Highlights

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

M. Cobos, P. Otiñar, P. Magaña, A. Baquerizo

- A general solution for a continuous piecewise-defined non-stationary probability model is obtained.
- The time-varying component probability models are described by means of generalized Fourier series expansions.
- The goodness of the characterization and its utility for simulating purposes is assessed using joint probability distributions and sojourn durations of several realizations.
- The method, with a suitable vectorial autoregressive model, can be used to obtain new statistically equivalent multivariate time series for climate, Earth or environmental vector random process.
A method to characterize climate, Earth or environmental vector random processes

M. Cobos, P. Otiñar*, P. Magaña and A. Baquerizo

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

ARTICLE INFO

Keywords:
Generalized Fourier series of parameters
non-stationary probability models
piecewise continuous probability density functions
stochastic characterization
time series of environmental processes

ABSTRACT

We propose a methodology to characterize a multivariate non-stationary vector random process that can be used for simulating random realizations that keep the probabilistic behavior of the original time series. The marginal probability distribution of each component 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. The probability model is assumed to vary periodically in time over a predefined time period 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. Three different types of sets of orthogonal functions are tested. The method is applied to three time series with different particularities. Firstly, it is shown its good behavior to capture the highly variable freshwater discharges at a dam located in a semiarid zone in Andalucía (Spain) which is influenced not only by the climate variability but also by management decisions. 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 (velocity and direction) wind time series observed at a location of the Atlantic Ocean. 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, the wind rose, and the duration of sojourns above given thresholds.

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 Walker (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 Mouël 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 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 (Chris-
takos, 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 which 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 charac-
terize 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). Monbet et al. (2007) provided a
survey about the state of the art about models for wind and sea state time series characterization and simulation, as
well as a method to validate the ability of models to capture statistical features like marginal distributions, covariance
functions and sojourns durations above/below a threshold level for analyze stormy/calm conditions for a given dataset.

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 and a VAR
model to capture the multivariate time series dependence. The parameters of the distributions 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 number of orthogonal functions. This set can be, among others,
the trigonometric functions that arise in the periodic Sturm Liouville problem (SLP) as in Solari and Losada (2011)
and the eigenfunctions of regular SLPs. 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. If needed,
a set of constraints are imposed on the sign of the parameters due to the intrinsic nature of the variables.

The article is organized as follows. Section 2 presents the theoretical foundations of the methodology. Section 3
illustrates its application to three environmental time series with different particularities. In Section 3.1 is analyzed
the freshwater river discharge from a dam which watershed shows a strong seasonal variability and where rain events
are scarce and management decisions add complexity. In Section 3.2 are shown the results of the analysis of Wolf or
Zurich sunspot number time series where time variability expands to several years. Later, in Section 3.3, a bivariate
series that includes wind velocities and the circular variable that describes the incoming direction in a site where wind
varies significantly at a wide range of scales, shows also the goodness of the methodology for simulation purposes. In
Section 4 some of the key points of the methodology are discussed and, finally, Section 5 concludes the study.

2. Theoretical background

We consider a vector random process, \( \vec{X} = (X_1(t), ..., X_j(t), ..., X_N(t)) \), 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 \( t_j \):
\[
\vec{x}(t_j) = (x_1^o(t_j), ..., x_j^o(t_j), ..., x_N^o(t_j)).
\]
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 \( \vec{X} \) includes the fit of the marginal NS distribution functions of each random variable \( X_j \).

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) (see Appendix) 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 \( X_i \) (\( i = 1, \ldots, N \)), 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, \( N_f \), of weighted probability models (PMs) fit within a partition of the real axis into intervals: \( \{ I_a : a = 1, \ldots, N_f \} \) where \( I_a = (u_{a-1}, u_a) \) for \( j = 2, \ldots, N_f - 1, I_1 = (-\infty, u_1] \) and \( I_{N_f} = (u_{N_f - 1}, +\infty) \). That is:

\[
\begin{align*}
  f_X(x) = \left\{ 
  \begin{array}{ll}
    \omega_1 f_1(x) & x \leq u_1 \\
    \omega_2 f_2(x) & u_1 < x \leq u_2 \\
    \vdots \\
    \omega_a f_a(x) & u_{a-1} < x \leq u_a \\
    \vdots \\
    \omega_{N_f} f_{N_f}(x) & u_{N_f - 1} \leq x 
  \end{array}
\right.
\end{align*}
\]  

(1)

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

\[
\omega_a f_a(u_a) = \omega_{a+1} f_{a+1}(u_a), \quad \text{for} \quad a = 1, \ldots, N_f - 1. 
\]  

(2)

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

\[
\omega_1 F_1(u_1) + \ldots + \omega_a \left( F_a(u_a) - F_a(u_{a-1}) \right) + \ldots + \omega_{N_f} \left( 1 - F_{N_f}(u_{N_f-1}) \right) = 1 
\]  

(3)

where \( F_a \) denotes the corresponding probability distribution function.

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

\[
\omega_a = \frac{a_1}{b_1} \cdots \frac{a_{a-1}}{b_{a-1}} \left[ c_1 + c_2 \frac{b_1}{a_1} + c_3 \frac{b_2}{a_2} + \ldots + c_{a-1} \frac{b_{a-1}}{a_{a-1}} + \ldots + c_{N_f} \frac{b_{N_f-1}}{a_{N_f-1}} \right]^{-1}
\]  

(4)

where \( a_a = f_a(u_a), b_a = f_{a+1}(u_a) \) and \( c_a = F_a(u_a) - F_a(u_{a-1}) \), provided that \( a_a \) and \( b_a \) 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:

\[
a(t) = \sum_{n=1}^{N_F} a_n \phi_n(t) \quad t \in [0, N_y],
\]  

(5)

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 periodic or regular Sturm Liouville problem (SLP) with ordinary differential equation:

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

(6)

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

The orthogonality is interpreted in regards to the inner product \( < f(t), g(t) > = \int_a^b \omega(t) f(t) g(t) dt \). Table 1 presents some plausible sets for series expansion that can be used with the appropriate linear transformation of the domain into \([0, N_y]\). It also includes the nomenclature used for the presentation of results in section 3.
The negative log-likelihood function (NLLF) is used as the objective function in the optimization algorithm. It reads:

$$\text{NLLF}(\vec{\xi}) = - \sum_{j=1}^{N_o} \log f \left( x^o(t_j); \vec{\xi} \right),$$

(7)

where $\vec{\xi}$ 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 $j = 1, ..., N_o$ are the observations.

The optimization problem is defined as the search for values of $\vec{\xi}$ 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 marked variability of the range of values along the year. The second one presents a marked 22- and 11-year periodicity and rather clear shorter terms. Finally, a bivariate time series that includes a circular variable is analyzed.

3.1. Fresh-water river discharge at the Guadalquivir river estuary

This first application is devoted to a univariate time series, hereinafter denoted by $Q(t)$, which stands for the daily fresh-water river discharge from the Alcalá del Río dam (37.29° N, -6.06° W), the last regulation point of the Guadalquivir river (Andalucía, Spain) after its flow into the Atlantic Ocean from July 1st, 1931 to April 27th, 2016 (Source: Andalusian Water Agency, Junta de Andalucía). 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 a 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 salinity below a given threshold for the irrigation of rice crops in the estuary (Cobos, 2020). As a result, the series varies from very low values (usually in summer $Q < 40$ m$^3$/s) to those that are almost squared in winter ($Q \approx 1000$ m$^3$/s) with sporadic sudden changes. To deal with this high variability, a Box-Cox transformation with $\lambda = 0.1756$ parameter is used.

Several combinations of PMs such as Normal - Weibull of maxima, Log-normal - Normal, Normal - Generalized Pareto, with different initial guesses of the percentiles of the threshold, as well as single models like Weibull of maxima, Log-normal or Normal were used. The best visual fits were obtained for a Weibull of maxima distribution. When trying the fit with more than one distribution, for all those combinations where this distribution was one of the PMs, the final percentile of the common endpoint was negligible for the accompanying PM. This indicates that the methodology is capable to distinguish when a single PM works adequately for all the range of values and to skip needless PMs. For the

| Orthogonal set of eigenfunctions | Differential equation and domain | Conditions imposed |
|----------------------------------|----------------------------------|-------------------|
| Trigonometric series expansion   | $p(t) = q(t) = u(t) = 1, t \in [0, 1]$ | Periodic SLP: $\phi(0) = \phi(N_o)$ and $\phi'(0) = \phi'(1)$ |
| Modified Fourier series expansion | $p(t) = q(t) = 1, t \in [-1, 1]$ | Regular SLP: $\phi'(-1) = \phi'(1) = 0$ |
| Sinusoidal series expansion      | $p(t) = q(t) = 1, t \in [-1, 1]$ | Regular SLP: $\phi(-1) = \phi(1) = 0$ |
A method to characterize climate, Earth or environmental vector random processes

Figure 1: Non-stationary CDF of the freshwater river discharge for several choices of the time expansion of the parameters for a Weibull of maxima PM. a) Modified with 20 terms; b) Sinusoidal with 17 terms; c) Trigonometric with 21 terms; and d) Trigonometric with 25 terms.

Weibull of maxima single distribution, different sets of basic functions and number of terms retained in the expansion have been assessed.

Figure 1 compares the empirical distribution with some of the theoretical ones obtained with the expansions of the parameters given by the modified Fourier series with 10 terms ($N_F = 10$), the sinusoidal series with 16 terms and the trigonometric expansion with 10 and 12 terms, respectively from panels a to d. The body is so well fitted for all the options that the differences can only be spotted in figure 1 for the upper tail (percentiles > 75%) which is also rather well reproduced along the whole year. As it is observed, it is hard to visually select the best fit between these models, however, it may be accounted for that the expansion with a sinusoidal basis needs a considerably smaller number of parameters (51 vs 63 for a) and c) and 75 for d)).

The performance of the most representative options considered is analyzed in terms of the dimension of the optim...
A method to characterize climate, Earth or environmental vector random processes

![Graph](image)

**Figure 2:** 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 freshwater river discharge at the Guadalquivir river estuary. \( N_o = 30983 \) observations.

minimization problem \((N_d)\) and the BIC (Schwarz, 1978) (see Figure 2), which is related to the optimum value, \( \text{NLLF}^* \) and \( N_d \) through the mathematical expression \( \text{BIC} = 2\text{NLLF}^* + \log(N_o)N_d. \)

For the trigonometric and the set of sines series, the BIC starts diminishing until 8 terms. Then it increases and does not show any improvement by including one additional oscillation and, finally, it starts to decrease again at the expenses of a large dimension problem. The modified trigonometric functions, however, show a more consistent descending behavior. The last apparent improvement of the fits does not reveal itself in the visual comparison (see Figure 1.b) due to the Gibbs phenomenon associated to the limitation of working with discrete, and therefore, non-smooth data.

### 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 up to 2 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 22/6 years. However, as it is known that the semiannual component is significant, we force the analysis to optimize up to 0.5 (equal to 22/44 year period). 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 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.
A method to characterize climate, Earth or environmental vector random processes

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

3.3. Wind regimen at the Gulf of Cádiz

The third example analyzes the multivariate time series of the wind regimen (mean wind velocity, $W_v$, and mean incoming wind direction, $W_d$) hindcasted at 10 m above the mean sea level at the SIMAR point 1052048 located at 37° N, 7° W in the Gulf of Cádiz (Source: Puertos del Estado, Spain). The time series has $\approx$56 years duration, with data that spans from 1958/01/05 to 2011/12/15 with a 3-hours temporal cadence.

The univariate analysis of $W_v$ was carried out with a piecewise function that uses three PMs, a generalized Pareto PMs for the lower and upper tails and a Log-normal PM for the body. The initial guesses of the percentiles of the thresholds between PMs were 0.1 and 0.85. The wind direction, $W_d$, was fitted using a piecewise function with two Gaussian PMs with an initial guess of the partition at 0.5. For $W_v$ and $W_d$, the trigonometric and sinusoidal expansions were performed, respectively. In both cases, a basic period of one year ($N_y = 1$) with four terms for the wind magnitude and eight terms for wind direction were used. No Box-Cox transformation was required for the analysis.

Figure 4.a and .b shows the marginal fits of the two RPs. For the wind velocity, the percentiles shown are 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95 and 0.99, so that it can be appreciated the goodness of fit for the lower and upper tails. For the direction, different representative percentiles have been chosen for a better visualization of this circular variable. As it observed, the models adequately reproduce the non-stationary pattern.

With the results from the marginal and multivariate analysis (see Appendix), 100 simulations were obtained in order to verify the goodness of the method with the methodology proposed by Monbet et al. (2007).

Figure 5 shows the wind roses of the observations and one of the random simulations. The model adequately reproduces the minimum and maximum values as well as its magnitude at any given direction. The angular sector that comprises 225 - 270 degrees is slightly underestimated. Also, winds from the sector 90 - 135 degrees are more evenly distributed along the 1st and 2nd quadrants (0° - 180°, figure 5). This might be associated to the influence of the breezes from the Strait of Gibraltar that show a large interannual variability of easterly winds with oscillations ranging from 2 - 4 yr and also at decadal scales (Hidalgo and Gallego, 2019). These variabilities cannot be detected with $N_y = 1$, however, the use of a larger value of the basis period to capture them is unfeasible as only a few complete cycles would be available for the fit.

The joint distribution of the wind magnitude and direction is also assessed in figure 6 where it is represented the joint density functions of observations (panel a) and the simulation (panel b). The pdf of the simulation shows the
A method to characterize climate, Earth or environmental vector random processes

Figure 4: a) Non-Stationary Cumulative Distribution Function of wind magnitude, and b) and wind direction from SIMAR point 1052048 at the Gulf of Cádiz.

Figure 5: Wind roses for a) the observed data from SIMAR point 1052048 at the Gulf of Cádiz, and b) for a random simulation.

same bumps than the original data, centered at 90° (eastern) and 300° (western) for relatively strong winds close to 15 m/s, typical of that zone. It also captures those western winds are more frequent than eastern ones. The correlation coefficient obtained with the values of those functions is $R^2 = 0.981$, which shows that there is a good agreement between the simulation and the original time series bivariate distributions.

Finally, in figure 7 the estimations of the distributions of the sojourns duration above/below levels $1/2 \max(V_w)$ and $1/3 \max(V_w)$ (panels a and b, respectively) obtained for the original series and the simulations are compared. As pointed out by Monbet et al. (2007), these last plots give information about the persistence of stormy and calm conditions and are strongly related to the capability of the models to reproduce the severity of the climate conditions. The figures include the curves of the observations as well as an envelope band with the minimum and maximum values of the simulations. The autocorrelation functions show a similar behavior consisting in a decreasing trend with slightly daily oscillations. The differences are small for both variables, but larger for the wind direction. The wind direction temporal pattern, as usually, is quite difficult to reproduce (Monbet et al., 2007). This is particularly true for this area where wind veers in a rather sudden way. The temporal dependency of simulations is quite stable showing small envelope bands. In regard to sojourns durations, they are slightly underestimated/overestimated for the calm/stormy
Figure 6: Joint distribution of $W_v$ and $W_d$. a) Observed data, and b) one random simulation.

Figure 7: a) CDFs of sojourn durations over $1/2 \max(W_v) = 10.16$ m/s and, b) below $1/3 \max(W_v) = 6.78$ m/s.

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 pro-
gression, 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., 2021).

5. Conclusions

We have proposed a general procedure for the NS analysis of multivariate vector random processes. It uses NS
piecewise functions for the marginal distributions of the vector components which parameters 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 application of the method to three time series with different particularities shows it goodness to reproduce the
stochastic features of the original data for processes of different nature, being able to identify the appropriate values
of the thresholds and if 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 discharges at a dam located in a semi-arid zone in Andalucía (Spain) which
variation depends not only on seasonal and yearly time climate variability but also on management decisions. 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 Schawbe cycle and oscillations that vary close to 7.5 and 17 years. Finally, the
application of the method to a 10-meter height observed wind at the Andalusian Atlantic Ocean shows its capability to
reproduce different statistical properties inferred from the original series such as the marginal and joint distributions,
the wind rose, the duration of sojourns above given thresholds.

Acknowledgements

This work was performed within the framework of the following projects: (1) AQUACLEW, which is part of
ERA4CS, an ERA-NET initiative by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD
(DK), MINECO (ES), ANR (FR) with co-funding by the European Commission [Grant 690462] 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].

A. Appendices

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.
A1. 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 semi-arid 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 A-parameter Box-Cox transformation (Box and Cox, 1964). Other power transformations can also be applied (Yeo and Johnson, 2000).

A2. Temporal dependence

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

\[ Z_{X_i}(t_j) = \Phi^{-1} \left[ F_{X_i} (x^o(t_j); t_j) \right], \]

where \( \Phi^{-1} \) 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 \( y^i_j = Z_{X_i}(t_j) \) and \( Y_j = (y^1_j, ..., y^q_j, ..., y^N_j)^T \) where \( T \) stands for the vector transposition. The dependence in time between variables in the VAR(q) model is given by:

\[ Y_j = c + A^1Y_{j-1} + A^2Y_{j-2} + ... + A^qY_{j-q} + e_j, \]

where \( c = (c_1, ..., c^1, ..., c^N)^T \) contains the mean values of the variables, \( A^m, m = 1, ..., q \) are the \( N \times N \) coefficients matrices and \( e_j = (e^1_j, ..., e^j_q, ..., e^N_j)^T \) 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 \( j = q + 1, ..., N_o \), we obtain \( Y = AX + E \), where \( Y = (Y_{q+1}, Y_{q+2}, ..., Y_N)^T \), \( X = (X_{q+1}, X_{q+2}, ..., X_N)^T \), with \( X_j = (Y^T_{j-q} Y^T_{j-q})^T \), \( A = (A^1 A^2 \ldots A^q) \) and \( E = (e^1 e^2 \ldots e_N) \).

The solution is obtained by means of minimum least square errors as \( A = YX^T(XX^T)^{-1} \),491\( E = Y - AX \) and \( Q = \text{cov}(E) \) is the covariance matrix of the error. A detailed description can be found e.g. in Lütkepohl (2005).

A3. 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 A2. Then, the VAR model is used to generate a multivariate Gaussian stationary time series \( (y^1(t), ..., y^N(t)) \) at regular time instants. The corresponding non-stationary time series is then recovered by using the following transformation:

\[ X_i(t) = \Phi^{-1} \left[ F_{X_i} (y_i(t); t) \right]. \]

Conflict 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.

CRediT authorship contribution statement

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.
References

Baquerizo, A., Losada, M.A., 2008. Human interaction with large scale coastal morphological evolution. an assessment of the uncertainty. Coastal Engineering 55, 569–580.

Bocotti, P., 2000. Wave mechanics for ocean engineering. volume 64. Elsevier.

Box, G.E., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological) 26, 211–243.

Callaghan, D., Nielsen, P., Short, A., Ranasinghe, R., 2008. Statistical simulation of wave climate and extreme beach erosion. Coastal Engineering 55, 375–390. doi:https://doi.org/10.1016/j.coastaleng.2007.12.003.

Christakos, G., 2000. Modern spatiotemporal geostatistics. volume 6. Oxford university press.

Christakos, G., 2017. Spatiotemporal random fields: theory and applications. Elsevier.

Cobos, M., 2020. A model to study the consequences of human actions in the guadalquivir river estuary. Tesis Univ. Granada. URL: http://hdl.handle.net/10481/65374.

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). Journal of Geophysical Research: Oceans 125.

Cobos, M., Lira-Loarca, A., Christakos, G., Baquerizo, A., 2019. Storm characterization using a bme approach. Contributions to Statistics. Springer, Cham. doi:https://doi.org/10.1007/978-3-030-26036-1_19.

Cobos, M., Otiñar, P., Magaña, P., Lira-Loarca, A., Baquerizo, A., 2021. Marinetools.temporal: A python package to simulate earth and environmental time serie. under review. Environmental Modelling and Software X, XXX–XXX.

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, D.D., 2012. Multivariate return periods of sea storms for coastal erosion risk assessment. Natural Hazards and Earth System Sciences 12, 2699–2708.

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 Engineering 167, 103896. doi:https://doi.org/10.1016/j.coastaleng.2021.103896.

De Luca, D.L., 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 Engineering 54, 734–751.

Díez-Minguito, M., de Swart, H.E., 2020. Relationships between chlorophyll-a and suspended sediment concentration in a high-nutrient load estuary: An observational and idealized modeling approach. Journal of Geophysical Research: Oceans 125.

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. Journal of Environmental Management 112, 252–266. doi: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. Science of The Total Environment 794, 148670.

He, J., Kolovos, A., 2018. Bayesian maximum entropy approach and its applications: a review. Stochastic Environmental Research and Risk Assessment 32, 859–877.

Hidalgo, P., Gallego, D., 2019. A historical climatology of the easterly winds in the strait of gibraltar. Atmosfera 32, 181–195.

Hildebrandsson, H.H., 1897. Quelques recherches sur les centres d'action de l'atmosphère. Norstedt & Söner.

Izaguirre, C., Méndez, F.J., Menéndez, M., Luceño, A., Losada, I.J., 2010. Extreme wave climate variability in southern europe using satellite data. Journal of Geophysical Research: Oceans 115.

Jäger, W.S., Nagler, T., Czado, C., McCall, R.T., 2019. A statistical simulation method for joint time series of non-stationary hourly wave parameters. Coastal Engineering 146, 14–31.

Jalón, M.L., Baquerizo, A., Losada, M.A., 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 Engineering 62, 91–109.

Le Mouël, J.L., Lopes, J., Courtillot, V., 2019. A solar signature in many climate indices. Journal of Geophysical Research: Atmospheres 124, 2600–2619. doi:https://doi.org/10.1002/2018JD028939, arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018JD028939.

Lira-Loarca, A., Cobos, M., Losada, M.A., Baquerizo, A., 2020. Storm characterization and simulation for damage evolution models of maritime structures. Coastal Engineering 156, 103620.

Loarca, A.L., Cobos, M., Besio, G., Baquerizo, A., 2021. Projected wave climate temporal variability due to climate change. Stochastic Environmental Research and Risk Assessment, 1–17.

López-Ruiz, A., Bergillos, R.J., 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, F.J., 2006. The effect of temporal dependence on the estimation of the frequency of extreme ocean climate events. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 462, 1683–1697.

Lütkepohl, H., 2005. New introduction to multiple time series analysis. Springer Science & Business Media.

Méndez, F.J., Menéndez, M., Luceño, A., Losada, I.J., 2006. Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. Journal of Geophysical Research: Oceans 111, 561.

Méndez, F.J., Menéndez, M., Luceño, A., Medina, R., Graham, N.E., 2008. Seasonality and duration in extreme value distributions of significant wave height. Ocean Engineering 35, 131–138.

Monbet, V., Ailliot, P., Prevosto, M., 2007. Survey of stochastic models for wind and sea state time series. Probabilistic engineering mechanics 22, 113–126.

M Cobos et al.: Preprint submitted to Stochastic Environmental Research and Risk Assessment
A method to characterize climate, Earth or environmental vector random processes

Pawłowicz, R., Beardsley, B., Lentz, S., 2002. Classical tidal harmonic analysis including error estimates in matlab using t_tide. Computers & Geosciences 28, 929–937.

Payo, A., Baquerizo, A., Losada, M., 2004. Uncertainty assessment of long term shoreline prediction. 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. Journal of hydraulic research 46, 96–104.

Ranasinghe, R., D., C., M.J.F., S., 2012. Estimating coastal recession due to sea level rise: beyond the bruun rule. Climatic Change, 561–574 doi: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 resources management 31, 2065–2083.

Schwarz, G., 1978. Estimating the dimension of a model. The annals of statistics 6, 461–464.

Solari, S., Losada, M., 2011. Non-stationary wave height climate modeling and simulation. Journal of Geophysical Research: Oceans 116.

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 Technology and Engineering 1, 217–230.

Usoskin, I., Mursula, K., 2003. Long-term solar cycle evolution: review of recent developments. Solar Physics 218, 319–343.

Von Stryk, O., 1993. Numerical solution of optimal control problems by direct collocation, in: Optimal control. Springer, pp. 129–143.

Walker, G., 1924. Correlations in seasonal variations of weather, ix, a further study of world weather (world weather ii). Memoirs of India Meteorological Department 24, 275–332.

Walker, G.T., Bliss, E.W., 1932. World weather v. Mem. Roy. Meteor. 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.

Yeo, I.K., Johnson, R.A., 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. doi:https://doi.org/10.1016/j.newast.2016.09.004.