



A method to characterize climate, Earth or environmental vector random processes

M. Cobos^{1,2} · P. Otiñar^{1,2} · P. Magaña^{1,2} · A. Baquerizo^{1,2}

Accepted: 6 May 2022
© The Author(s) 2022

Abstract

We propose a general methodology to characterize a non-stationary random process that can be used for simulating random realizations that keep the probabilistic behavior of the original time series. The probability distribution of the process is assumed to be a piecewise function defined by several weighted parametric probability models. The weights are obtained analytically by ensuring that the probability density function is well defined and that it is continuous at the common endpoints. Any number of subintervals and continuous probability models can be chosen. The distribution is assumed to vary periodically in time over a predefined time interval by defining the model parameters and the common endpoints as truncated generalized Fourier series. The coefficients of the expansions are obtained with the maximum likelihood method. Different sets of orthogonal basis functions are tested. The method is applied to three time series with different particularities. Firstly, it is shown its good behavior to capture the high variability of the precipitation projected at a semiarid location of Spain for the present century. Secondly, for the Wolf sunspot number time series, the Schwabe cycle and time variations close to the 7.5 and 17 years are analyzed along a 22-year cycle. Finally, the method is applied to a bivariate time series that contains (1) freshwater discharges at the last regulation point of a dam located in a semiarid zone in Andalucía (Spain) which is influenced not only by the climate variability but also by management decisions and (2) the salinity at the mouth of the river. For this case, the analysis, that was combined with a vectorial autoregressive model, focus on the assessment of the goodness of the methodology to replicate the statistical features of the original series. In particular, it is found that it reproduces the marginal and joint distributions and the duration of sojourns above/below given thresholds.

Keywords Generalized Fourier series of parameters · Non-stationary probability models · Piecewise continuous Probability density functions · Stochastic characterization · Time series of environmental processes

1 Introduction

The long-term analysis of a natural phenomenon is usually done from observations of multivariate time series whose statistical properties are representative of the conditions during regular time intervals known as states. For meteorological and wave climate, the duration of a state usually ranges from several minutes to a few hours. Those time

series, particularly if forced by climatic conditions, exhibit different probabilistic behavior along time associated to natural variations at different scales including daily, synoptic, seasonal and yearly. At longer temporal scales, variations are related to climatic oscillations usually described by indexes (Monbet et al. 2007) such as the South Oscillation that was first identified by Hildebrandsson (1897), the North Atlantic Oscillation recognized by (see Walker (1924)) and the North Pacific Oscillation first noticed by Walker and Bliss (1932) and ultimately to solar activity (see e.g. Zhai (2017); Le Mouél et al. (2019)).

For the stochastic characterization of those vector random processes, it is essential to take into account the time variability for the whole range of values. This type of analysis is usually aimed at simulating time series with the

✉ P. Otiñar
potinar@ugr.es

¹ Department of Structural Mechanics and Hydraulic Engineering, University of Granada, Edificio Politécnico, Campus de Fuentenueva, 18071 Granada, Spain

² Andalusian Institute for Earth System Research, Avda. del Mediterráneo s/n, Granada, Spain

same probabilistic structure, so that they can be used to infer the random response of a given system. Some examples of applications are (i) the study of beach evolution (Payo et al. 2004; Baquerizo and Losada 2008; Callaghan et al. 2008; Félix et al. 2012; Ranasinghe et al. 2012), (ii) the optimal design and management of an oscillating water column system (Jalón et al. 2016; López-Ruiz et al. 2018), (iii) the planning of maintenance strategies of coastal structures (Lira-Loarca et al. 2020), and (iv) the assessment of water quality management strategies in an estuary according to density variations and recovery time (Cobos 2020). It has also been used for the analysis of observed wave climate variability in the preceding century and the expected changes in projections under a climate change scenario (Loarca et al. 2021).

In environmental sciences, there are many proposals for the simulation of time series that focus on the generation of the values above a given threshold (known as storm conditions for climate variables) or the full time series. Some of them treat the series as stationary while more recent approaches consider their non-stationarity. The earliest attempts to reproduce stormy conditions in sea state wave climate analysis treated the occurrence of storms as Poisson events with exponential interarrival times. Their persistence was usually obtained by means of the joint distribution of peaks and durations and they used idealized storm shapes (e.g. (Callaghan et al. 2008; Boccotti 2000; De Michele et al. 2007; Fedele and Arena 2009; Corbella and Stretch 2012)). Payo et al. (2008) reproduced the growth and decay of wave energy in the storms using empirical orthogonal functions.

In the field of Geostatistics, a full theoretical framework for spatiotemporal processes has been developed (Christakos 2017; Wu et al. 2021; Christakos 2000). The analysis of this type of random fields is based in the space-time covariance and the bayesian maximum entropy. Several examples can be found (He and Kolovos 2018; He et al. 2021; Cobos et al. 2019). Another approach is followed in the present paper whose analysis is limited to time variability at a specific location. In this regard, several works analyze the time variability in maxima attained during a given time interval (De Leo et al. 2021; Izaguirre et al. 2010; De Luca and Galasso 2018), in peaks over threshold (Méndez et al. 2006, 2008; Jonathan and Ewans 2013) and frequencies of exceedances (Luceño et al. 2006; Razmi et al. 2017) in different climatic time series.

Solari and Losada (2011) proposed a non-stationary parametric distribution to characterize the whole range of values with a piecewise distribution that uses a log-normal distribution for the central body and two generalized Pareto distributions for the lower and upper tails. Based on this work, Solari and Van Gelder (2011) proposed a similar approach to deal with wave periods and mean incoming

wave direction in addition to the simulation of multivariate time series with a vectorial autoregressive model (VAR). In them, they use specific combinations of probability models that do not necessarily work for other type of data. Also, the non-stationary character of the random variables is considered by expressing the free parameters of the distribution and the percentiles of the common endpoints as a truncated trigonometric expansion taking the year as the largest cyclic periodicity.

In relation to this last aspect, the expansion in trigonometric series may give inaccurate results when the derivatives at the limits of the interval do not coincide. In addition, the existence of a discontinuity produces the so-called Gibbs phenomenon that brings unwanted oscillations and, at the same time, lowers the convergence rate of the series, not only at the discontinuity point but also over the entire interval. In general, any singularity affects the approach. This aspect has been studied for the trigonometric expansion by Lighthill et al. (1958) and is also applicable to other basis functions by virtue of the Darboux (1878) which allows to state that the rate of convergence in a real domain of the series expansion of a function depends on the location on the complex plane of the singularities and their gravity (Boyd 2000).

Certain times series, such as river discharges and precipitation in semiarid basins show strong time variations that reflect themselves as sudden changes on the time dependent empirical distribution. Also, the slopes of the trends of the percentiles at the extremes of the interval are usually not equal. Under these conditions it is expected that the trigonometric basis functions fail to reproduce the overall behavior. In this context it seems important to choose a suitable set of basis functions to minimize this inconvenience. Moreover, when the statistical characterization of several time series needs to be done (for example as a first step to characterize a spatial temporal random field), the choice also needs to attend for the dimension of the optimization problem. Intuitively, the similarity between a function and its best approach, depends on the shape of the functions of the basis (Mead and Delves 1973). In fact, the behavior of the basis functions at the boundaries of the interval determines the rate of convergence of the series expansion. Apart from this fact, there is a lack of knowledge that might serve as a guide for the choice of the basis for which a good approximation is faster and, accordingly, the dimension of the optimization problem is smaller.

In this work we propose a general procedure that is based on the research line initiated by Solari and Losada (2011). It uses non-stationary piecewise functions for the marginal distributions of the vector components. The theoretical probability models are fitted to data by solving a constrained optimization problem where the negative log-

likelihood function (NLLF) is used as the objective function. We explore with three different environmental time series the adequacy of probability models and basis functions to reproduce the statistical behavior of the data.

The novelties of the present formulation with respect to the abovementioned contributions are:

- Previous works (see e.g. Solari and Losada (2011); Solari and Van Gelder (2011)) use 2 or 3 intervals with specific probability models (e.g. a lognormal or a Weibull for the central part and two generalized Pareto for the tails), and the expression of the probability density function (pdf) is given in terms of the relationships between the parameters of those particular models, obtained from the continuity conditions imposed on the pdf and a restriction on the support of the model selected for the lower tail. The proposed procedure is a general formulation, valid for any number of intervals and any combination of continuous probability models. The restrictions on the sample space, if required, are imposed as constrictions in the optimization problem. The model is capable to detect whether a smaller number of intervals (or probability models) are needed as it gives a partition of the real axis with very close, almost indistinguishable values.
- In regard to the non-stationary characterization, existing works (see e.g. Solari and Losada (2011, 2014)) use the trigonometric expansion while we propose the use of the best approach in any subspace spanned by a set of a orthogonal basis functions (generalized Fourier expansion). This set can be, among others, the functions that arise in the periodic Sturm Liouville problem (SLP) as in Solari and Losada (2011) and the eigenfunctions of SLPs. Moreover, instead of taking the year as the reference time interval, the expansion can be done over an arbitrary integer number of years.

The article is organized as follows. Section 2 presents the theoretical foundations of the methodology. Section 3 illustrates its application to three time series with different particularities. In Section 3.1 is analyzed the daily mean precipitation projected over the period 2006-2100 in a location of a semiarid basin with a clear time variability with two main seasons. Section 3.2 shows the results of the analysis of Wolf or Zurich sunspot number time series where time variability expands over several years. Further on, Section 3.3 also shows the goodness of the methodology for simulation purposes with data from a bivariate vector random process. This series includes the freshwater river discharges at the last regulation point of a river and the salinity at the river mouth. In Section 4 some of the key points of the methodology are discussed, including its advantages regarding existing methods and, finally, Section 5 concludes the study.

2 Theoretical background

We consider a vector random process, $\vec{X} = (X_1(t), \dots, X_i(t), \dots, X_N(t))$, that can be multivariate or univariate (for $N=1$), where t belongs to a certain set of index, and a matrix that contains N_o observations made at discrete values t_j : $\vec{x}^o(t_j) = (x_1^o(t_j), \dots, x_i^o(t_j), \dots, x_N^o(t_j))$. Because t is usually time, for the sake of simplicity, from now on we will speak about time series, and we will assume that the random process is observed at equally spaced instants.

The characterization of \mathbf{X} includes the fit of the marginal NS distribution functions of each random variable X_i . This information can be used to simulate NS multivariate time series. In this work, we used a vectorial autoregressive model (VAR) as described in Lütkepohl (2005) to obtain realizations and to assess with them the goodness of fit of VRPs.

2.1 Fit of data to marginal NS distributions

We assume that each variable X_i ($i = 1, \dots, N$), from now on denoted by X , is a continuous random variable whose probability density function $f_X(x)$ can be expressed as a piecewise function where a finite number, N_I , of weighted probability models (PMs) fit within a partition of the real axis into intervals: $\{I_\alpha : \alpha = 1, \dots, N_I\}$ where $I_\alpha = (u_{\alpha-1}, u_\alpha]$ for $j = 2, \dots, N_I - 1$, $I_1 = (-\infty, u_1]$ and $I_{N_I} = (u_{N_I-1}, +\infty)$. That is:

$$f_X(x) = \begin{cases} \omega_1 f_1(x) & x \leq u_1 \\ \omega_2 f_2(x) & u_1 < x \leq u_2 \\ \dots & \dots \\ \omega_\alpha f_\alpha(x) & u_{\alpha-1} < x \leq u_\alpha \\ \dots & \dots \\ \omega_{N_I} f_{N_I}(x) & u_{N_I-1} \leq x \end{cases} \quad (1)$$

where f_α denotes the probability density function of the model selected for I_α . The function defined in eq. (1) is required to be continuous at the common matching points of the intervals by imposing the following conditions:

$$\omega_\alpha f_\alpha(u_\alpha) = \omega_{\alpha+1} f_{\alpha+1}(u_\alpha), \alpha = 1, \dots, N_I - 1 \quad (2)$$

Also, in order to guarantee that eq. (1) is well defined, the parameters are required to fulfil the following condition:

$$\begin{aligned} &\omega_1 F_1(u_1) + \dots + \omega_\alpha (F_\alpha(u_\alpha) - F_\alpha(u_{\alpha-1})) + \dots \\ &+ \omega_{N_I} (1 - F_{N_I}(u_{N_I-1})) \\ &= 1 \end{aligned} \quad (3)$$

where F_α denotes the corresponding probability distribution function.

The solution to eqs. (2) and (3) is:

$$\omega_x = \frac{a_1}{b_1} \dots \frac{a_{x-1}}{b_{x-1}} \left[c_1 + c_2 \frac{b_1}{a_1} + c_3 \frac{b_1 b_2}{a_1 a_2} + \dots + c_x \frac{b_1 b_2 \dots b_{x-1}}{a_1 a_2 \dots a_{x-1}} + \dots + c_{N_x} \frac{b_1 b_2 \dots b_{N_x-1}}{a_1 a_2 \dots a_{N_x-1}} \right]^{-1} \quad (4)$$

where $a_x = f_x(u_x)$, $b_x = f_{x+1}(u_x)$ and $c_x = F_x(u_x) - F_x(u_{x-1})$, provided that a_x and b_x and the denominator in eq. (4) are both different from zero.

In eq. (1), the parameters of the distributions are assumed to be unknown time dependent functions which largest periodic variation is N_y years. Any of these functions, generically denoted by $a(t)$, can be expanded into a Generalized Fourier series over the interval $[0, N_y]$ which expression, truncated to N_F terms, is:

$$a(t) \approx \sum_{n=1}^{N_F} a_n \phi_n(t) \quad (5)$$

$t \in [0, N_y]$,

where a_n are the coefficients of the best approach in the subspace spanned by a set of orthogonal functions, $\{\phi_n(t)\}_{n=1}^{N_F}$. This set may be, among others, the set of eigenfunctions of a Sturm Liouville problem (SLP) with ordinary differential equation:

$$\frac{d}{dt} \left(p(t) \frac{d\phi}{dt} \right) + (\lambda w(t) - q(t)) \phi(t) = 0, \quad (6)$$

where $p(t)$, $\omega(t) > 0$ and $p(t)$, dp/dt , $w(t)$ and $q(t)$ are continuous functions over the interval $[0, N_y]$.

The orthogonality is interpreted in regards to the inner product $\langle f(t), g(t) \rangle = \int_a^b \omega(t) f(t) g(t) dt$. Table 1 presents some plausible sets for series expansion that can be used with the appropriate linear transformation of the domain into $[0, N_y]$.

The negative log-likelihood function (NLLF) is used as the objective function in the optimization algorithm. It reads:

$$\text{NLLF}(\vec{\zeta}) = - \sum_{j=1}^{N_o} \log f(x^o(t_j); \vec{\zeta}), \quad (7)$$

where $\vec{\zeta}$ is a vector of dimension N_d that contains the Fourier coefficients of the expansion of the parameters and the percentiles of the common matching points, and $x^o(t_j)$ for $j = 1, \dots, N_o$ are the observations.

The optimization problem is defined as the search for values of ζ that minimize the NLLF. When necessary, the optimization problem will be subject to conditions imposed on the sign of certain parameters of the distributions involved. An approximation of the solution is found by means of the Sequential Least Squares Programming (SLSQP) (Von Stryk 1993), and by using as initial solution a first guess of the values of the coefficients obtained from stationary conditions and also a guess of the percentiles of the common endpoints of the intervals.

The resulting distributions where the parameters are those obtained from the optimization problem, are NS and, therefore, hereinafter denoted by $F_{X_i}(x^o(t); t)$ for each X_i .

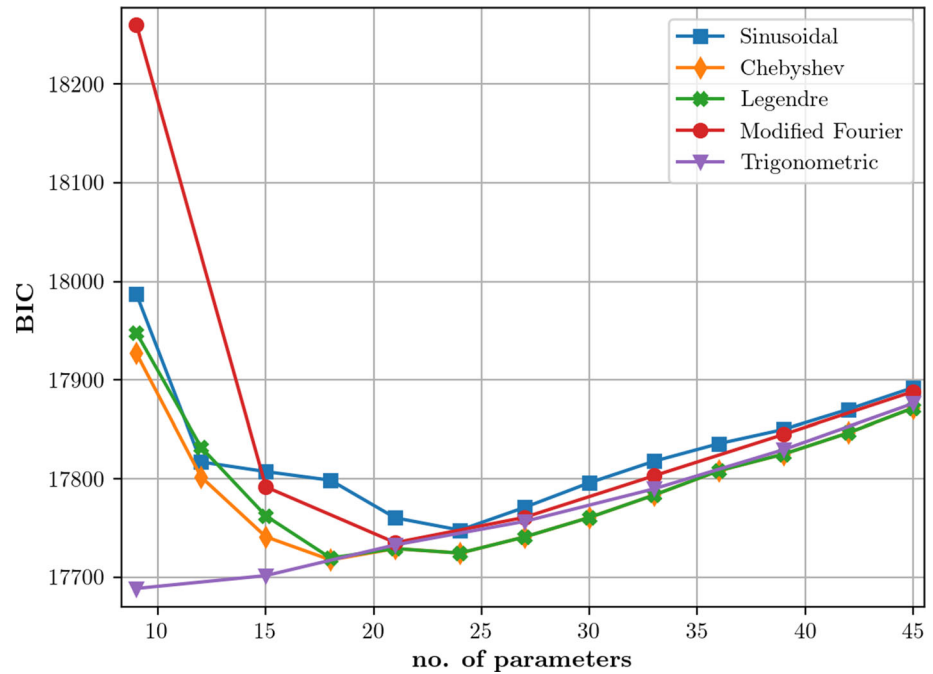
3 Application to climate time series

In the following subsections, the results of the application of the method to different time series is presented. Two univariate time series and a multivariate one is analyzed. The first one shows a significant yearly cycle and a strong variability of the range of values along the year. The second one presents marked 22- and 11-year periodicities and rather clear shorter terms. Finally, the analysis focus on a

Table 1 Sets of basis expansion solutions (first column) that solves the differential eq. (6) with the functions (second column) and conditions (third column) given

Orthogonal set of eigenfunctions	Differential equation and domain	Conditions imposed
Trigonometric series expansion $\{1\} \cup \{\cos(2n\pi t), \sin(2n\pi t) : n = 1, 2, \dots\}$	$p(t) = q(t) = w(t) = 1, t \in [-1, 1]$	Periodic SLP: $\phi(-1) = \phi(1)$ and $\phi'(-1) = \phi'(1)$
Modified Fourier series expansion $\{1\} \cup \{\cos(n\pi t), \sin((n - \frac{1}{2})\pi t) : n = 1, 2, \dots\}$	$p(t) = q(t) = 1, t \in [-1, 1]$	Regular SLP: $\phi'(-1) = \phi'(1) = 0$
Sinusoidal series expansion $\{\sin(n\pi t) : n = 1, 2, \dots\}$	$p(t) = q(t) = w(t) = 1, t \in [0, 1]$	Regular SLP: $\phi(0) = \phi(1) = 0$
Legendre polynomials $\{P_0 = 1\} \cup \{P_1 = t\} \cup \{P_n : nP_n = (2n - 1)tP_{n-1} - (n - 1)P_{n-2}, n = 2, 3, \dots\}$	$p(t) = w(t) = 1,$ $q(t) = -1/4 \tan^2(\pi t/2) - 1/2, t \in [-1, 1]$	Regular SLP: $\phi(-1)$ and $\phi(1)$ are finite
Chebyshev polynomials $\{P_0 = 1\} \cup \{P_1 = t\} \cup \{P_n : P_n = 2tP_{n-1} - P_{n-2}, n = 2, 3, \dots\}$	$p(t) = w(t) = 1,$ $q(t) = -1/4 \tan^2(\pi t/2)$ $-3/4 \sec^2(\pi t/2) \sin^2(\pi t/2) - 1, t \in [-1, 1]$	Regular SLP: $\phi(-1)$ and $\phi(1)$ are finite

Fig. 1 Optimum value, BIC, versus the number of parameters for the marginal fit with different choices for the time expansions of the parameters of the PMs for daily precipitation projected at Sierra Nevada location



bivariate time series that links a variable with a strong variability along the year due to climate variations and management decisions with a time series that is also influenced by other physical processes.

3.1 Precipitation at Sierra Nevada mountain (Andalusia, Spain)

This first application is devoted to an univariate time series, hereinafter denoted by $P(t)$, which stands for the daily mean precipitation projected at the position (3.546°W - 36.706°N) in Sierra Nevada (Andalusia, Spain) from 2006 to 2100. Data comes from EUROCORDEX project and has been obtained with the climate model SMHI-CNRM-CERFACS-CNRM-CM5 for a Representative Concentration Pathway RCP4.5 scenario. The point is located at the Guadalfeo river watershed, an area of semiarid Mediterranean climate where precipitation events are scarce and usually torrential, mainly concentrated during the period ranging from October to April. Due to this behavior, the empirical distribution function obtained by taking the year as the reference period (see dots in Figure 2), shows steep changes close to the end of April and at the end of September. The curves also have marked peaks at the beginning and at the end of the year and, therefore, the trends at the limits of the intervals have different slopes. To deal with this high variability, a Box-Cox data transformation with $\lambda = -0.1186$ parameter was used.

Several combinations of PMs such as Normal - Weibull of maxima, Log-normal - Normal, Normal - Generalized Pareto, with different initial guess of the percentiles of the

threshold, as well as single models like Weibull of maxima, Log-normal or Normal were used for testing. The best visual fits were obtained for a Weibull of maxima distribution. In addition, when trying the fit with more than one distribution, for all those combinations where this distribution was one of the PMs, the percentiles of the final support of this PM were almost 0 or 1. This indicates that the methodology is capable to distinguish when a single PM works adequately for all the range of values and when it is worth to skip needless PMs. The performance of different sets of basis functions is analyzed for all the expansions included in Table 1 in terms of the dimension of the optimization problem (N_d) and the BIC (Schwarz 1978) (see Figure 1), which is related to the optimum value, NLLF* and N_d through the mathematical expression $\text{BIC} = 2\text{NLLF}^* + \log(N_o)N_d$.

It can be observed that for a small number of parameters, the best approach in terms of the BIC is obtained with the trigonometric expansion. As the number of terms in the series increase, the differences between the approaches become smaller. For larger dimensions of the optimization problem, the other expansions show minima at values ranging from 18 to 24 parameters.

Figure 1 compares the empirical distribution with some of the theoretical ones obtained with the expansions of the parameters of the distribution for four of the sets in Table 1. For all of them the BIC is close to the minimum, $N_d = 21$. From panels a) to d), it includes the Legendre's polynomial approximation up to degree 7, the sinusoidal with 7 terms, and ultimately the modified Fourier and the trigonometric with 3 oscillatory components. A logarithmic

scale has been used for the vertical axis so that the goodness of the fits can be clearly visualised for all the percentiles. Despite they all have the same number of parameters and the similarities in the BIC values, the

expansion that gives an overall better fit with smoother curves is Legendre's. All the basis are capable to give rather accurate and similar descriptions of the behavior of the lower, intermediate and upper percentiles. However, for

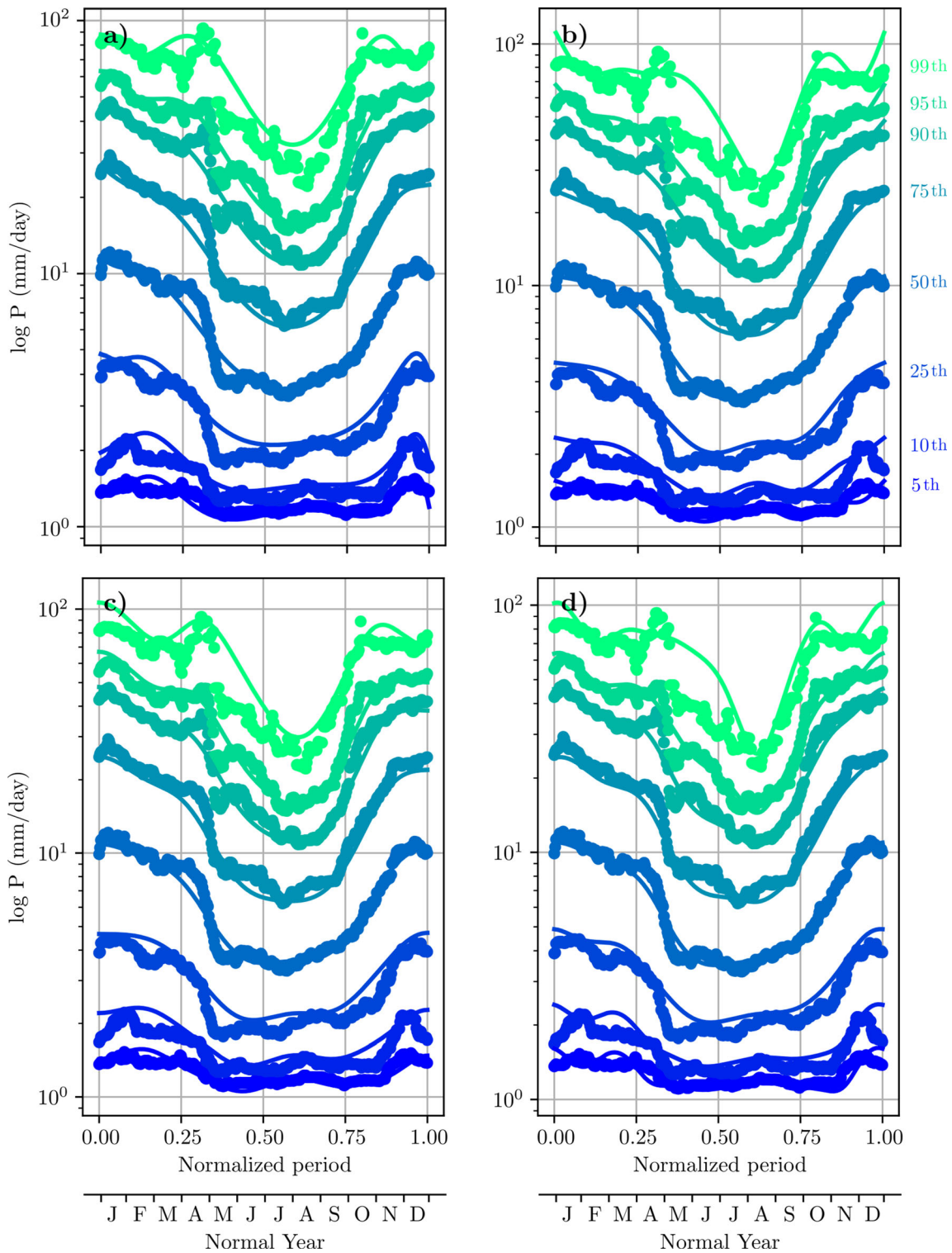


Fig. 2 Non-stationary empirical CDF of the precipitation and theoretical model fits for different choices of the basis functions for the time expansion of the parameters of a Weibull of maxima. *a)* Legendre polynomials; *b)* Sinusoidal; *c)* Modified Fourier and *d)* Trigonometric

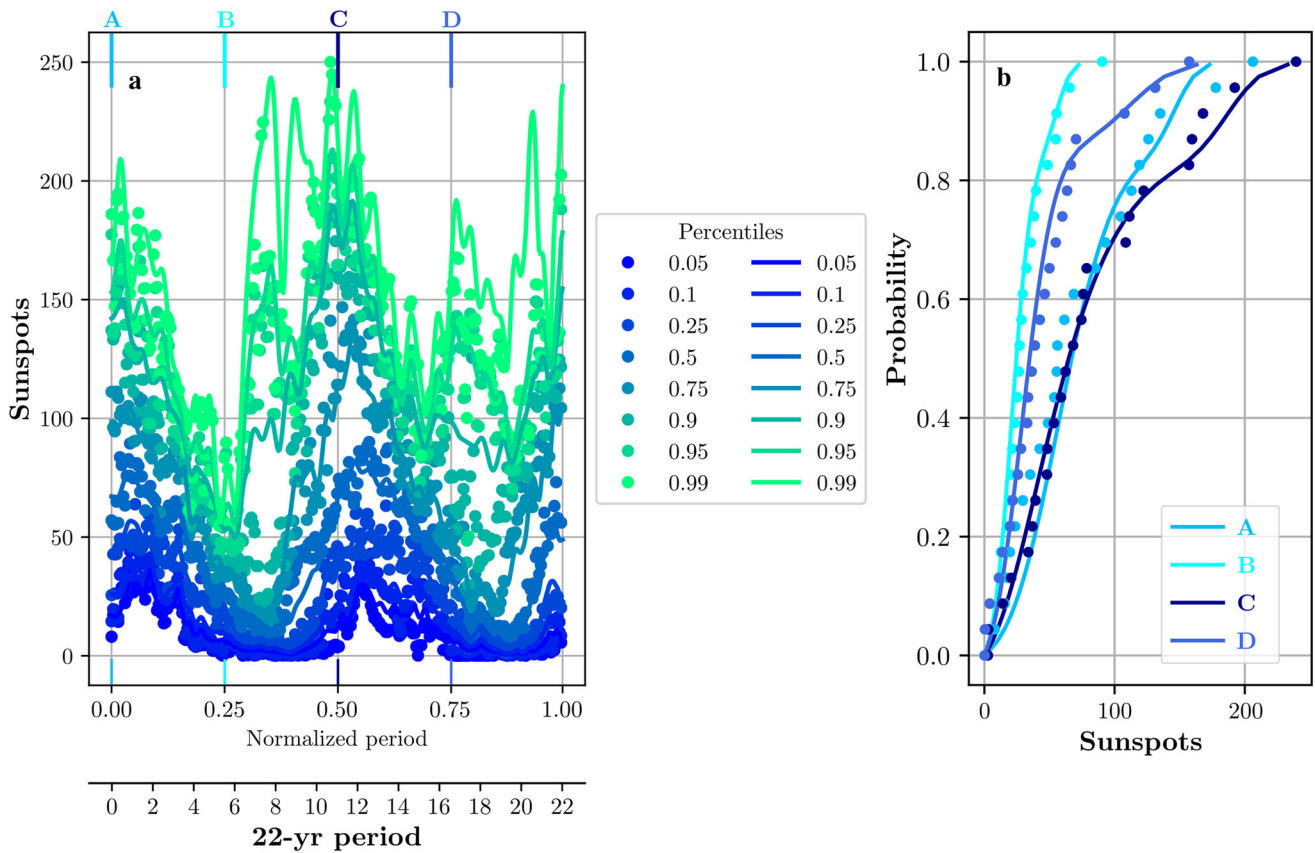


Fig. 3 a) Non-Stationary CDF of sunspots, and b) stationary cumulative distribution functions at sections given in panel a)

the rapid changes happening between April and May and between September and October, they show slightly different behaviours. On the one hand, for the largest percentiles and for the period between April and May, the sudden change is better captured by the modified Fourier one while the trigonometric one is not so good at this steep transition. On the other hand, it is the trigonometric basis which better reproduce the rapid variations in precipitation from September to October. Regarding the infraestimation / overestimation at the upper percentiles, the Legendre basis is the one that more fairly reproduce the magnitude of the precipitation of these extreme events, while the trigonometric and sinusoidal approaches overestimate this magnitude.

3.2 Wolf sunspot number

In the second example, we analyze the monthly time series of Wolf or Zurich sunspot number, available from 1749 (Source: WDC-SILSO, Royal Observatory of Belgium, Brussels). The signal contains the well-known 11 years Schwabe cycle and also the 22 years one described in Usoskin and Mursula (2003).

In order to detect the time random variability up to the seasonal scale, a basic period of $N_y = 22$ years is taken for the analysis. A piecewise function composed of two PMs, a log-normal and a normal, were used in eq. (1). Several initial guesses were tried as the percentiles of the common matching points and the final values always were close to 0.85.

Figure 3 shows the fit with a sinusoidal expansion retaining $N_F = 44$ terms (covering frequencies ranging from $1/44$ to 1 yr^{-1}) that was the option that gave similar values of the optimum NLLF and the BIC with a considerable smaller number of parameters (442 versus more than 600). In this example, it is highlighted that the minimum BIC is found for $N_F = 6$, which means that the minimum oscillatory period included in the analysis would be $22/3$ years. However, as it is known that the annual component is significant, we force the analysis to optimize up to 1 yr^{-1} . No Box-Cox transformation was required for the analysis. As it is observed, all the percentiles show a peak associated to the 11 years cycle which is asymmetric as pointed out by Usoskin and Mursula (2003), who detected that it has a shorter ascending phase and a longer descending phase. This asymmetry is particularly visible in the lower percentiles. The upper tails show two additional peaks that are

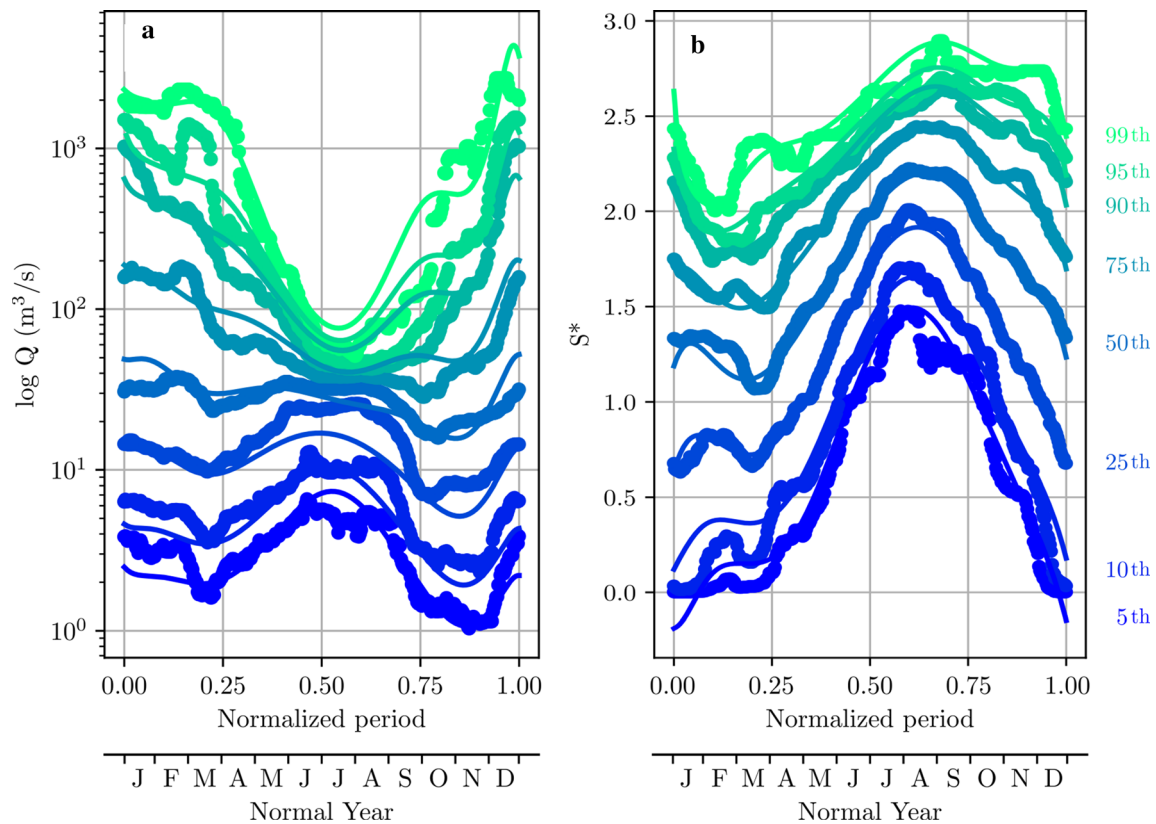


Fig. 4 a) Non-Stationary Cumulative Distribution Function of river discharge at Alcalá del Río and b) and the Yeo-Johnson transformation of salinity at the river mouth

related to the 7.5 to about 17 years also mentioned in Usoskin and Mursula (2003).

In Figure 3.b it is shown the empirical and theoretical stationary cumulative distribution functions at sections A to D indicated in panel (a) of the same figure. This graph allows to observe not only the goodness of fit of the theoretical model but also the capability of the theoretical PMs to distinguish the behavior of the body and the upper tail.

3.3 Fresh-water river discharge and salinity at the Guadalquivir river estuary

The third example analyzes the bivariate time series of the following variables: (a) the fresh-water mean daily river discharge ($Q(t)$) at Alcalá del Río dam ($6.06^\circ\text{W} - 37.29^\circ\text{N}$), the last regulation point of the Guadalquivir river estuary, and (b) the mean daily sea water salinity ($S(t)$) at its mouth ($6.5^\circ\text{W} - 36.83^\circ\text{N}$) at 0.5 meters depth from SWL. The time series of $Q(t)$ is available from July 1st, 1931 to April 27th, 2016 (Source: Andalusian Water Agency, Junta de Andalucía). At the mouth, the time series of S was obtained from Marine Copernicus service, specifically, the IBI MULTI-YEAR PHY 005 002 TDS ocean reanalysis service and cmems_mod_ibi_phy_my_0.083deg-3D_P1D-m product.

In this case, it ranges from January 1st of 1993 to December 31st of 2019. The regulation of this dam is aimed not only at controlling floods but also at fulfilling, among others, the following management objectives: i) the maintenance of an ecological river discharge, ii) the avoidance of unwanted turbidity conditions (Cobos et al. 2020; Díez-Minguito and de Swart 2020), and iii) the maintenance of $S(t)$ below a given threshold for the irrigation of rice crops in the estuary (Cobos 2020). As a result, $Q(t)$ varies from very low values (usually in summer $Q < 40 \text{ m}^3/\text{s}$) to those that are almost squared in winter ($Q \approx 1000 \text{ m}^3/\text{s}$) with sporadic sudden changes. Salinity variations are also related to sun radiation and variations associated to spring-neap tidal conditions.

The univariate analysis of Q and S were carried out with a Generalized Extreme Value function. For Q and S , the Chebyshev and Legendre expansions were performed, respectively. In both cases, a basic period of one year ($N_y = 1$) with degree equal to twelve for Q and S were used. A Box-Cox transformation (Box and Cox 1964) was required for the analysis of Q with $\lambda = -6.84 \cdot 10^{-3}$ and a Yeo-Johnson (Yeo and Johnson 2000) with $\lambda = 20.869$ to S . Figure 4.a and .b shows the marginal fits of the two RPs.

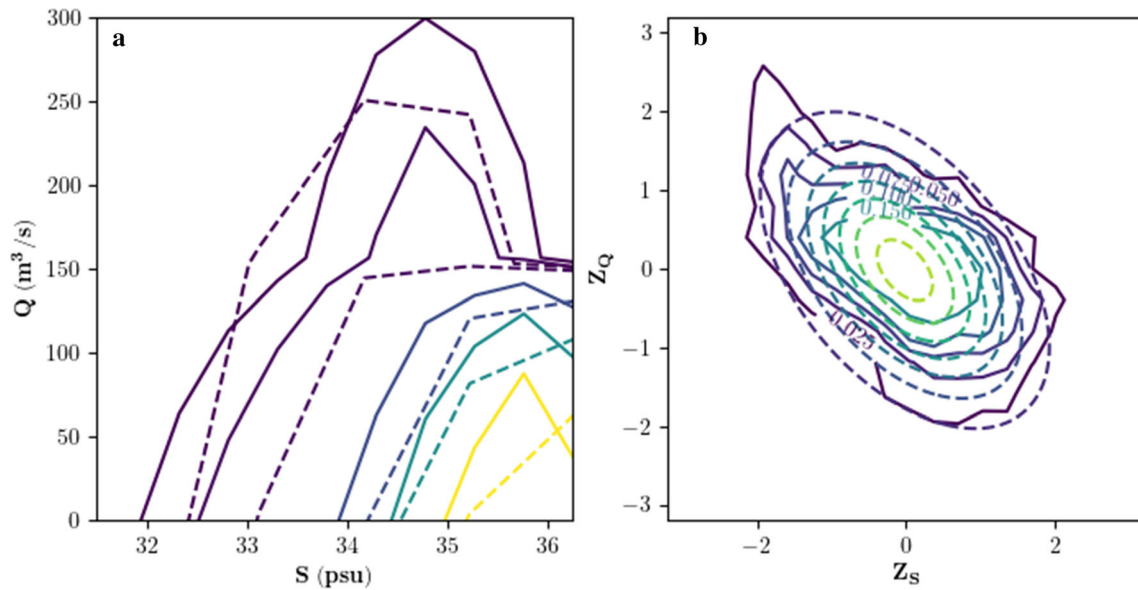


Fig. 5 Comparison of joint distribution. *a*) Joint distribution of of Q and S for observations (solid lines) and one of the random simulations (dashed lines). *b*) Joint distribution of normalised variables for observations (solid lines) and theoretical Gaussian distribution (dashed lines)

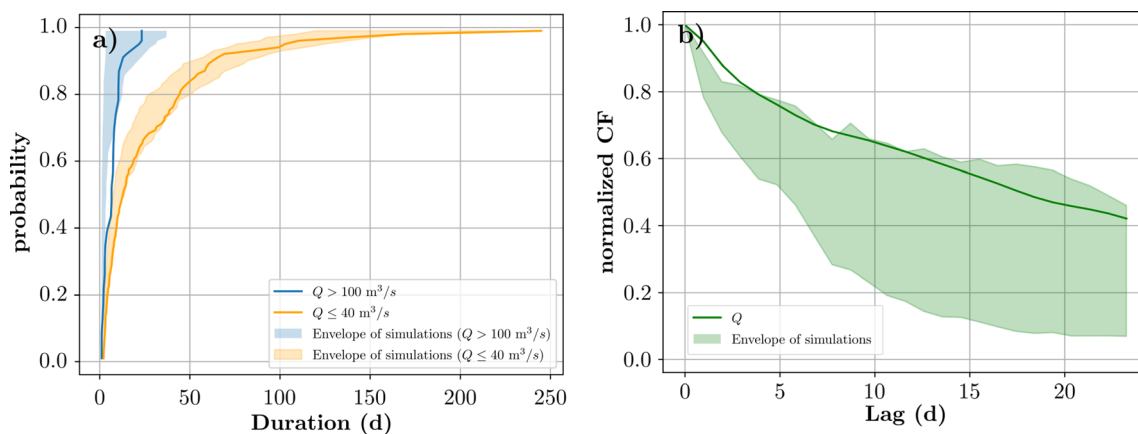


Fig. 6 *a*) CDFs of sojourn durations below $40\text{m}^3/\text{s}$ and above $100\text{m}^3/\text{s}$ and *b*) Autocorrelation function

As it observed, the models adequately reproduce the non-stationary pattern.

A VAR(6) model (see Appendix) was fit to the data from the common period between both series. With those results and the marginal distributions, 100 simulations were obtained in order to verify the goodness of the method with the methodology proposed by Monbet et al. (2007).

The joint distribution of river discharge and salinity is assessed in figure 5 where the joint density functions of observations and one of the simulations are compared in panel a). Panel b) shows the pdf of the normalized variables (eq. (8)) obtained with observations and a theoretical gaussian bivariate fit. The pdf of the normalized data resembles a standardized bivariate gaussian density function with a correlation coefficient $\rho = -0.75$ indicating that the VAR assumption regarding the gaussian behavior

is valid. The pdf of the simulated data shows a bump rather similar to the original data but with smaller values for the modal points. The correlation coefficient obtained with the values of those functions is $R^2 = 0.839$, which shows that there is a good agreement in bivariate distributions between the simulation and the original time series.

Finally, in figure 6 the estimations of the distributions of the sojourns durations below/above $40\text{m}^3/\text{s}$ and $100\text{m}^3/\text{s}$ obtained for the original series and the simulations are compared in panel a). These levels, according to Díez-Minguito et al. (2012, 2014), correspond to critical states of the estuary. Indeed, under low-river flow conditions ($Q < 40\text{m}^3/\text{s}$) the estuary is tidally dominated and turbidity and hypoxia events occur. Discharges with $Q > 100\text{m}^3/\text{s}$ helps water renovation, promotes life in the estuary and lowers the salinity values to acceptable levels for rice crops

(Cobos 2020). These plots give information about the persistence of extreme events in this particular environment, and, as pointed out by Monbet et al. (2007), are strongly related to the capability of the models to reproduce the severity of the conditions. The figures include the curves of the observations as well as an envelope band with the minimum and maximum values of the simulations. It is found that the model is capable to fairly reproduce the duration of both types of events. The autocorrelation function shown in panel b) shows a similar behavior consisting in a decreasing trend with smaller lags than the observations. This might be related to the influence of management decisions on river discharges that are not only related to climate conditions and that the VAR model is not capable to capture. Conversely, the salinity shows a high value (higher than 0.99) in both cases, which means that a almost perfect correlation is found. This behavior is the expected since the governing process that modify the salinity is relate to short time variations, i.e., tidal frequency M2 (12.42 hr). The strong water discharge also modified the salinity pattern, however these strong events are rare.

4 Discussion

The temporal description of the parameters (eq. 6) has been done in terms of SLPs. However, the expansion may also be the orthogonal projection of $a(t)$ in a subspace of any Hilbert function space of finite dimension. Among others, it can be the best polynomial approach of degree $N_F - 1$ by virtue of the Weierstrass theorem, that can be obtained with any set of orthogonal polynomials defined over bounded intervals such as Jacobi and Gegenbauer (that generalize Legendre and Chebyshev polynomials). In the examples shown in this work, oscillatory functions were used because climate forced time series have intrinsic oscillations that can be directly associated to the terms in the expansion. The consideration of alternative functions to the commonly used trigonometric basis is found to be particularly useful for the description of large dimension multivariate time series like those usually needed in coastal engineering, as the number of coefficients used in the approach can be significantly reduced. This is the case for the analysis of time series measurement projections of joint wave and wind climate conditions. It must be noticed that the better the fit of the marginal NS distributions, the better the temporal dependency obtained and, consequently, more accurately representative new random realizations would be obtained.

For some climate variables such as sea level, the oscillatory behavior is governed by some well-known periods associated to the gravitational attraction on the Earth by the

Sun and the Moon. In these cases, it is also possible to use a harmonic expansion of the time series with the identified significant periods, in a similar way than for tidal analysis (Pawlowicz et al. 2002; Codiga 2011).

The optimization problem increases its dimension with the number of PMs chosen in eq. (1) in a geometric progression, making the analysis impractical. To the authors experience, the selection of three PM's is usually enough to describe the central body as well as the lower and upper tails. The use of Generalized Pareto PMs for modeling the tails is highly recommended to properly simulate the higher and lower values of the variables. In applications where the interest is focused on the exceedances over a threshold, as it is the case for many engineering studies, the discretization in three regimes and the use of those PMs fairly reproduces the body and the upper tail. In addition, and following the suggestions given by Lira-Loarca et al. (2020); Jäger et al. (2019), some physical conditions might limit the event space, for example the wave height in shallow waters due to breaking. In those cases, it should be convenient to impose constrictions in the optimization problem.

The selection of the basis period for the analysis depends on the length of the available time series. The choice of the year does not allow to capture the longer-term variations described by climatic oscillations that have indeed shown to be relevant in the solar activity that strongly affects climate. It is important to note that when the chosen base period is larger than one year, the initial date for the simulation must be properly set-up so that the phase of the larger scale variability obtained is coherent with the original data.

A Python tool that guide users along all the steps required for making the NS analysis for VRPs and the simulation can be found in <https://github.com/gdfa-ugr/marinetools> (Cobos et al. 2022).

5 Conclusions

We have proposed a general procedure for the NS analysis of a random processes. It uses a NS piecewise function whose parameters and common endpoints are allowed to vary periodically in time over a certain number of years. That time dependence is described with the best approach in the subspace spanned by a subset containing a finite amount of eigenfunctions of a SLP. The parameters of the theoretical PMs are fitted to data by solving a constrained optimization problem where the NLLF is the objective function and, if needed, constrains are imposed on the sign of the parameters due to the intrinsic nature of the variables.

The novelty of this procedure, brings up some advantages with respect to previous works. First, from a mathematical point of view, the general formulation allows to extend the definition of the piecewise density function to any type of data sets. It is highlighted the importance of the selection of the appropriate sets of basis functions which might also significantly reduce the dimension of the optimization problem. Finally, the treatment of an arbitrary integer number of years makes possible to explore the presence of pluriannual cycles of variation whenever a large enough period of time is available for the analysis.

The application of the method to three time series with different particularities shows its goodness to reproduce the stochastic features of the original data for processes of different nature, being able to identify the appropriate values of the partition of the real axis and whether any of the models at the outer intervals is strictly necessary. More precisely, it is shown that it is capable to capture the highly variable precipitation projected at a mountainous environment with a semiarid climate where two main seasons are clearly observed. It is also found that it can capture a wide range of time scale variations already known along a 22 years cycle for the Wolf sunspot number time series, such as the Schwabe cycle and oscillations that vary close to 7.5 and 17 years. Finally, the joint variation of river discharges at the last point of regulation and the salinity at the river mouth is analyzed. The dam is located in a semiarid zone in Andalucía (Spain) and its regulatory activities depend not only on seasonal and yearly time climate variability but also on management decisions. The salinity at the mouth of the estuary is strongly related to river discharges and also to other processes such as tidal propagation and sun radiation. The application of the method combined with a VAR model to that bivariate data shows its capability to reproduce different statistical properties inferred from the original series such as the autocorrelation, the marginal and joint distributions and the duration of sojourns below/above given thresholds.

A. Appendix

The following sections show some methods that are used in this work in order to ease the analysis and simulate new random realizations of the VRPs.

A.1. Pretreatment of time series with persistent low values

Some climate related time series usually show very large differences between the smaller and the larger values. This is the case of river discharges at dams that regulate rivers in semiarid zones where most of the time the flow is the

minimum ecological discharge. Those low values are exceptionally exceeded when intense and persistent precipitation events occur and the dam releases for safety purposes. Those differences are also not evenly distributed along time due to, for example, to strong seasonal and yearly climate variation. Under such circumstances, it is convenient to transform the data into Gaussian distributed values using a λ -parameter Box-Cox transformation (Box and Cox 1964). Other power transformations can also be applied (Yeo and Johnson 2000).

A.2. Temporal dependence

The Vector Auto-regressive, VAR(q) model is applied to the normalized series obtained from the observations as:

$$Z_{X_i}(t_j) = \Phi^{-1} [F_{X_i}(x^o(t_j); t_j)] \tag{8}$$

where Φ^{-1} is the inverse of the Gaussian cumulative distribution function with zero mean and unit standard deviation and $F_{X_i}(x^o(t); t)$ is the NS probability distribution function of X_i .

We denote the values of the normalized series (eq. 8) at time t_j as $y_j^i = Z_{X_i}(t_j)$ and $Y_j = (y_j^1, \dots, y_j^i, \dots, y_j^N)^T$ where T stands for the vector transposition. The dependence in time between variables in the VAR(q) model is given by:

$$Y_j = c + A^1 Y_{j-1} + A^2 Y_{j-2} + \dots + A^q Y_{j-q} + e_j \tag{9}$$

where $c = (c_1, \dots, c^i, \dots, c_N)^T$ contains the mean values of the variables, $A^m, m = 1, \dots, q$ are the $N \times N$ coefficients matrices and $e_j = (e_j^1, \dots, e_j^i, \dots, e_j^N)^T$ is the vector with the white noise error terms. Using eq. (9) to relate data at an instant t_j to their previous q values, for $j = q + 1, \dots, N_o$, we obtain $Y = A\chi + E$, where $Y = (Y_{q+1} Y_{q+2} \dots Y_N)$, $\chi = (\chi_{q+1} \chi_{q+2} \dots \chi_N)$, with $\chi_j = (1 Y_{j-1}^T \dots Y_{j-q}^T)^T$, $A = (A^1 A^2 \dots A^q)$ and $E = (e_{q+1} e_{q+2} \dots e_N)$.

The solution is obtained by means of minimum least square errors as $A = Y\chi^T(\chi\chi^T)^{-1}$, where $E = Y - A\chi$ and $Q = cov(E)$ is the covariance matrix of the error. A detailed description can be found e.g. in Lütkepohl (2005).

A.3. Simulation

In order to obtain a realization of the vector random process, the first q -values of the time series are obtained with a Monte Carlo simulation using a Gaussian multivariate distribution with mean vector c and the covariance matrix Q given in section 1. Then, the VAR model is used to generate a multivariate Gaussian stationary time series $(y^1(t), \dots, y^i(t), \dots, y^N(t))$ at regular time instants. The

corresponding non-stationary time series is then recovered by using the following transformation:

$$X_i(t) = F_{X_i}^{-1}[\Phi(y_i(t); t)] \quad (10)$$

Acknowledgements This work was performed within the framework of the following projects: (1) Programa Operativo FEDER de Andalucía (30BE61F301) and (2) Flooding and erosion works in coastal areas of Andalusia under a climate change scenario, funded by the Ministry of Agriculture, Livestock, Fisheries and Sustainable Development of the Junta de Andalucía [Contrat No. CONTR 2018 66984]. The first author is also indebted to Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía (POSTDOC_ 21_00724) which partially funded his work.

Author Contributions M. Cobos: Conceptualization, Methodology, Software, Writing - Original draft preparation. P. Otiñar*: Data curation, Software, Writing - Original draft preparation. P. Magaña: Data curation, Software. A. Baquerizo: Conceptualization, Methodology, Writing - Original draft preparation.

Funding Funding for open access charge: Universidad de Granada / CBUA.

Declarations

Conflicts of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baquerizo A, Losada MA (2008) Human interaction with large scale coastal morphological evolution. an assessment of the uncertainty. *Coastal Eng* 55:569–580
- Boccotti P (2000) *Wave mechanics for ocean engineering*, vol 64. Elsevier, Netherland
- Box GE, Cox DR (1964) An analysis of transformations. *J Royal Stat Soc: Ser B (Methodol)* 26:211–243
- Boyd J (2000) *Chebyshev and Fourier Spectral Methods*. Courier Corporation, North Chelmsford
- Callaghan D, Nielsen P, Short A, Ranasinghe R (2008) Statistical simulation of wave climate and extreme beach erosion. *Coastal Eng* 55:375–390. <https://doi.org/10.1016/j.coastaleng.2007.12.003>
- Christakos G (2000) *Modern spatiotemporal geostatistics*, vol 6. Oxford University Press, Walton
- Christakos G (2017) *Spatiotemporal random fields: theory and applications*. Elsevier, Netherland
- Cobos M (2020) A model to study the consequences of human actions in the Guadalquivir river estuary. Universidad de Granada, Spain
- Cobos M, Baquerizo A, Díez-Minguito M, Losada M (2020) A subtidal box model based on the longitudinal anomaly of potential energy for narrow estuaries an application to the Guadalquivir river estuary (sw Spain). *J Geophys Res: Oceans*. <https://doi.org/10.1029/2019JC015242>
- Cobos M, Lira-Loarca A, Christakos G, Baquerizo A (2019) Storm characterization using a bme approach. Springer, Cham, *Contributions to Statistics*. https://doi.org/10.1007/978-3-030-26036-1_19
- Cobos M, Otiñar P, Magaña P, Lira-Loarca A, Baquerizo A (2022) Marinertools temporal: a python package to simulate earth and environmental time series. *Environ Model Softw* 150:105359
- Codiga D (2011) Unified tidal analysis and prediction using the “UTide” Matlab functions. Graduate School of Oceanography, University of Rhode Island Technical Report. Tech. Rep. 2011-01
- Corbella S, Stretch DD (2012) Multivariate return periods of sea storms for coastal erosion risk assessment. *Nat Hazards Earth Syst Sci* 12:2699–2708
- Darboux G (1878) Mémoire sur l'approximation des fonctions de très-grands nombres, et sur une classe étendue de développements en série. *J de mathématiques pureset appliquées série* 4:5–56
- De Leo F, Besio G, Briganti R, Vanem E (2021) Non-stationary extreme value analysis of sea states based on linear trends analysis of annual maxima series of significant wave height and peak period in the mediterranean sea. *Coastal Eng* 167:103896. <https://doi.org/10.1016/j.coastaleng.2021.103896>
- De Luca DL, Galasso L (2018) Stationary and non-stationary frameworks for extreme rainfall time series in southern Italy. *Water* 10:1477
- De Michele C, Salvadori G, Passoni G, Vezzoli R (2007) A multivariate model of sea storms using copulas. *Coastal Eng* 54:734–751
- Díez-Minguito M, Baquerizo A, De Swart H, Losada M (2014) Structure of the turbidity field in the Guadalquivir estuary: Analysis of observations and a box model approach. *J Geophys Res: Oceans* 119:7190–7204
- Díez-Minguito M, Baquerizo A, Ortega-Sánchez M, Navarro G, Losada M (2012) Tide transformation in the Guadalquivir estuary (sw Spain) and process-based zonation. *J Geophys Res: Oceans*. <https://doi.org/10.1029/2011JC007344>
- Díez-Minguito M, de Swart HE (2020) Relationships between chlorophyll-a and suspended sediment concentration in a high-nutrient load estuary: an observational and idealized modeling approach. *J Geophys Res: Oceans*. <https://doi.org/10.1029/2019JC015188>
- Fedele F, Arena F (2009) The equivalent power storm model for long-term predictions of extreme wave events. In: ASME 2009 28th International Conference on Ocean, Offshore and Arctic Engineering, American Society of Mechanical Engineers, pp. 401–411
- Félix A, Baquerizo A, Santiago J, Losada M (2012) Coastal zone management with stochastic multi-criteria analysis. *J Environ Manage* 112:252–266. <https://doi.org/10.1016/j.jenvman.2012.05.033>
- He J, Christakos G, Wu J, Li M, Leng J (2021) Spatiotemporal bme characterization and mapping of sea surface chlorophyll in Chesapeake Bay (USA) using auxiliary sea surface temperature data. *Sci Total Environ* 794:148670

- He J, Kolovos A (2018) Bayesian maximum entropy approach and its applications: a review. *Stoch Environ Res Risk Assess* 32:859–877
- Hildebrandsson HH (1897) Quelques recherches sur les centres d'action de l'atmosphère. Norstedt Soner, Stockholm
- Izaguirre C, Mendez FJ, Menendez M, Luceño A, Losada IJ (2010) Extreme wave climate variability in southern europe using satellite data. *J Geophys Res: Oceans*. <https://doi.org/10.1029/2009JC005802>
- Jäger WS, Nagler T, Czado C, McCall RT (2019) A statistical simulation method for joint time series of non-stationary hourly wave parameters. *Coastal Eng* 146:14–31
- Jalón ML, Baquerizo A, Losada MA (2016) Optimization at different time scales for the design and management of an oscillating water column system. *Energy* 95:110–123
- Jonathan P, Ewans K (2013) Statistical modelling of extreme ocean environments for marine design: a review. *Ocean Eng* 62:91–109
- Le Mouél JL, Lopes F, Courtillot V (2019) A solar signature in many climate indices. *J Geophys Res: Atmos* 124:2600–2619. <https://doi.org/10.1029/2018JD028939>
- Lighthill MS, Lighthill MJ, Lighthill MJ et al (1958) *Int Fourier Anal Generalised funct.* Cambridge University Press, England
- Lira-Loarca A, Cobos M, Losada MA, Baquerizo A (2020) Storm characterization and simulation for damage evolution models of maritime structures. *Coastal Eng* 156:103620
- Loarca AL, Cobos M, Besio G, Baquerizo A (2021) Projected wave climate temporal variability due to climate change. *Stoch Environ Res Risk Assess* 35:1–17
- López-Ruiz A, Bergillos RJ, Lira-Loarca A, Ortega-Sánchez M (2018) A methodology for the long-term simulation and uncertainty analysis of the operational lifetime performance of wave energy converter arrays. *Energy* 153:126–135
- Luceño A, Menéndez M, Méndez FJ (2006) The effect of temporal dependence on the estimation of the frequency of extreme ocean climate events. *Proc Royal Soc A: Math, Phys Eng Sci* 462:1683–1697
- Lütkepohl H (2005) *New introduction to multiple time series analysis.* Springer Science Business Media, New York
- Mead K, Delves M (1973) On the convergence rate of generalized fourier expansions. *IMA J Appl Math* 12:247–259. <https://doi.org/10.1093/imamat/12.3.247>
- Méndez FJ, Menéndez M, Luceño A, Losada IJ (2006) Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. *J Geophys Res: Oceans* 111:561
- Méndez FJ, Menéndez M, Luceño A, Medina R, Graham NE (2008) Seasonality and duration in extreme value distributions of significant wave height. *Ocean Eng* 35:131–138
- Monbet V, Ailliot P, Prevosto M (2007) Survey of stochastic models for wind and sea state time series. *Probab eng mech* 22:113–126
- Pawlowicz R, Beardsley B, Lentz S (2002) Classical tidal harmonic analysis including error estimates in matlab using `t_tide`. *Compt Geosci* 28:929–937
- Payo A, Baquerizo A, Losada M (2004) Uncertainty assessment of long term shoreline prediction. In: *Proc. of the 29th Int. Conf. on Coastal Engineering'2004*. 2, 2087–2096
- Payo A, Baquerizo A, Losada M (2008) Uncertainty assessment: application to the shoreline. *J Hydraul Res* 46:96–104
- Ranasinghe RDC, Stive MJF (2012) Estimating coastal recession due to sea level rise: beyond the bruun rule. *Clim Change*. <https://doi.org/10.1007/s10584-011-0107-8>
- Razmi A, Golian S, Zahmatkesh Z (2017) Non-stationary frequency analysis of extreme water level: application of annual maximum series and peak-over threshold approaches. *Water Res Manage* 31:2065–2083
- Schwarz G (1978) Estimating the dimension of a model. *Annals Statist* 6:461–464
- Solari S, Losada M (2011) Non-stationary wave height climate modeling and simulation. *J Geophys Res: Oceans*. <https://doi.org/10.1029/2011JC007101>
- Solari S, Losada M (2011) Non-station wave height clim model simulation. *Oceans, J Geophys Res*. <https://doi.org/10.1029/2011JC007101>
- Solari S, Van Gelder P (2011) On the use of vector autoregressive (var) and regime switching var models for the simulation of sea and wind state parameters. *Marine Technol Engi* 1:217–230
- Usoskin I, Mursula K (2003) Long-term solar cycle evolution: review of recent developments. *Solar Phys* 218:319–343
- Von Stryk O (1993) *Numerical solution of optimal control problems by direct collocation.* Optimal control. Springer, New York, pp 129–143
- Walker G (1924) Correlations in seasonal variations of weather, ix, a further study of world weather (world weather ii). *Mem India Meteorol Depa* 24:275–332
- Walker GT, Bliss EW (1932) *World weather v.* Mem Roy Mete Soc 4:53–84
- Wu J, He J, Christakos G (2021) *Quantitative analysis and modeling of earth and environmental data: space-time and spacetime data considerations.* Elsevier, Netherland
- Yeo IK, Johnson RA (2000) A new family of power transformations to improve normality or symmetry. *Biometrika* 87:954–959
- Zhai Q (2017) Evidence for the effect of sunspot activity on the el niño/southern oscillation. *New Astronomy* 52:1–7. <https://doi.org/10.1016/j.newast.2016.09.004>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.