



HAL
open science

Comparative approach of multivariate statistics to decipher trade off in temporal datasets

Lisa Baboin, Servane Berton, Dylan Dunand, Quentin M. Dutertre, Sabrina Gilardi, Jean-Baptiste R. Guy, Bérangère Leys

► To cite this version:

Lisa Baboin, Servane Berton, Dylan Dunand, Quentin M. Dutertre, Sabrina Gilardi, et al.. Comparative approach of multivariate statistics to decipher trade off in temporal datasets. *Ecologia mediterranea*, 2022, 48 (2), pp.33-40. hal-03992246

HAL Id: hal-03992246

<https://amu.hal.science/hal-03992246v1>

Submitted on 16 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparative approach of multivariate statistics to decipher trade off in temporal datasets

Approche statistique multivariée pour décrire les données temporelles long terme (millénaire)

Lisa BABOIN¹, Servane BERTON¹, Dylan DUNAND¹, Quentin M. DUTERTRE¹,
Sabrina GILARDI¹, Jean-Baptiste R. GUY¹, Bérangère LEYS^{2,*}

1. Université Aix-Marseille Campus Saint-Jérôme, Marseille

2. IMBE (CNRS, AMU, IRD), Technopole Arbois Méditerranée, Aix-en-Provence

* Corresponding author: berangere.leys@imbe.fr

Received: 6 July, 2022; First decision: 12 Oct., 2022; Revised: 14 Nov., 2022; Accepted: 10 Dec., 2022

Abstract

Paleorecords are replete with examples of long-term biotic responses to past climate change, and disturbances such as fire, herbivores, insect outbreaks or ice storms. The now frequent challenges in palaeoecological studies are to better understand both the trends and the breaking points of biotic variables in response to disturbances. The aim of this article is to expand the capacity of new and existing methods to provide greater insight into ecosystem trajectories and functioning.

Our analyses focus on biomes where fire is a key structural disturbance for vegetation establishment and evolution. We aim to determine: 1) explanative variances between biotic and abiotic variables at both short and millennial time scales, and 2) whether a temporal relationship exists between fire regimes and plant diversity change in both systems.

Our statistical approach focused on PCA (Principal Component Analysis), co-inertia, MFA (Multiple Factor Analysis) and Hill & Smith methods, which provide us new information about our data, and bring new conclusions to the former studies.

We advise to preferentially use the MFA when the environmental variables are numerous, while the co-inertia is a better tool to explain variance of the datasets.

Keywords: vegetation, fire, long-term, co-inertia, MFA, PCA.

Introduction

Paleorecords are replete with examples of long-term biotic responses to past climate change, and disturbances such as fire, herbivores, insect outbreaks or ice storms to cite a few (Whitlock *et al.* 1997; Whitlock & Larsen 2002; McLauchlan *et al.* 2014). One particularly useful set of records are those preserved in lake and marine sediments, recording both past environmental changes and corresponding biotic responses (*e.g.* Leys *et al.* 2018; Genet *et al.* 2021).

The now frequent challenges in palaeoecological studies are to better understand both the trends and the breaking points of biotic variables in response to disturbances. Recently there has been increasing focus on the need for conservation of ecological and evolutionary processes in the face of climate change based on long-term records.

The lake sedimentary archives, due to their anoxic environment, preserve the biotic and abiotic factors of the ecosystems particularly well (McLauchlan *et al.* 2014). Generally, the sediments are deposited since the formation or the outcrop of the wetland, *i.e.*,

around 20,000 years for most moraine lakes in Europe, and continuously. They are cut into thin slices (0.5cm or 1cm) to reconstruct past environments over time steps ranging from a few years to a few decades depending on the rate of sedimentation (which depends on the production of organic matter in the lake, and inputs from the watershed) (Scheffer *et al.* 2001; Leys *et al.* 2013, 2018). Thus, it is possible to quantify both the post-disturbance response of the vegetation, and the millennial trajectories (stability and changes) of the vegetation and the disturbance regimes.

For instance, the simple analysis of correlation between the biomass burned (through charcoal records) and the vegetation (by pollens) of the last millennia in central and southern Europe has clearly shown that fires have been a major factor in the change of vegetation 4,500 years ago leading to the establishment of meadows and scrubland, the reduction of mesophyte forests and the extension of arable and pastoral land (Connor *et al.* 2012). Recent work on the Iberian Peninsula has shown more precisely that some of the changes observed in the millennial trajectories of vegetation in response to the fire regime constitute change points. Quasi-permanent transitions in vegetation status are characterized by shifts from wooded to open vegetation and from a fire-adapted system (xerophilic bushes and fire-adapted pines) to a fire-resistant system fire (*Pinus nigra* J.F.Arnold. and the mesophytes) (Connor *et al.* 2012).

The challenge is to expand the capacity of new and existing methods to provide greater insight into ecosystem trajectories and functioning. Here we examine Mediterranean and grassland fire and vegetation diversity histories by developing alternative approaches based on multivariate analyses. These two case studies allow to analyze temporal datasets at both decennial and millennial scales. Our analyses focus on biomes where fire is a key structural disturbance for vegetation establishment and evolution. We aim to determine: 1) explanative variances between biotic and abiotic variables at both short and millennial time scales, and 2) whether a temporal relationship exists between fire regimes and plant diversity change in both systems. To achieve these aims, we will present different and complementary statistical approaches using objective criteria. Finally, we discuss the likely drivers of fire regime and diversity change on millennial and decennial timescales at the light of the different multivariate analyses.

Material and methods

Study sites

The two site cases have been chosen to understand the charcoal production used to illustrate the relationships between fire and vegetation diversity, at both short and long-time scales.

(1) The Flint Hills ecoregion (37°17'00"N 96°40'31"W) is the largest remaining contiguous tallgrass prairie ecosystem in North America, encompassing over 50,000km². The current climate conditions provide enough moisture to support woody plants, and grassland is frequently and extensively burned throughout the region to maintain open conditions (Reichman 1987). The study site of Konza Prairie (39°05'N, 96°35'W) is in the Flint Hills region of northeastern Kansas (hereafter, "Konza"). Konza is a 3,500 hectares native tallgrass prairie preserve. Sediments in traps from Konza Prairie (LTER) represents the last 4 years of vegetation composition, fire activity and other environmental parameters (climate, herbivores) in the charcoal production (table 1).

(2) Lac de Creno (1,310m a.s.l.; 42°12'17"N, 08°56'45"E) is in the southeastern part of the Rotondo massif (western Corsica, west Mediterranean basin). Macrofossils of *Pinus laricio* Poir. provide evidence for its growth around the Lac de Creno since 13,200 cal year BP, which is consistent with its natural origin in Corsica. Sediments from Lac de Creno in Corsica (France) extend the Holocene period and from which a reconstruction of vegetation composition and fire history is available (table 2).

All details are available in (Leys *et al.* 2014) and (Leys *et al.* 2016).

Statistical approaches

PCA

Principal component analyses (PCA), standardized by taxa with log-transformed concentration data, were used to visualize the relationships among taxa inferred from macrofossils accumulation rates (table 2). The log-transformation allow to minimize extreme value and stabilize the variance within the data, which is particularly required in this case study where number of macroremains can vary of a factor 25 from one taxa to another (Ammann *et al.* 2000). To limit the biases of

Table 1 – Groups of data used for statistical analyses for the site of Konza Prairie: charcoal measurements, fire parameters, vegetation and landscape data, and climate.

Parameter types	Name	Unit	Formula	Sources
Charcoal	Morphotypes: Cell, Porous, Fibrous, Sponge, Branch, Dark	pieces		Trap content
	Non-Arboreal	pieces	Cell + Porous + Fibrous	
	Total	pieces	Sum of all morphotypes	
	Non-Arboreal/Total	pieces	Ratio of non-arboreal charcoal pieces on total amount of charcoal pieces	
	125µm-250µm	pieces		Trap content
	250µm-1mm	pieces		Trap content
Fire	100 to 500m buffer area burn	km ²	Sum of area burn of watershed within the 100 (to 500) m radius of a trap	Konza website
	watershed burn	km ² .watershed-1		Konza website
	Konza area burn	km ²	Sum of area burn of watersheds	
	FlintHills area burn	km ²		FlintHills annual report
	month of burn			Konza website
	Fire return interval	years	Average fire return intervals of watersheds within the 500m radius distance of a given trap	Konza website
	Fire frequency	fire.100years-1	Inverse of fire return interval multiplied by 100	
	Type of fire	3 classes		Konza website
	Time since last fire	year	Average number of years since last fire of watershed burn within the 500m radius distance of a given trap	
	Fuel burn index		Sum of 500m buffer area multiplied by time since last fire	
Vegetation and landscape	Woody	5 classes		Field survey
	Herbaceous	5 classes		Field survey
	Shrubbery	yes/no		Field survey
	Height	dm and 2 classes		Field survey
	Bare Soil	5 classes		Field survey
	Limestone rocks	yes/no		Field survey
	Bison Manure	yes/no		Field survey
	Elevation	m		Field survey
	Landscape position	3 classes		Field survey
Climate	Wind Speed	m.s-1	Average on 8 days following the date of burn	Konza Prairie Meterorological station
	Relative Humidity	%	Average on 8 days following the date of burn	Konza Prairie Meterorological station
	Precipitation	mm	Average on 8 days following the date of burn	Konza Prairie Meterorological station
	Solar Radiation	W	Average on 8 days following the date of burn	Konza Prairie Meterorological station
	Air temperature (min and max)	°C	Average on 8 days following the date of burn	Konza Prairie Meterorological station
	Mean Air temperature	°C	Average on 8 days following the date of burn	Konza Prairie Meterorological station
	Soil temperature (min and max)	°C	Average on 8 days following the date of burn	Konza Prairie Meterorological station
	Evapotranspiration	mm	Average on 8 days following the date of burn	Konza Prairie Meterorological station

erosions events that could have created artefactual increase of taxa abundances though increase of matter in the lacustrine system, we used the taxa accumulation rate, which correct the taxa abundance by the accumulation rate of sediments through time. To visualize the

relationships between the local fire-episodes regime and the local vegetation inferred from macrofossil assemblages, CHAR_{background} (proxy of the burning area) and FRI (proxy of local fire episodes) were added passively in the PCA ordination plots. All ordinations

Table 2 – Description of all variables used in the following analyses for Lac de Creno.

Parameter types	Name	Description	Unit
Age		²¹⁰ Pb and ¹³⁷ Cs measurements were used to build an accurate age-depth model	cal. year BP
Charcoal	CHAR _{area}	Charcoal particles > 250µm ² were identified. Charcoal measurements are reported as charcoal accumulation rates (CHAR _a)	mm ² cm ⁻² year ⁻¹
	CHAR _{background}	2 Components were derived from Char _{area} : CHAR _{peak} which represents the charcoal peaks for local fire episodes. (CHAR _p) CHAR _{background} which indicated the charcoal background for regional area burned. (CHAR _r)	mm ² cm ⁻² year ⁻¹
Fire	Fire Return Interval (FRI)	Number of fire between two consecutive years (derived from CHAR _p)	year.fire ⁻¹
	Fire Frequency (FF)	Number of fires per unit of time, smoothed over a 1000-year smoothing window.	year
Vegetation	Plant macrofossil	Plant macrofossils were extracted by water-sieving (250-µm mesh size) and identified to genus, species when possible	number of fragments cm ⁻² .year ⁻¹

were carried out in R Studio (version 4.1.1), package ade4 (Dray & Dufour 2007), using the function dudi.pca().

Co-inertia

We performed a co-inertia analysis (Dray *et al.* 2003) based on a Correspondence Analysis (CA) on vegetation data and a normalized Principal Component Analysis (PCA) on fire data to assess the covariance between these two components (table 2). The PCA was performed using the CA's rows weight to match co-inertia analysis requirement. The co-inertia analysis was performed with the coinertia() function from the ade4 package (Dray & Dufour 2007) on the last 6,000 years of history where fire and vegetation data were the most complete.

To enhance the interpretation of co-inertia results we performed a Monte-Carlo test (Heo & Ruben Gabriel 1998) with 9,999 permutations on both CA and PCA. It allowed us to assess the correlation between vegetation and fire data. The function RV.rtest() from the ade4 package was used.

Hill and Smith

We used multivariate analyses to visualize the relationships among environmental parameters, fire parameters, and charcoal measurements. Because our dataset included both quantitative and qualitative variables (table 1), we used a Hill and Smith analysis (Hill & Smith 1976) to identify a specific set of variables that avoid redundancy.

MFA

All the variables (table 1) were grouped in four categories: Charcoal (8 variables), Fire (2 variables), Climate (4 variables), Vegetation (5 variables) (table 1).

To represent the relationships among all groups of individual variables, we performed a multiple factor analysis [MFA, (Escofier & Pages 1994; Bécue-Bertaut & Pagès 2008)] using the MFA() function from the FactoMineR package (Husson *et al.* 2007).

The MFA is a symmetrical ordination based on correlation, and this particular analysis was specifically chosen to ordinate the different variables (not assumed to be explanative in the first place) with no risk of overfitting. We plotted the MFA results using the fviz_mfa() function from the factoextra package (Kassambara & Mundt 2019). We graphically presented the quality of the representation (cos²) of each variable for the four first axes of the MFA, i.e. the proportion of the variable represented by each axis using the corrplot() function from the corrplot package (Wei *et al.* 2017). To analyze the relationships among the four groups of variables, we extracted the RV coefficients from the MFA, measuring the similarity (squared correlation) between each pair of groups. Then, we tested these coefficients using the coeffRV() function from the FactoMineR package.

Results and discussion

MFA pinpointed new relationship between fire and climate parameters

At Konza Prairie, from the precedent study, the Hill and Smith analysis (Figure 1) demonstrated no direct relationships between charcoal production and fire variables < 500m around the trap, as indicated by non-significant correlations (i) between different measurements of charcoal and (ii) between total, non-arboreal charcoal or NA/T ratio and fire parameters (burn area at 500m distance from the trap, fuel burn estimation, fire frequency

listed variables except solar radiation and air temperature for climate parameters). MFA axis 1 (hereafter MFA 1) is represented by vegetation at 21.50%, while the second axis (hereafter MFA 2) is represented by charcoals at 14.15%.

The relationships between the fire and climate groups follow two different trajectories given the explanative variables of interest. First, the fire variables were positively correlated with the evapotranspiration variable which brings together the measurements of maximum temperature, minimum temperature, solar radiation rate and evapotranspiration. Conversely, the variables negatively correlated to fires

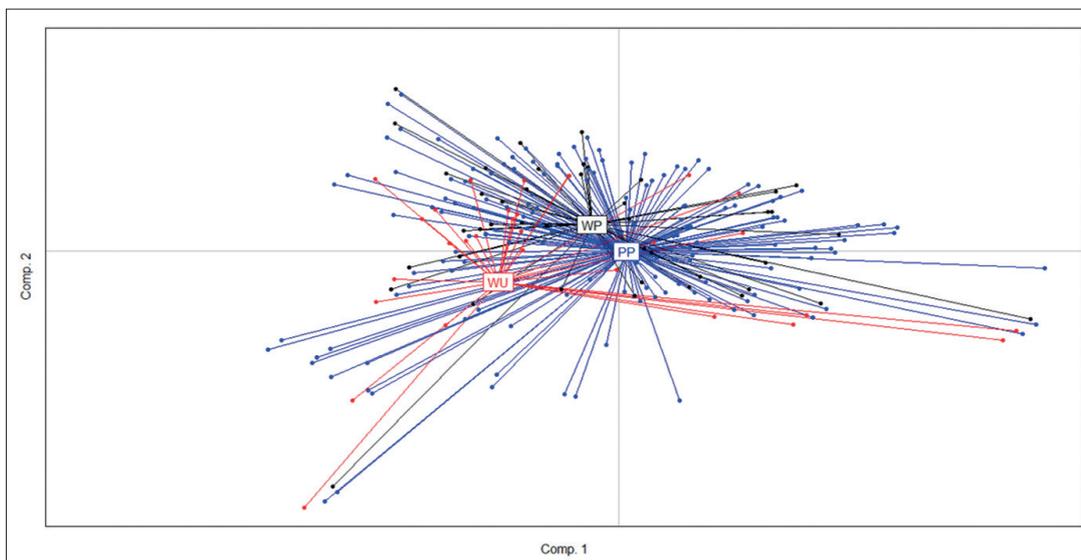


Figure 1 – Ordination plots from Hill and Smith multivariate analysis of numerical data in function of fire type. (WU = wildfire burn unplanned for a given year; PP = prescribed burn planned; WP = wildfire burn planned for a given year).

and time since last fire). In addition, the authors performed a random forest analysis (more details in Leys *et al.* 2016) which highlighted that total charcoal or NA/T ratio are most strongly predicted using fire parameters at a larger spatial scale and with environmental variables, both of which varied during each of the four years of the study. Charcoal variables, including Total charcoal count and NA/T ratio, did not correlate with other fire parameters, vegetation cover, landscape, or climate variables.

Here, we performed an MFA (Figure 2) to test whether the charcoal data (all morphotypes and 125µm) in one hand and the fire parameters (buffer at 300m, all other parameters were strongly correlated) on the other hands are explained by the two other categories: climate and vegetation (table 1, all

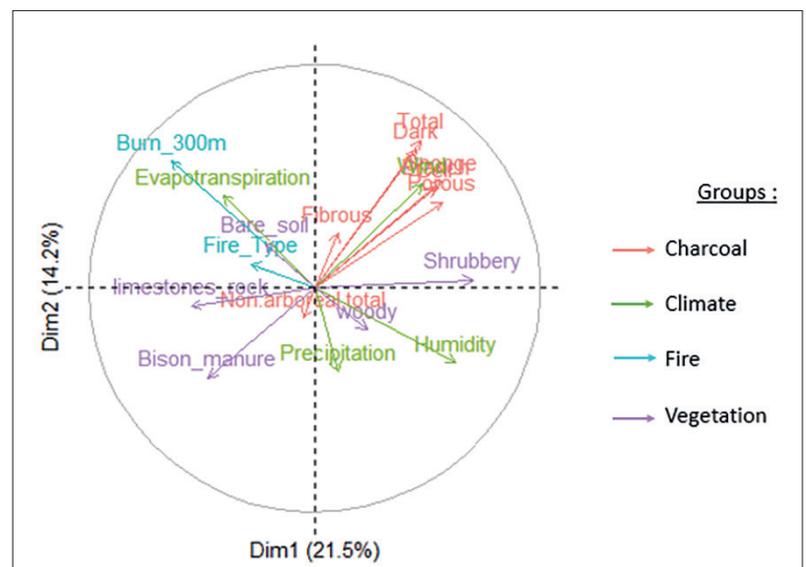


Figure 2 – Multiple factor analysis (MFA) of 4 parameters: charcoal, climate, fire and vegetation.

were humidity and precipitation measurements. In addition to precedent results, the MFA pinpoint different relationships between charcoal, vegetation and climate groups. A positive correlation between the charcoal variables and the wind measurement appears positively correlated with charcoal variables. A positive correlation appears between the charcoal variables and the wind measurement. The RV coefficient, which measures the similarity between pairs of variables, were 12% between the vegetation and fire variable groups, and 7% between the climate and vegetation groups. Thus, fire parameters seem to correlate more the vegetation parameters than does the climate.

Co-inertia allows to identify time period where fire played a major role in the vegetation assemblage

The onset of the Holocene in Corsica (~11,700 cal year BP) was marked by a rapid shift in the treeline, the establishment of *P. laricio*-dominated woodlands with low species turnover, and a long-term increase in taxa richness as a result of successive expansions of broad-leaved deciduous trees. In spite of fire-return intervals (FRIs) fluctuating between 30 and 490 years during the Holocene, *P. laricio*-dominated woodlands persisted for several millennia, and fires likely played an ecological role in the functioning of these woodlands. The 12,000-year record of mean FRI is of 80 years between fires; i.e., frequency of 12.5 fires 1000 year-1 (Leys *et al.* 2014). Previous analysis with PCA method on the

relationship among fire and vegetation displays a weak explanative percentage of variance explained (4%).

The co-inertia explores the covariances between both fire and vegetation categories, based respectively on PCA for charcoal parameters (FRI and CHARa) and CA for vegetation data. From the PCA, axis 1 is well explained by the Fire Return Interval (FRI) while axis 2 is well explained by the charcoal accumulation rate (Figure 3 B). From the CA, axis 1 is explained positively with *Quercus* and negatively with *Alnus glutinosa* (L.) Gaertn and *Juniperus communis* L., while axis 2 is positively correlated with *Alnus cordata* (Loisel.) Duby and negatively with *Berberis* (Figure 3 A). Combined, co-inertia axis 1 explained 60% of the variance, the second axis explained 11%.

The graphical representation of the co-inertia results allows to associate at each observation represented by the score of the CA in the ordination plot, with an arrow representing the predicting trend from PCA ordination (Figure 4). The longer the arrow, the more the observation (sample with vegetation assemblage) was unexpected given the fire condition at this time. For instance, the fire conditions were consistent over the last millennia (Leys *et al.* 2014), although the vegetation assemblages at 127 (125-129) cal year BP differed by the presence of *Juniperus communis* (Figure 4 B). Besides, at 5,422 (5,344-5,579) cal year BP, the charcoal accumulation rate was particularly high given the vegetation assemblage. This result suggests that the vegetation is not following the expected assemblages given the fire regime recorded, maybe reflecting the beginning of a fire regime change appearing before a vegetation change (Figure 4 B). On the other side, at 4,148 (4,045-4,230) cal year BP, the arrow is not even visible, meanings that the vegetation was consistent with the fire regime (Figure 4 B).

To strength the results statistically, we can calculate a p-value through Monte-Carlo test on the RV-coefficients (RV.rtest() in ade4 package). Given the non-significance of our results of our analysis we decided to notshow this part here.

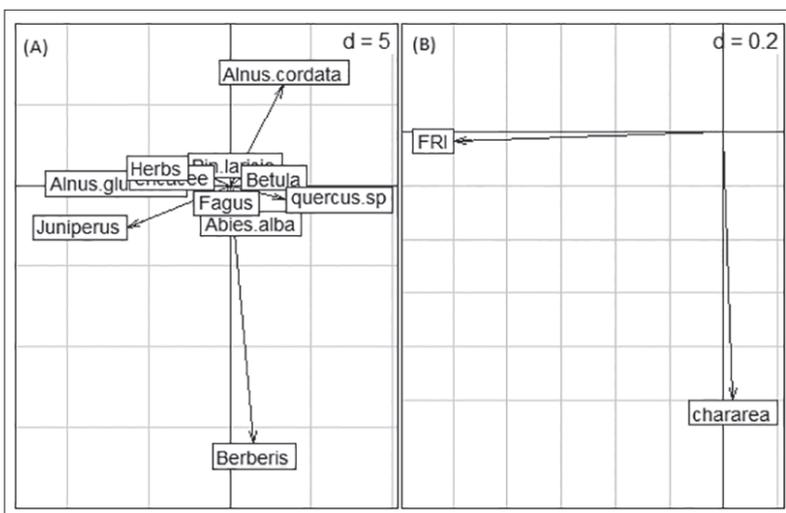


Figure 3 – Direction of the CA variables (A) and the PCA variables (B).

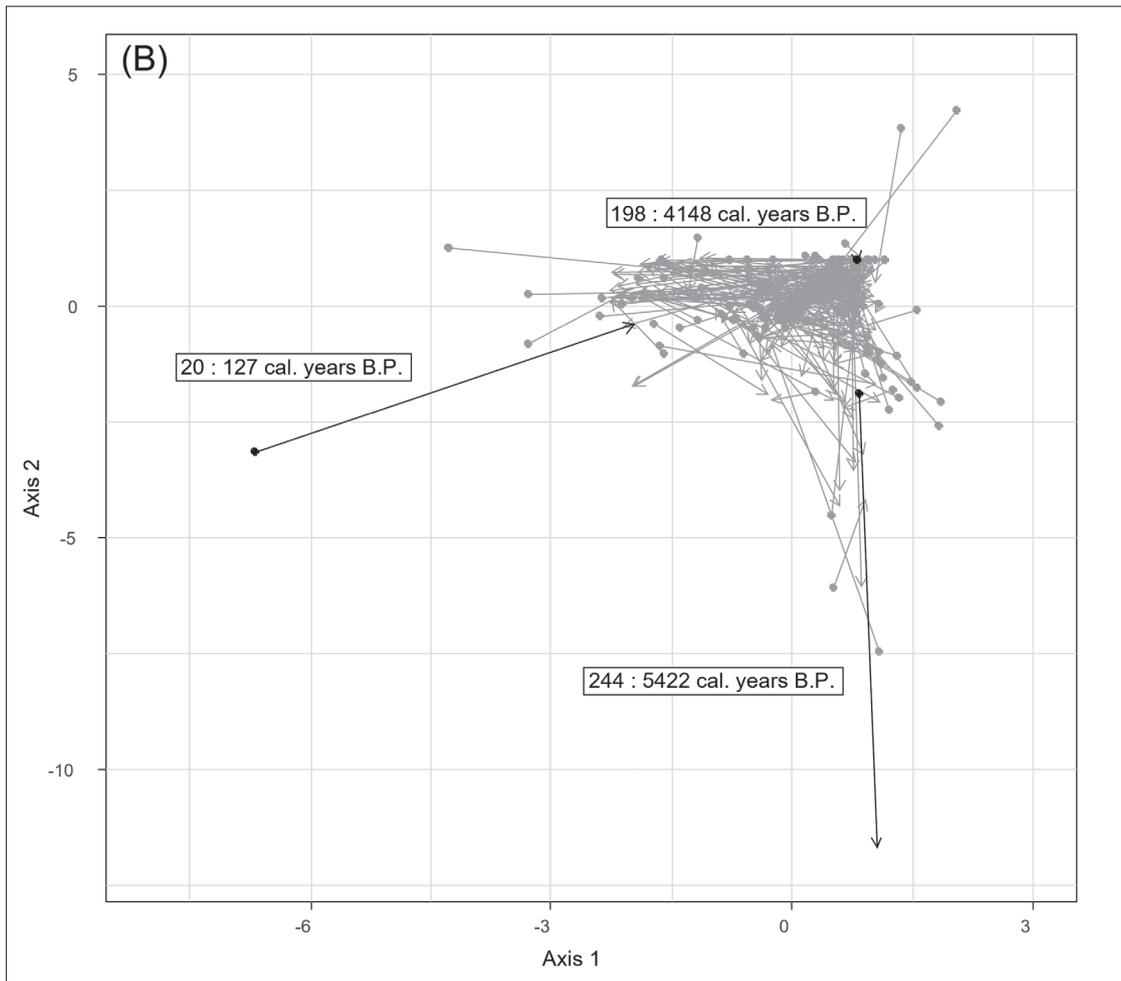
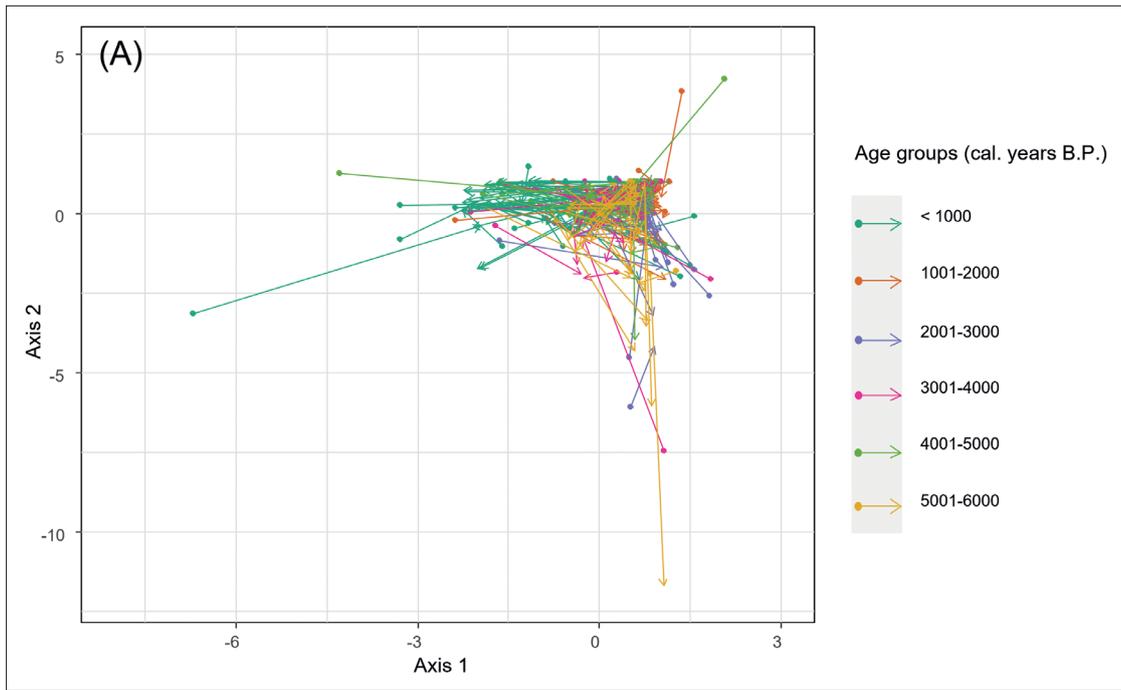


Figure 4 – Scores of all the observations on the co-inertia axis (A); scores of the observations n° 20, 198 and 244 on co-inertia axis (B).

Conclusion

Multivariate analyses can bring new information on palaeoecological data, and complete efficiently the description of long-term trajectories. These analyses found their spot to answer key questions relative to ecological questions, on both long- and short-time windows studied. We advise to preferentially use the MFA when the environmental variables are numerous, while the co-inertia is a better tool to explain variance of the datasets. We presented chosen multivariate analyses, although we strongly encourage the use of all multivariate analyses that answer the best to the dataset studied and their related questions.

Acknowledgments

We gratefully thank Franck Torre, Eric Meineri and Marie Finocchiaro for their insights and the interesting discussions while preparing the manuscript. We also thank Mathis Degand and Manon Berchet for their implication on the project and their help with statistical tests realization.

Author contribution

BL conceived and ran the research project. All authors contributed equally to data analysis and writing.

Bibliography

- Ammann, B., Birks, H.J.B., Brooks, S.J., Eicher, U., Von Grafenstein, U., Hofmann, W., *et al.* (2000). Quantification of biotic responses to rapid climatic changes around the Younger Dryas - a synthesis. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 159, 313-347.
- Bécue-Bertaut, M. & Pagès, J. (2008). Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. *Comput. Stat. Data Anal.* 52, 3255-3268.
- Connor, S.E., Araujo, J., Van der Knaap, W.O., van Leeuwen, J.F.N., Araújo, J., Van der Knaap, W.O., *et al.* (2012). A long-term perspective of biomass burning in the Serra da Estrela, Portugal. *Quat. Sci. Rev.* 55, 114-124.
- Dray, S., Chessel, D. & Thioulouse, J. (2003). Co-inertia analysis and the linking of ecological data tables. *Ecology* 84, 3078-3089.
- Dray, S. & Dufour, A.-B. (2007). The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22, 1-20.
- Escofier, B. & Pages, J. (1994). Multiple factor analysis (AFMULT package). *Comput. Stat. Data Anal.* 18, 121-140.
- Genet, M., Daniau, A.-L., Mouillot, F., Hanquiez, V., Schmidt, S., David, V., *et al.* (2021). Modern relationships between microscopic charcoal in marine sediments and fire regimes on adjacent landmasses to refine the interpretation of marine paleofire records: An Iberian case study. *Quat. Sci. Rev.* 270, 107148.
- Heo, M. & Ruben Gabriel, K. (1998). A permutation test of association between configurations by means of the RV coefficient. *Commun. Stat. Comput.* 27, 843-856.
- Hill, M.O. & Smith, A.J.E. (1976). Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon*, 249-255.
- Husson, F., Josse, J., Lê, S. & Mazet, J. (2007). Factor analysis and data mining with R. *FactoMineR Packag.*
- Kassambara, A. & Mundt, F. (2019). Factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0. 5. 2017.
- Leys, B., Carcaillet, C., Dezileau, L., Ali, A.A. & Bradshaw, R.H.W. (2013). A comparison of charcoal measurements for reconstruction of Mediterranean paleo-fire frequency in the mountains of Corsica. *Quat. Res. (United States)* 79.
- Leys, B., Finsinger, W. & Carcaillet, C. (2014). Historical range of fire frequency is not the Achilles' heel of the Corsican black pine ecosystem. *J. Ecol.* 102, 381-395.
- Leys, B., Higuera, P.E., McLauchlan, K.K. & Dunnette, P.V. (2016). Wildfires and geochemical change in a subalpine forest over the past six millennia. *Environ. Res. Lett.* 11.
- Leys, B., Marlon, J.R., Umbanhowar, C. & Vannièrè, B. (2018). Global fire history of grassland biomes. *Ecol. Evol.* 8, 8831-8852.
- McLauchlan, K.K., Higuera, P.E., Gavin, D.G., Perakis, S.S., Mack, M.C., Alexander, H., *et al.* (2014). Reconstructing disturbances and their biogeochemical consequences over multiple timescales. *Bioscience* 64, bit017.
- Reichman, O.J. (1987). *Konza Prairie: a tallgrass natural history*. University Press of Kansas, Lawrence.
- Scheffer, M., Carpenter, S., Foley, J. a, Folke, C. & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature* 413, 591-596.
- Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y. & Zemla, J. (2017). Package "corrplot". *Statistician* 56, 316-324.
- Whitlock, C. & Larsen, C. (2002). Charcoal as a fire proxy. *Track. Environ. Chang. using lake sediments*, 75-97.
- Whitlock, C. & Anderson, R.S. (1997). 1. Fire History Reconstructions Based on Sediment Records from Lakes and Wetlands. *History*, 1-29.