first dimension over time: the higher this is, the lower is the difference between the chronologies in the node.

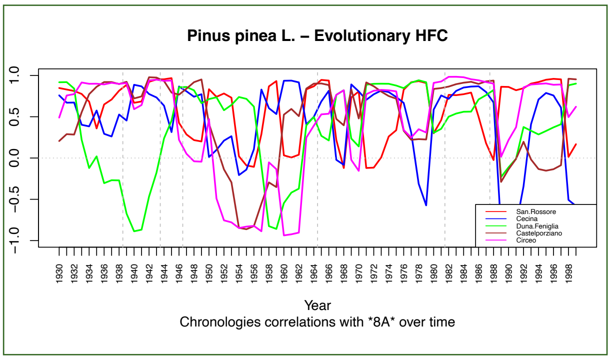
Note also the similarity between the pattern of the *9A* with that of the first eigenvalues of *EPCA* (as well as that of the *9B* with the second eigenvalues): this shows the near optimality of *HFC* at the upper level. As for the patterns of *8A* and *7A*, it should be borne in mind that their different composition over time prevents interpretations. More interesting is the pattern of the fusion levels, say of *9B*, *8B*, and *7B*, given that on them depend the dendrograms' cuts: it may be observed that in 1939-1942 and 1982-86 *09B* is constantly lower than 0.7, thus, during those periods the homogeneity is maximum and one may consider all chronologies forming a single group – or better a dipole in 1939-42, since *Duna Feniglia* behaviour is opposite to the others. *8B* is lower in 1939-49, 1981-91 and occasionally in other dates: therefore, in these periods two groups are supposed to be a consistent classification, whereas in the rest three groups might be considered, given that four groups might result only in 1951 and 1962, the two dates in which *07B* is little higher than 0.7.

Figure 8 - *EHFC* of the five chronologies of *P. pinea* performed on a 10-years mobile window: Robinson-Foulds distances of each dendrogram to the previous. In green: unweighted distance; in blue: weighted distance. The dashed vertical lines are the cut-points of the sequence of first eigenvalues of *EPCA*.



In Figure 8 the two Robinson-Foulds distances between each indicated date and the previous one are reported: considering first the trees topology through the unweighted distance (in green in the figure), the pairs of adjacent identical topologies are frequent (30), but not continuous: only three adjacent dates result in 1954-56 and 1983-85, four in 1972-75 and 1976-79, and six just in 1992-1997. Other 27 pairs differ for one branch only: note in particular the period 1936-48 in which the topologies are either equal or have distance 1. Then, 8 have distance 2 and 5 have 3, the maximum. Instead, the weighted distances are never zero, due to the varying branches length, and raise above 1 for 17 pairs of adjacent dates only. Comparing these patterns with the cut-points, we may notice that in periods 2, 3, and 6 only the weighed distance keeps constantly under 1: nevertheless, this does not mean a constant topology, as already underlined.

Figure 9 - The time series of the correlations of the five chronologies of *P. pinea* with the representative variable of node *08* issued by their *EHFC*. These variables are issued by the *HFC*s performed on a 10-years mobile window.



The patterns over time of the correlations of the five chronologies with both representative and differences chronologies of node *9* in each period (not shown) are nearly equal to those of the first two *EPCA* components, therefore no particular remark is needed. Of more interest is the graphic concerning the correlations with the representative variable *8A*, shown in Figure 9, whose variations are depicted by the node's composition over time. There, in the first three time intervals from 1930 to 1946 and the two from 1965 to 1987 all chronologies tend to be positively correlated with it – even if not always significantly, but *Duna Feniglia* in the first and occasionally others in the last ones. This would reflect a change of meaning of the second node, which in the first period isolates *Duna Feniglia*, that in the others keeps positively correlated with most chronologies. On the other hand, in the remaining intervals, the structure of correlation varies too much to be fully interpreted: it may be better understood by looking at the variation of the dendrograms over time.

In Table 4, the topological structure of the dendrogram issued by *EHFC* of every window is shown in Newick format: each node is represented by a pair of parentheses, enclosing the pair of merged nodes. Thus, e.g. in window centred in 1930, the nodes *(San Rossore, Duna Feniglia)* and *(Castelporziano, Circeo)* are first created, then *Cecina* merges with the first, giving *((San Rossore, Duna Feniglia, Cecina)*, and eventually this merges with *(Castelporziano, Circeo)*. Indeed, this topological notation does not show which pair has been created first: the incorporation of the

Table 4 - The topological structure of the 70 dendrograms built by the EHFC of the five chronologies in Newick format (Felsenstein 2021, Wikipedia contributors 2021). Each group within parentheses corresponds to a node, regardless of the fusion level and therefore of a tentative partition. The cut-points are indicated by the bold horizontal lines.

Year	Topology	Year	Topology
1930	((Castelporziano,Circeo),(Cecina,(San Rossore,Duna Feniglia)))	1965	((((San Rossore,Castelporziano),Cecina),Circeo),Duna Feniglia)
1931	(Castelporziano,(Cecina,((San Rossore,Duna Feniglia),Circeo)))	1966	(((Castelporziano,(San Rossore,Cecina)),Circeo),Duna Feniglia)
1932	(Castelporziano,(Cecina,(Circeo,(San Rossore,Duna Feniglia))))	1967	((Castelporziano,(San Rossore,Cecina)),(Circeo,Duna Feniglia))
1933	(((San Rossore,Circeo),Castelporziano),(Cecina,Duna Feniglia))	1968	((Castelporziano,(San Rossore,Cecina)),(Duna Feniglia,Circeo))
1934	((San Rossore,(Castelporziano,Circeo)),(Cecina,Duna Feniglia))	1969	((San Rossore,(Cecina,Castelporziano)),(Duna Feniglia,Circeo))
1935	((San Rossore,Duna Feniglia),(Cecina,(Castelporziano,Circeo)))	1970	((San Rossore,Cecina),(Circeo,(Duna Feniglia,Castelporziano)))
1936	((San Rossore,(Castelporziano,Circeo)),(Cecina,Duna Feniglia))	1971	(San Rossore,(Cecina,((Duna Feniglia,Castelporziano),Circeo)))
1937	((San Rossore,(Castelporziano,Circeo)),(Cecina,Duna Feniglia))	1972	(San Rossore,((Cecina,Circeo),(Duna Feniglia,Castelporziano)))
1938	(((San Rossore,(Castelporziano,Circeo)),Cecina),Duna Feniglia)	1973	(San Rossore,((Cecina,Circeo),(Duna Feniglia,Castelporziano)))
1939	(((San Rossore,(Castelporziano,Circeo)),Duna Feniglia),Cecina)	1974	(San Rossore,((Cecina,Circeo),(Duna Feniglia,Castelporziano)))
1940	((San Rossore,(Castelporziano,Circeo)),(Cecina,Duna Feniglia))	1975	(San Rossore,((Cecina,Circeo),(Duna Feniglia,Castelporziano)))
1941	((San Rossore,(Castelporziano,Circeo)),(Cecina,Duna Feniglia))	1976	(((San Rossore,Duna Feniglia),Cecina),(Castelporziano,Circeo))
1942	(((San Rossore,(Castelporziano,Circeo)),Cecina),Duna Feniglia)	1977	(((San Rossore,Duna Feniglia),Cecina),(Castelporziano,Circeo))
1943	(((San Rossore,(Castelporziano,Circeo)),Cecina),Duna Feniglia)	1978	(((San Rossore,Duna Feniglia),Cecina),(Castelporziano,Circeo))
1944	(((Castelporziano,(San Rossore,Circeo)),Cecina),Duna Feniglia)	1979	(((San Rossore,Duna Feniglia),Cecina),(Castelporziano,Circeo))
1945	((Castelporziano,(San Rossore,Circeo)),(Cecina,Duna Feniglia))	1980	((San Rossore,Duna Feniglia),(Cecina,(Castelporziano,Circeo)))
1946	((San Rossore,Circeo),((Cecina,Duna Feniglia),Castelporziano))	1981	((San Rossore,Duna Feniglia),(Cecina,(Castelporziano,Circeo)))
1947	((San Rossore,Circeo),(Cecina,(Duna Feniglia,Castelporziano)))	1982	(Duna Feniglia,(Cecina,(San Rossore,(Castelporziano,Circeo))))
1948	((San Rossore,Circeo),(Cecina,(Duna Feniglia,Castelporziano)))	1983	((San Rossore,(Cecina,(Castelporziano,Circeo))),Duna Feniglia)
1949	((San Rossore,Circeo),(Duna Feniglia,(Cecina,Castelporziano)))	1984	((San Rossore,(Cecina,(Castelporziano,Circeo))),Duna Feniglia)
1950	(((San Rossore,Duna Feniglia),Circeo),(Cecina,Castelporziano))	1985	((San Rossore,(Cecina,(Castelporziano,Circeo))),Duna Feniglia)
1951	(((San Rossore,Duna Feniglia),Circeo),(Cecina,Castelporziano))	1986	(San Rossore,((Cecina,(Castelporziano,Circeo)),Duna Feniglia))
1952	(((San Rossore,Circeo),Duna Feniglia),(Cecina,Castelporziano))	1987	(San Rossore,((Cecina,(Castelporziano,Circeo)),Duna Feniglia))
1953	(((San Rossore,Circeo),Duna Feniglia),(Cecina,Castelporziano))	1988	(San Rossore,(Cecina,(Duna Feniglia,(Castelporziano,Circeo))))
1954	((San Rossore,Cecina),((Castelporziano,Circeo),Duna Feniglia))	1989	((San Rossore,Cecina),(Duna Feniglia,(Castelporziano,Circeo)))
1955	((San Rossore,Cecina),((Castelporziano,Circeo),Duna Feniglia))	1990	((San Rossore,Cecina),(Circeo,(Duna Feniglia,Castelporziano)))
1956	((San Rossore,Cecina),((Castelporziano,Circeo),Duna Feniglia))	1991	((San Rossore,Cecina),(Circeo,(Duna Feniglia,Castelporziano)))
1957	((San Rossore,Duna Feniglia),((Cecina,Circeo),Castelporziano))	1992	(((San Rossore,Circeo),Cecina),(Duna Feniglia,Castelporziano))
1958	(((San Rossore,Duna Feniglia),Castelporziano),(Cecina,Circeo))	1993	(((San Rossore,Circeo),Cecina),(Duna Feniglia,Castelporziano))
1959	(((San Rossore,Duna Feniglia),Castelporziano),(Cecina,Circeo))	1994	(((San Rossore,Circeo),Cecina),(Duna Feniglia,Castelporziano))
1960	((San Rossore,Duna Feniglia),((Cecina,Circeo),Castelporziano))	1995	((Duna Feniglia,Castelporziano),(Cecina,(San Rossore,Circeo)))
1961	((San Rossore,Duna Feniglia),((Cecina,Circeo),Castelporziano))	1996	(((San Rossore,Circeo),Cecina),(Duna Feniglia,Castelporziano))
1962	(San Rossore,(Duna Feniglia,((Cecina,Circeo),Castelporziano)))	1997	(((San Rossore,Circeo),Cecina),(Duna Feniglia,Castelporziano))
1963	(((San Rossore,Castelporziano),Duna Feniglia),(Cecina,Circeo))	1998	((San Rossore,Circeo),(Cecina,(Duna Feniglia,Castelporziano)))
1964	(((San Rossore,Castelporziano),Duna Feniglia),(Cecina,Circeo))	1999	((San Rossore,Circeo),(Cecina,(Duna Feniglia,Caselporziano)))

fusion level in the node description is of help in this respect, but for simplicity we did not show it here. Looking at the table, one may observe that, from 1934 to 1942 (except for 1935) the topology only slightly modifies, keeping constant the group *(San Rossore, Castelporziano, Circeo)*, to which both *Cecina* and *Duna Feniglia* gather in different ways. From 1976 to 1981, two groups appear, *(San Rossore, Duna Feniglia)* and *(Castelporziano, Circeo)*, with *Cecina* merging with either group. Eventually, from 1992 to 1999 the two groups *(San Rossore, Circeo, Cecina)* and *(Duna Feniglia, Castelporziano)* appear rather stable.

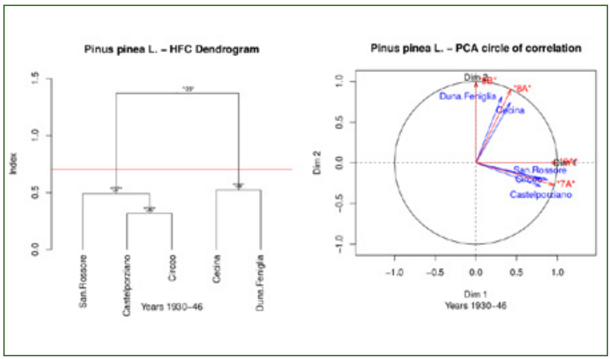
Discussion

The results of the evolutionary analyses induce to reconsider the outcomes proposed by Camiz et al. (2020). In fact, the new results show that the five chronologies proposed there are not so similar, despite of their close geographical proximity. Therefore, it is advisable to split the chronologies at hand in intervals and study them in detail.

Indeed, the group *(San Rossore, Castelporziano, Circeo)*, encompassing the northernmost and southernmost sites, is not so stable over time as expected, since both evolutionary analyses show different relevant groupings within this same area, highly depending on the seven detected time intervals (compare Figs. 4 and 5 with Tab. 4). During the first three, i.e., from 1930 to 1946, both groups *(San Rossore, Castelporziano, Circeo)* and *(Cecina, Duna Feniglia)* keep nearly constant. Therefore, it makes sense to run both static *HFC* and *PCA* limited to the interval 1930-46: there the two groups appear homogeneous – both dendrogram fusion indexes being around 0.5 – and well distinguished, as it may be observed both in Figure 10-*left* and in Figure 10-*right*, where their representative variables result nearly orthogonal and close to the first two principal components, respectively.

Between 1944 and 1964 the groups vary in number and composition, getting difficult any interpretation. Instead, in the intervals from 1965 to 1987 *Duna Feniglia* and *Castelporziano* show a similar pattern and all chro-

Figure 10 - *HFC* and *PCA* of the 5 chronologies of *P. pinea*, limited to the time interval 1930-46. Left: the dendrogram issued by the hierarchical factor classification. Right: the chronologies (in blue) and the representative variables of the first node and of those of the merging ones (in red) on the circle of correlations on the first factorial plane of *PCA*.

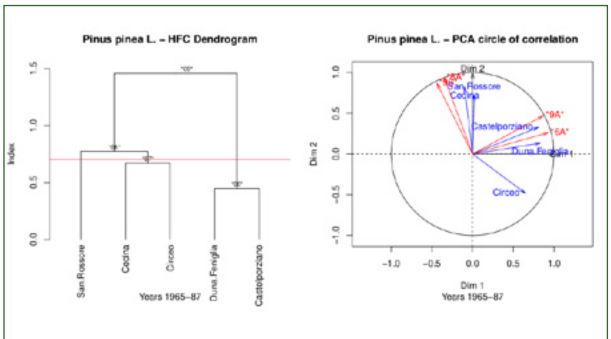


nologies are positively correlated with the first principal component, but episodically, so that that period may be studied in detail too. Indeed, based on the static analyses of that period, in Figure 11-*left* they appear grouped together, while *Cecina* and *Circeo* are grouped, including *San Rossore* in the further node *08*, which is not strictly unidimensional (fusion level 0.77). This may appear strange, and in fact in Figure 11-*right Cecina* and *Circeo* appear opposed, meaning that the node *07* is a dipole, i.e., the chronologies are negatively correlated. Anyway, the representative variables of the two nodes *06* and *08* are nearly orthogonal and oriented toward the first two principal components, respectively. Nevertheless, it must be observed that correlations in this period are lower than those in the interval 1930-46, a sign of instability.

Eventually, in the last interval, both *(San Rossore, Cecina)* and *(Duna Feniglia, Castelporziano)* remain nearly constant, while *Circeo* alternatively associated with either of them. Summarizing, the present studies confirm the transformation of the relations between the chronologies at hand over time, with only short periods of stability.

According to this complicated pattern, it is difficult to ascertain a common causalism. While the special behaviour of *Duna Feniglia* was attributed to its particular location – on a narrow sandy strip of dunes separating the sea from a lagoon, thus affected by salty water and exposed to both sides dominating winds (Camiz et al. 2020) – the three-chronologies group found there appears evident only during the first few years of the series and the newly found groups may be understood considering their geographical proximity, due to some common influencing factor, albeit limited in time.

Figure 11 - *HFC* and *PCA* of the 5 chronologies of *P. pinea*, limited to the time interval 1965-87. Left: the dendrogram issued by the hierarchical factor classification. Right: the chronologies (in blue) and the representative variables of the first node and of those of the merging ones (in red) on the circle of correlations on the first factorial plane of *PCA*.



Conclusion

The evolutionary study carried out on the five chronologies of *P. pinea* showed substantial differences over time of their correlation structure, hence suggested to study in detail some time sub-intervals. The two periods studied in detail – 1930-1946 and 1965-1987 – showed relevant differences, in particular concerning the more homogeneous structure of the earlier period and the different resulting groups of chronologies.

While the pattern of growth of *P. pinea* populations might be attributed to climate (Piraino et al. 2013), the chaotic behaviour of the five chronologies, described by Camiz et al. (2020), focuses on environmental changes at local level. Indeed, the results of the evolutionary analyses, show the consistence of such idea, giving way to a joint study of chronologies with meteorological and other local time series.

The investigation of the possible environmental determinants needs a causal insight, i.e., the study of the climatic conditions of the different sites at the time scale of few years. In Piraino et al. (2013) the response of *P. pinea* tree rings growth to climate has been addressed, finding relations with temperatures and humidity. Therefore, results similar to those proposed here might be expected by applying evolutionary exploratory methods to environmental local factors, such as meteorology, soil, edaphic conditions, diseases, and human impact, to reveal the sites' variations over time. Then, a comparison between the two sets of data, the chronologies, and the environmental determinants, seems necessary. For this task, other analysis methods are needed, exploring changes in the multivariate response: for this reason it has not been carried out here. Indeed, such study might take advantage from a deeper insight of the chronologies obtained from the methods proposed in this paper.

Acknowledgments

Thanks are due to Dr. Sergio Piraino who kindly provided the data of the chronologies. The authors are most obliged and grateful for the precious help of the anonymous referees, whose stringent critics and detailed revision highly contributed to significantly improve the first manuscript.

References

Aluja-Banet T., Morineau A., Sanchez G. 2018 - *Principal Component Analysis for Data Science.* [Online]. Available: https://pca4ds.github.io [September 2023].

Büntgen U., Frank D.C., Kaczka R.J., Verstege A., Zwijacz-Kozica T., Esper J. 2007 - Growth *responses to climate in a multi-species tree-ring network in the Western Carpathian Tatra Mountains, Poland and Slovakia.* Tree Physiology 27: 689–702. DOI: 10.1093/treephys/27.5.689.

Caliński T., Harabász J. 1974 - *A dendrite method for cluster analysis*. Communications in Statistics 3(1): 1–27. DOI: 10.1080/03610927408827101.

Camiz S. 2021 - *Exploratory classification of time-series.* In: "Handbook of Research on Emerging Theories, Models, and Applications of Financial Econometrics", Mercangöz B.A. Ed., Springer Nature, Cham (CH): 1–29. DOI: 10.1007/978-3-030-54108-8_1.

Camiz S., Diblasi A. 2013 - *Evolutionary principal component analysis*. In: "Trabajos Completos, XLI Coloquio Argentino de Estadística", 16-18 Octubre 2013, Universidad de Cuyo en Mendoza (Argentina, on CD-ROM): 680–685.

Camiz S., Maulucci R., Roig F.A. 2010 - *Exploratory analysis methods applied to dendrochronological series.* In: "Actas E-ICES 5", Malargüe (Mendoza, Argentina), 24-27 Octubre 2009, Mendoza, Papu O. Ed., Comisión Nacional de Energía Atómica: 61–67.

Camiz S., Roig F.A. 2011 - *Evolutionary analysis applied to tree-ring chronologies series.* In: "Actas E-ICES 6", CNEA 2011, Malargüe (Mendoza, Argentina), 4-8 Octubre 2010, Buenos Aires, Comisión Nacional de Energía Atómica: 30–37.

Camiz S., Spada F., Denimal J.-J., Piraino S. 2020 - *Hierarchical factor classification of dendrochronological time-series*. Annals of Silvicultural Research 45: 62–75. DOI: 10.12899/asr-1968.

Denimal J.-J. 2001 - *Hierarchical Factorial Analysis*. In: "Actes du 10th International Symposium on Applied Stochastic Models and Data Analysis", Compiègne, 12-15 Juin 2001.

Denimal J.-J. 2007 - *Classification hiérarchique optimisée d'un tableau de mesures*. Revue de Statistique Appliquée 148: 29–61.

Felsenstein J. 2021 - *The Newick tree format*. [Online]. Available: https://evolution.genetics.washington.edu/phylip/newicktree.html. [August 19, 2021].

Fisher W.D. 1958 - *On grouping for maximum homogeneity.* Journal of the American Statistical Association 53(284): 789–798.

Frank D., Esper J. 2005 - *Characterization and climate response patterns of a high-elevation, multi-species tree-ring network in the European Alps*. Dendrochronologia 22(2): 107–121. DOI: 10.1080/01621459.1958.10501479.

Jolliffe I. 2002 - *Principal Component Analysis*. Springer, Berlin. ISBN: 978-1-4757-1904-8.

Lebart L., Piron M., Morineau A. 2006 - *Statistiques exploratoire multidimensionnelle: Visualisations et inférences en fouille de données.* Paris, Dunod.

Leland C., Pederson N., Hessl A., Nachin B., Davi N., D'Arrigo R., Jacoby G. 2013 - *A hydroclimatic regionalization of central Mongolia as inferred from tree rings*. Dendrochronologia 31(3): 205–215. DOI: 10.1016/j.dendro.2012.11.003.

Peters K., Jacoby G.C., Cook E.R. 1981 - *Principal components analysis of tree-ring sites*. Tree-Ring Bulletin 41: 1–19.

Piraino S., Camiz S., Filippo A.D., Piovesan G., Spada F. 2013 - *A dendrochronological analysis of* Pinus pinea *L. on the italian mid-tyrrhenian coast*. Geochronometria 40: 77–89. DOI: 10.2478/s13386-012-0019-z.

Robinson D.F., Foulds L.R. 1979 - *Comparison of weighted labelled trees*. In: "Combinatorial Mathematics 6", Horadam A.F., Wallis W.D. Eds., Lecture Notes in Mathematics, Springer, Berlin 748: 119–126. DOI: 10.1007/BFb0102690.

Robinson D.F., Foulds L.R. 1981 - *Comparison of phylogenetic trees*. Mathematical Biosciences 53(1-2): 131–147. DOI: 10.1016/0025-5564(81)90043-2.

Wikipedia contributors 2021 - *Newick format*. Wikipedia, the free encyclopedia. [Online]. Available: https://en.wikipedia.org/wiki/Newick_format. [August 19, 2021].