Correspondence to:
A. Rueda,
ruedaac@unican.es

Multiscale climate emulator of multimodal wave spectra: MUSCLE-spectra

Ana Rueda[1] , Christie A. Hegermiller[2] , Jose A. A. Antolinez[1], Paula Camus[3], Sean Vitousek[2] , Peter Ruggiero[4] , Patrick L. Barnard[2] , Li H. Erikson[2] , Antonio Tomás[3] , and Fernando J. Mendez[1]

[1]Departamento de Ciencias y Tecnicas del Agua y del Medio Ambiente, Universidad de Cantabria, Santander, Spain, [2]Pacific Coastal and Marine Science Center, United States Geological Survey, Santa Cruz, California, USA, [3]Environmental Hydraulics Institute, IHCantabria, Universidad de Cantabria, Santander, Spain, [4]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA

Abstract Characterization of multimodal directional wave spectra is important for many offshore and coastal applications, such as marine forecasting, coastal hazard assessment, and design of offshore wave energy farms and coastal structures. However, the multivariate and multiscale nature of wave climate variability makes this complex problem tractable using computationally expensive numerical models. So far, the skill of statistical-downscaling model-based parametric (unimodal) wave conditions is limited in large ocean basins such as the Pacific. The recent availability of long-term directional spectral data from buoys and wave hindcast models allows for development of stochastic models that include multimodal sea-state parameters. This work introduces a statistical downscaling framework based on weather types to predict multimodal wave spectra (e.g., significant wave height, mean wave period, and mean wave direction from different storm systems, including sea and swells) from large-scale atmospheric pressure fields. For each weather type, variables of interest are modeled using the categorical distribution for the sea-state type, the Generalized Extreme Value (GEV) distribution for wave height and wave period, a multivariate Gaussian copula for the interdependence between variables, and a Markov chain model for the chronology of daily weather types. We apply the model to the southern California coast, where local seas and swells from both the Northern and Southern Hemispheres contribute to the multimodal wave spectrum. This work allows attribution of particular extreme multimodal wave events to specific atmospheric conditions, expanding knowledge of time-dependent, climate-driven offshore and coastal sea-state conditions that have a significant influence on local nearshore processes, coastal morphology, and flood hazards.

1. Introduction

Directional wave spectra are usually summarized in terms of simple aggregated parameters, such as significant wave height (H_s), peak wave period (T_p), and mean wave direction (D). However, this simplification lacks a description of multimodal sea states with short-period seas and long-period swell originating from multiple storm systems [*Boukhanovsky and Guedes Soares*, 2009], all of which are needed to accurately model local nearshore processes [*García-Medina et al.*, 2013], coastal morphology, and flood hazards. Accurate modeling of multimodal sea states is also relevant for analyzing wave energy resources, computing wave loads on offshore structures, and estimating the probability of rogue waves [*Trulsen et al.*, 2015]. Mixed sea states (with both sea and swell present) are very common on coastlines that are distant from swell-generation regions [*Semedo et al.*, 2011]. For example, *Crosby et al.* [2016] pointed out that 50% of the sea states in southern California are bimodal.

Long-term (30+ years) buoy records and wave hindcast models [*Rascle and Ardhuin*, 2013] have greatly improved our knowledge of multimodal directional spectra and thus our ability to generate synthetic, probabilistic forecasts. Viable statistical forecasting methods must treat (1) the multivariate nature of the sea state, (2) the chronology of the sea states at different temporal scales (daily, monthly, interannual, interdecadal, and climate-change time scales), and (3) extreme sea states.

Recent work addressing statistical wave forecasting has focused on parameterizing partitions of long-term directional spectra [*Boukhanovsky and Guedes Soares*, 2009], reproducing the short-term chronology of directional spectra types based on transition probabilities [*Boukhanovsky et al.*, 2007; *Lucas et al.*, 2011], and analyzing variability of wave spectra based on weather types [*Espejo et al.*, 2014].

Extreme sea states vary seasonally, interannually, and on longer time scales potentially related to climate change [*Milly et al.*, 2008]. Several extreme value models have been developed recently to deal with nonstationarity [*Katz et al.*, 2002; *Mendez et al.*, 2006; *Serafin and Ruggiero*, 2014]. Long-term projections of sea states are required for many offshore and coastal applications [*Solari and Losada*, 2011] or probabilistic estimation of wave-induced coastal erosion [*Walstra et al.*, 2013; *Callaghan et al.*, 2008, 2013; *Corbella and Stretch*, 2012; *Antolinez et al.*, 2016].

Classification of atmospheric conditions into a number of representative weather types (WT) is a technique widely used by climatologists [*Huth*, 2001], providing a framework to assess seasonal-to-interannual variability [*Guanche et al.*, 2013]. Statistical downscaling methods based on WTs are flexible frameworks that can explore climate variability of many physical processes at different time scales. These methods have been recently applied in *Camus et al.* [2014b] and *Bárdossy et al.* [2015] for the empirical distribution of sea-state parameters, in *Espejo et al.* [2014] for directional wave spectra, in *Rueda et al.* [2016a] for an extreme value model of daily significant wave height maxima, and in *Guanche et al.* [2013] for modeling the chronology of weather types.

The integration of (i) weather-types classification, (ii) chronology of weather types, (iii) multimodal directional spectra, and (iv) the joint distribution of multivariate extremes requires a novel framework composed of many recently developed methodologies. In this paper, we propose a weather-type-based framework [*Camus et al.*, 2014b; *Rueda et al.*, 2016b] to model daily multivariate events of sea and swell systems, each represented by H_s, T_p, and D, using Generalized Extreme Value (GEV) marginal distributions for H_s and T_p and the empirical distribution for D. Multivariate Gaussian copulas [*Nelsen*, 2006; *Salvadori et al.*, 2007] are used to capture the correlation between variables. The statistical dependence between consecutive days is addressed by defining a climate-based extremal index for each weather type [*Rueda et al.*, 2016a]. The chronology of multivariate sea states is modeled by a time-dependent Markov chain model for daily weather types.

The paper is organized as follows: section 2 describes the overall framework and methodology, section 3 details the preprocessing of the spectral data of the model applied to the case study, section 4 describes the statistical model for the predictor, section 5 describes the statistical model for the predictand, and section 6 demonstrates the results of the stochastic simulation. Application of the framework to the southern California coast is presented throughout sections 3–6 to facilitate comprehension of the approach. Finally, section 7 summarizes and concludes the work.

2. Overview of the Methodology

The goal of this work is to develop a time-dependent emulator based on a nonlinear statistical downscaling method that relates a local multivariate predictand (daily multimodal directional wave spectra) with a multivariate predictor (daily representative patterns of sea level pressure). The statistical downscaling method employed is based on clustering the atmospheric forcing into a number of representative daily patterns, namely weather types (WT) and analyzing their associated daily wave conditions. Cluster analysis (i.e., WTs) yields better fits of the statistical distributions of the predictand and increases the correlation between interdependent variables (e.g., H_s and T_p). This framework is conceived to model the intramonthly, seasonal, and interannual variability of wave conditions considering the chronology of the daily weather types based on a time-dependent categorical distribution [*Guanche et al.*, 2013]. The methodology (Figure 1) is based on a number of steps grouped into four main modules: (A) parameterization of spectral data, (B) statistical model for the predictor, (C) statistical model for the predictand, and (D) climate-based stochastic simulation of synthetic time series of the multivariate predictand. In relation to previous works [*Perez et al.*, 2014; *Camus et al.*, 2014a,2014b; *Rueda et al.*, 2016a,2016b], the primary new contributions of this work are highlighted in Figure 1 in grey boxes. The substeps of the methodology are listed below:

A1. Define areas of wave generation using ESTELA [*Perez et al.*, 2014].
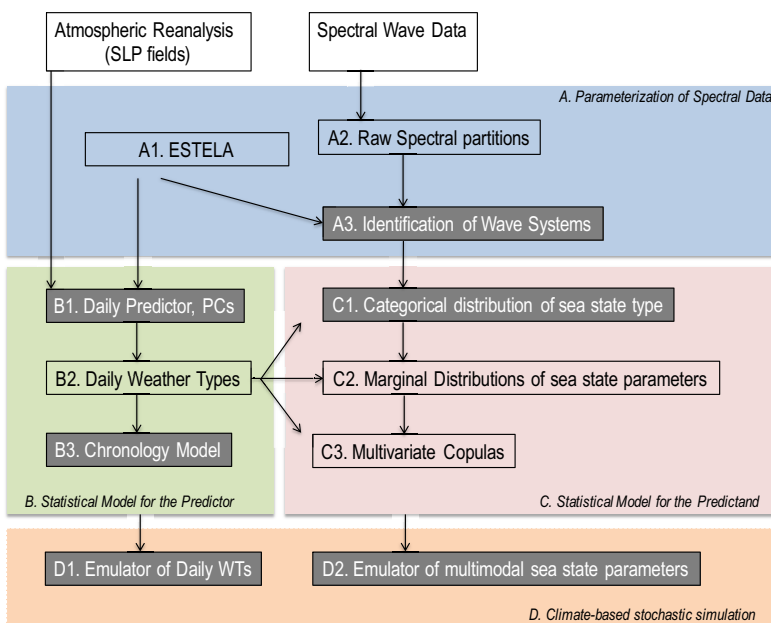
Figure 1. Flowchart of the methodology to obtain a multivariate climate-based emulator for multipeaked spectra.

A2. Obtain sea and swell partitions of the directional wave spectra.

A3. Identify the different areas of sea and swell generation based on ESTELA and the partitions.

B1. Construct the daily predictor, i.e., principal components of the daily sea level pressure (SLP) fields in the areas of sea and swell generation [*Perez et al.*, 2014; *Camus et al.*, 2014a] taking into account the area between daily isochrones, meaning areas between the same daily wave travel time to the target area [*Hegermiller et al.*, 2017].

B2. Define daily weather types from synoptic SLP patterns [*Camus et al.*, 2014b].

B3. Develop a chronology model for the sequencing of daily weather types using a time-dependent Markov chain.

C1. Obtain the categorical distribution of sea-state types (e.g., unimodal, bimodal, or multimodal wave conditions) for each weather type.

C2. Fit the marginal distributions of the daily sea-state parameters (Hs, Tp, and D) of the sea and swell systems for each weather type [*Rueda et al.*, 2016a].

C3. Model the dependence between predictand variables for each weather type using a multivariate Gaussian copula function [*Rueda et al.*, 2016b].

D1. Generate synthetic, long-term time series (possible realizations of climate) of daily weather types considering daily, monthly, and annual variability.

D2. Generate synthetic multivariate sea-state parameters (Hs, Tp, and D) considering the occurrence probability of the sea-state types and its dependence structure associated with each weather type.

3. Parameterization of Spectral Data

3.1. Finding the Large-Scale Swell-Generation Regions

We first identify the source regions and travel time of wave energy reaching the target area using ESTELA (Evaluating the Source and Travel-time of the wave Energy reaching a Local Area) method [*Perez et al.*, 2014]. ESTELA maps wave-generation areas that contribute to the wave conditions at a point of interest, based on a geographical criteria and information about two-dimensional wave spectra. In Figure 2a, applying ESTELA to a point located offshore of southern California reveals two important regions: a region covering the North Pacific from 30°N to 60°N driving swells in the sector 240°–360°, and a region in the southwest Pacific near Australia and New Zealand driving swells in the sector 150°–240°. This rough segmentation into two large-scale swell-generation regions will be used to identify the different wave systems at the target location.
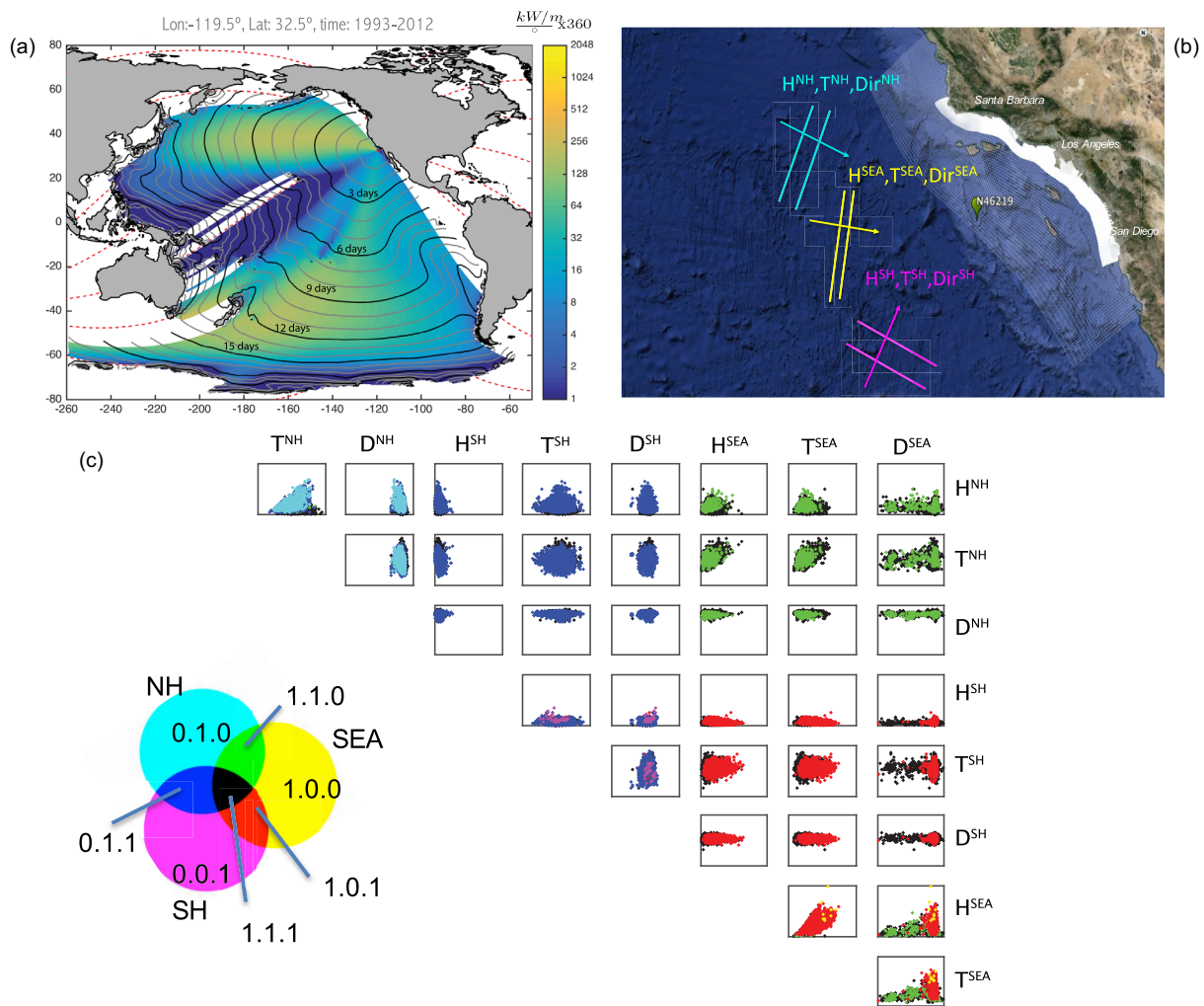
Figure 2. (a) ESTELA map [*Perez et al.*, 2014] applied to a target point in southern California; (b) sketch of three wave systems approaching to southern California; (c) scatterplots of the nine variables (*Hs*, *Tp*, and *D* for SEA, NH, and SH components). Color represents unimodal sea states (yellow, cyan, and magenta), bimodal sea states (red, green, and blue) and trimodal sea states (black). Scales: *H* [0, 11 m], *T* [0, 25 s], *D*[0, 360°].

3.2. Partitioning the Directional Wave Spectra

We assume that the directional wave spectra, $S(f, \theta)$, is composed of N + 1 wave systems, in the form:

$$S(f, \theta) = \sum_{i=0}^{N} S_i(f, \theta)$$

where the index value $i = 0$ is associated with local wind waves (SEA), N is the number of swell systems, f is the frequency, and θ is the direction. Recall that the total wave energy is given by $\int_{-\pi/2}^{\pi/2} \int_{0}^{\infty} S(f, \theta) \, df \, d\theta = m_0$, and, for example, the significant wave height is given by $H_s = 4\sqrt{m_0}$ [*U.S. Army Corps of Engineers*, 2002]. However, this aggregated representation is inaccurate for multimodal spectra, motivating our intention to partition the spectrum.

The ideal number of swell systems is estimated using partition algorithms [e.g., *Vincent and Soille*, 1991; *Hanson and Phillips*, 2001] or numerical optimization [*Boukhanovsky and Guedes Soares*, 2009]. Here we select two swell partitions and one sea partition from an hourly wave hindcast from 1979 to 2013 [*Perez et al.*, 2015] at location 33°N, 120°W offshore of southern California, generated by a global WaveWatch III model (version 4.18) [*Tolman*, 2014] with 0.5° horizontal resolution. The model setup is based on a multigrid

system and was forced with hourly wind and ice coverage from the Climate Forecast System Reanalysis (CFSR; 30 km spatial resolution approximately). Validation against buoys and altimeter data (not shown) demonstrated good agreement between this hindcast and observations [*Perez et al.*, 2015].

3.3. Identifying the Dominant Wave Systems

After identifying the regions of swell generation with ESTELA and partitioning the directional wave spectrum at our target location offshore of southern California, we relate the physical process of wave generation to the resulting wave systems [*Hanson and Phillips*, 2001]. Here we have applied a spatial criterion, splitting the directional wave spectrum into two large-scale generation regions, namely, the Northern Hemisphere ocean "paddle" and the Southern Hemisphere ocean "paddle." Therefore, we focus on one wind sea (SEA) and two swell partitions generated in the Northern (NH) and Southern (SH) Hemispheres,

$$S(f, \theta) = S_{SEA}(f, \theta) + S_{NH}(f, \theta) + S_{SH}(f, \theta)$$

Based on ESTELA model results (section 3.2), $S_{NH}(f, \theta)$ and $S_{SH}(f, \theta)$ are calculated by aggregating all of the wave systems within the sectors $240° \leq D \leq 360°$ and $140° \leq D \leq 240°$, respectively, where D is the mean direction of each wave system (see sketch of Figure 2b). For southern California, partitioning the wave spectrum according to wave direction alone is sufficient. However, other locations of interest might require a more sophisticated partitioning of the wave spectrum, for example, using both wave direction and frequency.

The directional spectrum is partitioned into three components where i = 0 represents local wind seas (denoted by SEA), i = 1 represents Northern Hemisphere swells (denoted by NH), and i = 2 represents Southern Hemisphere swells (denoted by SH). A spectrum partition is parameterized as $Si(f, \theta) = S(f; H_s^i T_p^i, \sigma_f^i)D(f, \theta; D^i, \sigma_\theta^i)$, where $S(f)$ is the frequency spectrum with significant wave height, H_s^i, peak period, T_p^i, and shape coefficient, σ_f^i (e.g., JONSWAP) and $D(f, \theta)$ is the directional spreading distribution where D^i is the mean direction and σ_θ^i is the directional spreading coefficient (cosine type distribution). As an example, the wave system from the Northern Hemisphere is defined as $(H_s^{NH}, T_p^{NH}, D^{NH}, \sigma_f^{NH}, \sigma_\theta^{NH})$. Without loss of generality, we will restrict our analysis to three variables (H_s, T_p, and D) for each wave system (Figure 2b). Therefore, our application to southern California considering one wind sea and two swells becomes a 9-D multivariate problem $(H_s^{SEA}, T_p^{SEA}, D^{SEA}, H_s^{NH}, T_p^{NH}, D^{NH}, H_s^{SH}, T_p^{SH}, D^{SH})$. In our application, the 9-D predictand is selected at a daily scale from the hourly records that produce the daily maximum of wave runup, calculated as $Z = 0.043Hs^{0.5} \cdot L_0$ using the empirical formulation of *Stockdon et al.* [2006] for wave runup on dissipative beaches, where H_s and L_0 are the bulk wave parameters (e.g., $H_s = \sqrt{\sum_{i=0}^{N} H_{s_i}^2}$). Figure 2c shows the scatterplot of the 9-D space of the WaveWatch III hindcast data at the southern California location. Colored dots (see Figure 2, left) represent unimodal sea states (yellow for SEA, cyan for NH, and magenta the SH), bimodal sea states (green for NH + SEA, blue for NH + SH, and red for SEA + SH), and trimodal sea states (black for SEA + NH + SH).

4. Statistical Model for the Predictor

4.1. Building the Predictor

SLP fields from the global Climate Forecast System Reanalysis h(CFSR) [*Saha*, 2010] define the predictor of the statistical downscaling model. The CFSR hindcast spans from 1979 to 2009 with hourly temporal resolution and 0.5° spatial resolution.

Based on previous work [*Camus et al.*, 2014a; *Perez et al.*, 2015; *Rueda et al.*, 2016a], the daily predictor is defined as the SLP and squared SLP gradient (SLPG) fields, representing the geostrophic wind conditions over a spatial domain that covers the corresponding wave-generation area (given by ESTELA). However, other predictors such as geopotential height [*Bárdossy et al.*, 2015] or SLP and winds [*Martínez-Asensio et al.*, 2016] could be used. In contrast to *Camus et al.* [2014a], instead of using the averaged n days SLP and SLPG fields, we build a predictor P_t considering isochrones which are defined as the averaged temporal lags of the energy arriving to the target area:

$$P_t = \{\ldots, SLP_{t-i+1,\Omega i}, SLPG_{t-i+1,\Omega i} \ldots\} \ \text{for} \ i=1,\ldots,p$$

where Ω_i represents the spatial domain between the daily isochrones $i-1$ and i, and p is the number of days of the last isochrone predicted by ESTELA (21 days in the current application), which represents the longest possible wave propagation time from generation until arrival at the target location. We apply principal component (PC) analysis to the spatial-temporal field P_t to reduce the high dimensionality of the predictor while retaining 95% of the variance.

Following *Cannon* [2012] and *Camus et al.* [2016], a regression-guided (RG) *k*-means classification with $n_{DWT} = 36$ daily weather types (DWT) is applied to the daily PCs of the predictor and the daily bulk wave parameters (H_s, T_p, and D) used to select the multivariate predictand (9-D sea state). The RG classification varies according to a parameter α that controls the weight of the predictor (by a factor $1-\alpha$) and the predictand, explained by a multivariate linear regression model, (by a factor α) on the classification. In large ocean basins, such as the Pacific, where the wave climate at a particular point is the combination of distant swells and local seas, the application of a regression-guided classification significantly improves the statistical downscaling performance. It helps to detect weather types with a stronger influence on the local wave conditions. In the current application, we obtain an optimal factor of $\alpha = 0.2$, meaning that a regression of the predictand is being introduced on the classification with 20% of the weight. This step increases the homogeneity of the multivariate predictand within each WT and therefore, improves the statistical downscaling performance. Figure 3 shows the DWT classification, as well as the seasonal occurrence probability. The WTs shown on Figure 3 are organized in a bidimensional lattice, where similar patterns are located together. Each daily weather type (DWT_i) represents the mean synoptic circulation conditions in each cluster of the regression-guided classification. The probability of each cluster (p_i) is calculated from the number of SLP fields belonging to each cluster in a particular season. A climate-based extremal index for each weather type [*Rueda et al.*, 2016a] accounts for the statistical dependence between consecutive days.

As seen in Figure 3, boreal winter (DJF) and summer (JJA) patterns appear on the left and right sides of the lattice, respectively. For example, note that DWT2 and DWT6 depict the largest low-pressure systems, driving strong, swell-generating westerlies directed at the target area.

4.2. Chronology Model

The chronology of DWTs drives the time dependence of the current model. To simulate the time dependence of the natural system, the chronology model must capture (1) the historical probability of occurrence of the different daily weather types, (2) the transition probabilities, (3) the persistence, and (4) the seasonal-to-interannual variability. To address items (1), (2), and (3), a stationary Markov model is sufficient, however, item (4) requires nonstationary approaches such as logit regression models for seasonal [*Jordan and Talkner*, 2000] and interannual variability [*Guanche et al.*, 2013]. In the current model, interannual variability is introduced by linear covariates representing well-known climate indices (e.g., PNA and SOI) or tailor-made monthly indices [*Camus et al.*, 2014a]. Here the monthly covariates are represented by a collection of monthly weather types MWT_i, $\{i=1,\ldots,n_{MWT}\}$ based on the PCs of SLP and SLPG at a monthly scale following *Camus et al.* [2014b]. We applied the classification algorithm to different numbers of monthly weather types, n_{MWT}, ranging from 12 to 36, analyzed the monthly and daily persistence and variability (see section 6.1), and finally selected $n_{MWT} = 16$ as the optimal value. Figure 4a shows the monthly patterns and seasonal occurrence probabilities (bottom plot). In this case, winter (DJF) conditions appear in the top plots of the lattice whereas summer (JJA) conditions appear in the bottom plots. Additionally, we obtain a probabilistic relationship between the daily and monthly predictors by obtaining the occurrence probabilities of the DWTs projected into the MWT lattice (see Figure 4b). Finally, stationary Markov models for each MWT are obtained. The chronology model is composed of two contemporaneous time processes: (1) the monthly scale, $X_\tau^m \in \{MWT_1,\ldots,MWT_{n_{MWT}}\}$, where τ represents each month, and (2) the daily scale, $X_t^d \in \{DWT_1,\ldots,DWT_{n_{DWT}}\}$, where t represents each day. In short, the transition probability matrix p_{ijk} models the transition from DWT_i to DWT_j during MTW_k (Figure 4c).

$$p_{ijk} = \text{Pr}\,(X_{t+1}^d = DWT_j | X_t^d = DWT_i, X_\tau^m = MWT_k)$$

For clarification, the transition probability matrix, p_{ijk}, for MWT9 (k = 9) is shown in Figure 4d. Note that, because MWT9 corresponds with a winter pattern, only certain DWTs can occur. For example, the most common daily weather types associated with MWT9 are DWT3 and DWT10 (as shown in Figure 4b). The
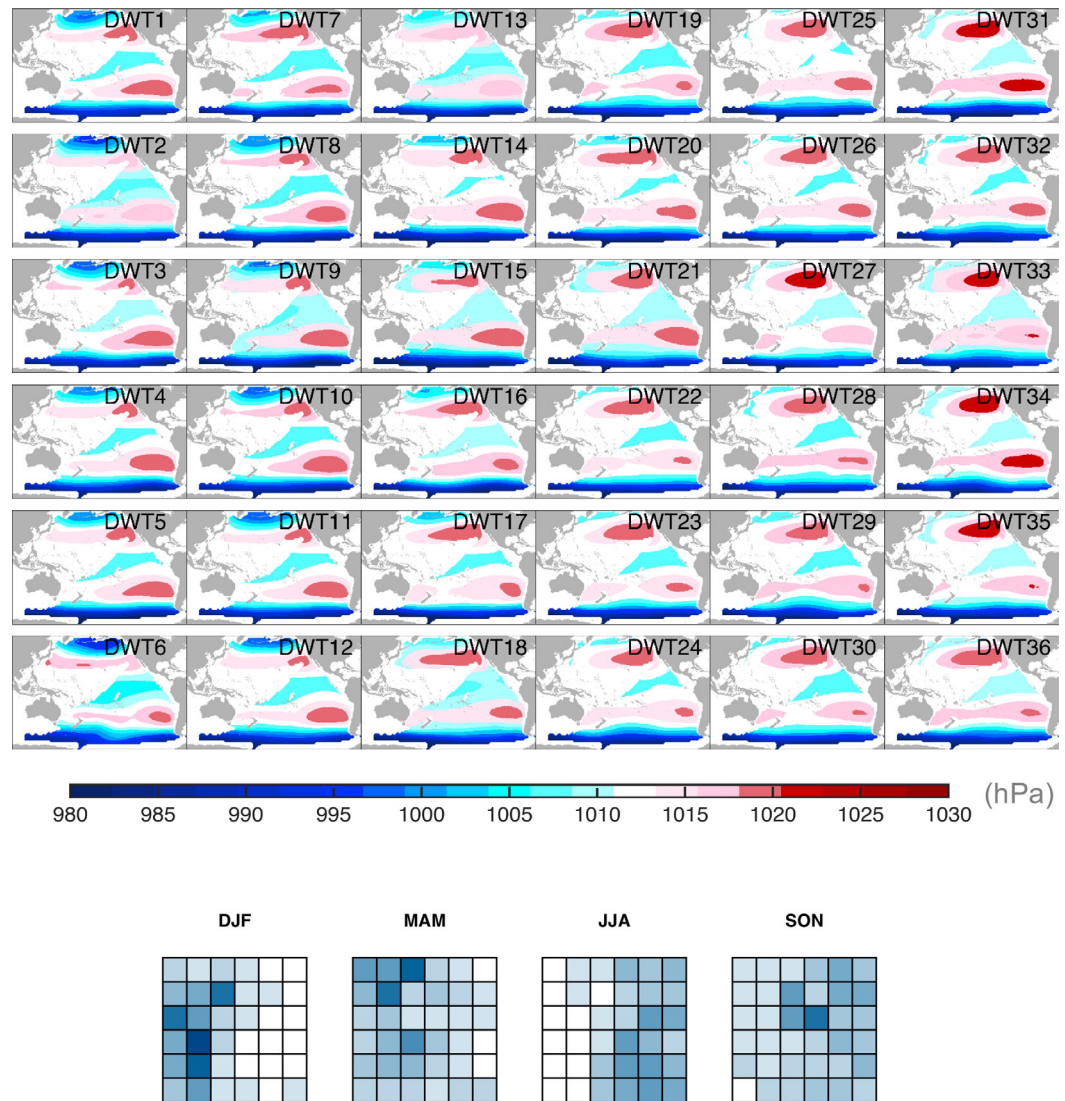
Figure 3. The regression-guided weather-types (WT) classification represented by SLP fields (hPa) corresponding to the predictor-to-predictand classification obtained for a factor $\alpha = 0.2$. The seasonal occurrence probability (p_i, in blue scale) is shown in the bottom plot.

diagonal dominance of the transition probability matrices, shown in Figures 4c and 4d, indicates that the most common behavior is to remain in the same DWT for a few days.

In this work, the sequencing of MWTs is based on the historical realization of the climate during the period 1979–2009. The stochastic modeling of the chronology of MWTs, although possible, is beyond the scope of the current work.

5. Statistical Model for the Predictand

The statistical downscaling model expresses the probabilistic relationship between the daily weather types and the daily multimodal spectra. In section 3, we partitioned the multimodal directional spectra into one wind sea and two swell components for the target area. The different partitions (SEA, NH, and SH) can be considered present (1) or not (0), producing $2^3 = 8$ possible combinations. Consequently, a sea state is considered as a mixed random variable: (i) the sea-state type (unimodal or multimodal) is modeled as a discrete random variable, and (ii) the magnitude of wave height, wave period, and wave direction for each mode is modeled as a continuous multivariate random variable (i.e., a scalar variable for H_s and T_p, and a circular variable for D).
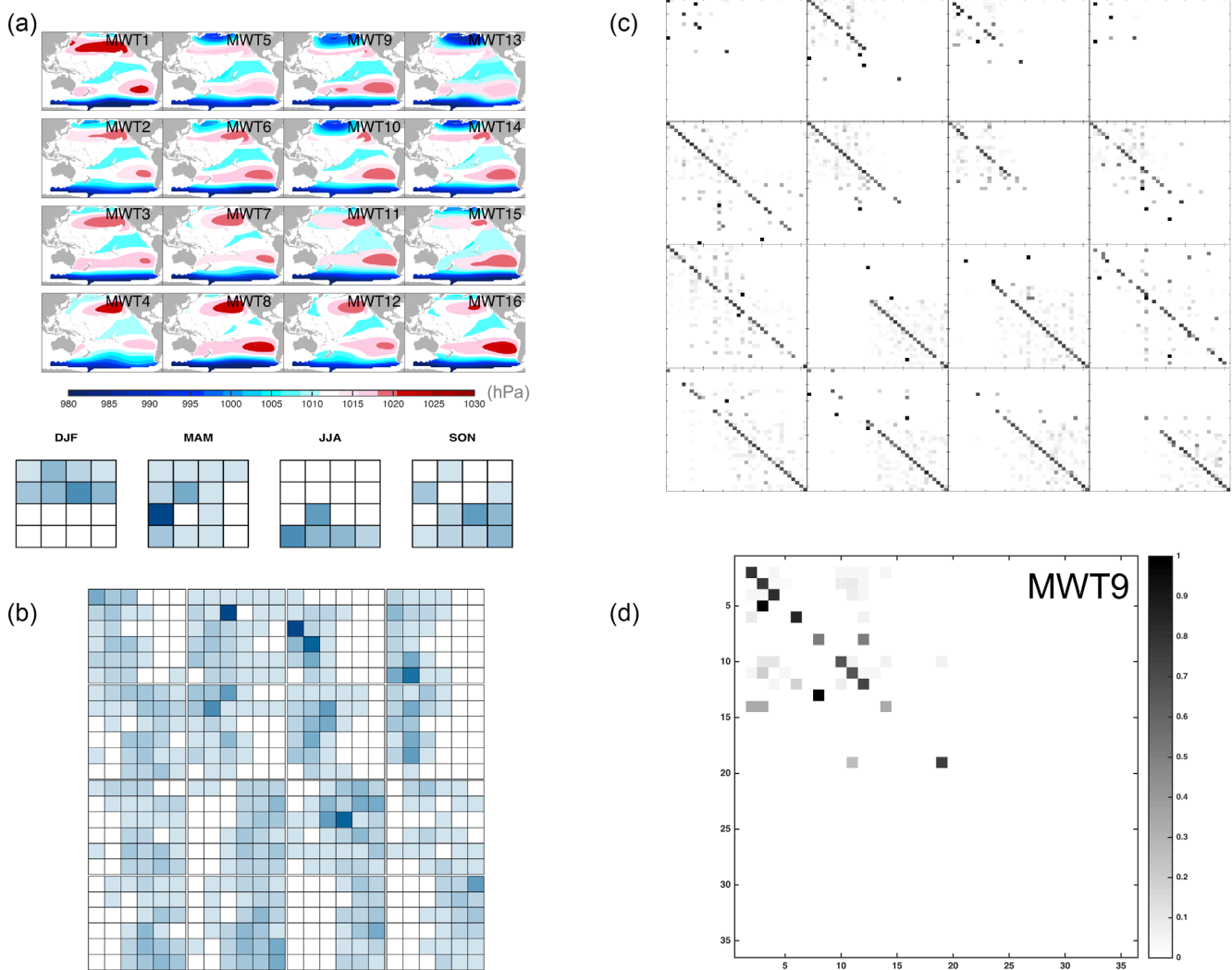
Figure 4. (a) Monthly weather types, MWT, and seasonal occurrence probability; (b) occurrence probability of DWTs projected into the MWT lattice; (c) transition probability matrices of DWTs projected into the MWT lattice; (d) zoom of the transition probability matrix for MWT9.

5.1. Categorical Distribution of Sea-State Type

The sea-state type is represented as a random variable $\mathbf{I}$, which is a sequence of three Bernouilli trials,

$$
\left.\begin{array}{l}
I_{SEA} = Bernoulli(p_{SEA}) \\
I_{SEA} = Bernoulli(p_{NH}) \\
I_{SEA} = Bernoulli(p_{SH})
\end{array}\right\} = \mathbf{I} = \{I_{SEA}, I_{NH}, I_{SH}\}
$$

where the sample space is composed of eight triples, $S = \{000, 100, 010, 001, 110, 101, 011, 111\}$, although the event 000 (no waves) is highly uncommon, if not impossible. Figure 2c depicts the sample space S following a CMYK color code, S = {white, yellow, cyan, magenta, green, red, blue, black}. Removing event 000, the random variable $\mathbf{I}$ follows a categorical distribution with seven possible outcomes, $\mathbf{I} \sim (\{p_{100}, p_{010}, p_{001}, p_{110}, p_{101}, p_{011}, p_{111}\} | DWT_i), \{i = 1, \ldots, n_{DWT}\}$, and $\sum p_i = 1$. In our statistical downscaling framework, $\mathbf{I}$ is conditioned to each daily weather type. Figure 5 shows the occurrence probabilities of the categorical distribution for each DWT. Unimodal sea states are represented by dots in the top corner (p_{010}, NH), bottom corner (p_{001}, SH), and right corner (p_{100}, SEA). Bimodal sea states are represented by dots in the top edge (p_{110}, NH + SEA), bottom edge (p_{101}, SH + SEA) and left edge (p_{011}, SH + NH). Trimodal sea states (p_{111}, SEA + NH + SH) are represented by the dot in the center of each triangle. Note that the contribution of each particular sea-state type strongly depends on the DWTs, reinforcing the validity of the
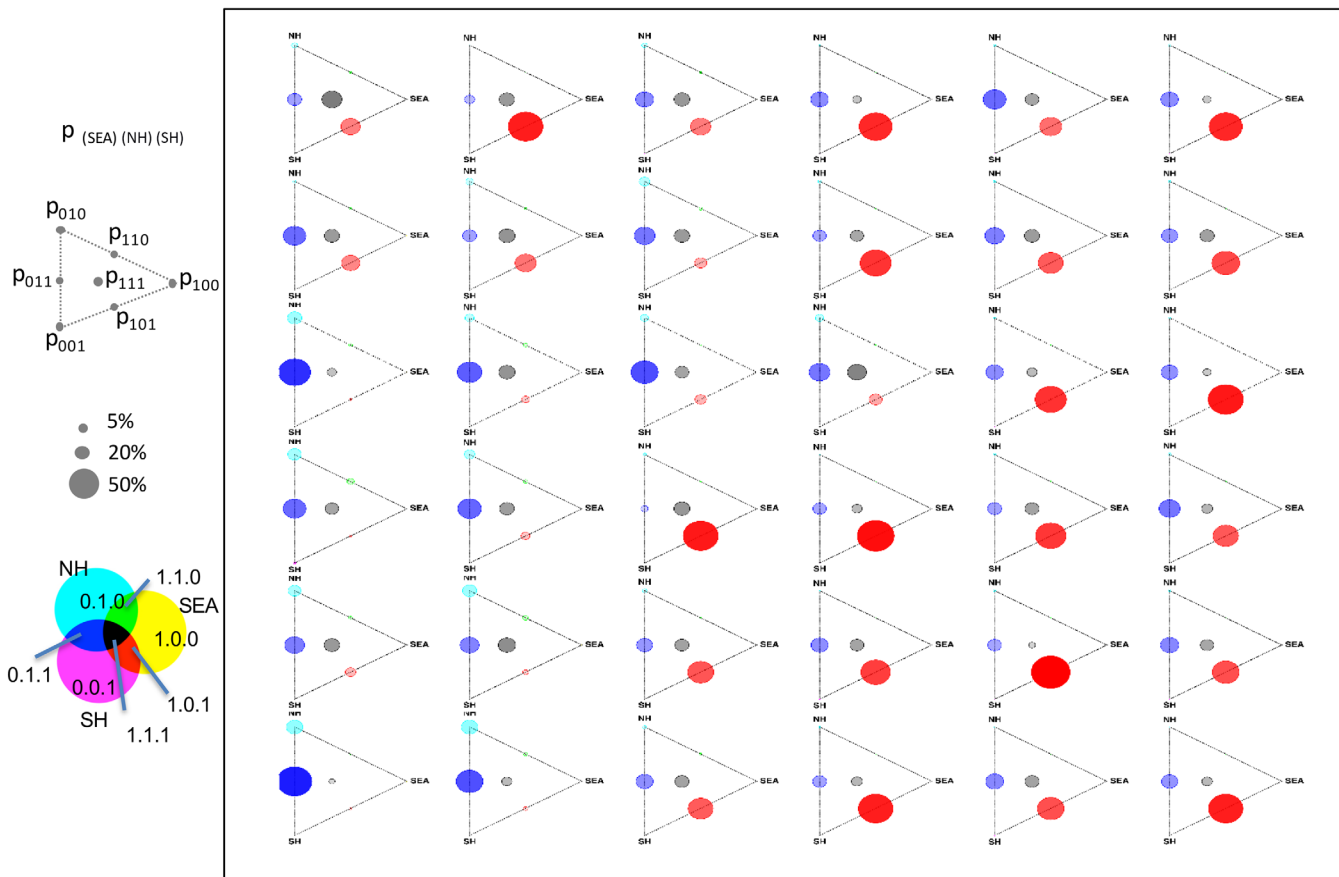
Figure 5. Occurrence probability of sea-state types (combination of wave systems) conditioned to each DWT. Unimodal sea states are represented by dots in the top corner (NH), bottom corner (SH), and right corner (SEA). Bimodal sea states are represented by dots in the top edge (NH + SEA), bottom edge (SH + SEA), and left edge (SH + NH). Trimodal sea states (SEA + NH + SH) are represented by the dot in the center of each triangle.

weather type framework to physically interpret and associate a large-scale synoptic weather pattern with a local wave predictand. As an example, the most common sea-state types are the bimodal NH + SH in the boreal winter (in 10 DWTs) and the bimodal SH + SEA in the austral winter (in 20 DWTs).

5.2. Marginal Distribution of Sea-State Parameters

The multivariate random variable is defined by the joint distribution of significant wave height, peak wave period, and mean wave direction for the unimodal {100,010,001}, bimodal {110,101,011}, or trimodal {111} sea-state types. The strategy for modeling the joint distribution is: (a) obtain marginal distributions (Generalized Extreme Value distributions for H_s and T_p [*Rueda et al.*, 2016a,2016b] and an empirical distribution for D, although the methodology is not limited to these distributions); (b) transform each marginal into a Normal distribution N(0,1) to account for the dependence between the variables; and (c) apply a multivariate copula. For example, sea-state type $\mathbf{I} = \{0,1,0\}$, a swell from NH, is a multivariate random variable with three components $\left\{H_s^{NH}, T_p^{NH}, D^{NH}\right\}$. In this particular case, H_s^{NH} follows $GEV\left(\mu_H^{NH}, \psi_H^{NH}, \xi_H^{NH}|DWT_i\right)$, T_p^{NH} follows $GEV\left(\mu_T^{NH}, \psi_T^{NH}, \xi_T^{NH}|DWT_i\right)$, and D^{NH} follows an empirical distribution for each DWT_i. In the GEV distribution, μ is the location parameter, ψ is the scale parameter and ξ is the shape parameter [*Coles et al.*, 2001]. As a second example, a bimodal sea state, composed of sea and SH swell, $\mathbf{I} = \{1,0,1\}$, has components $\left\{H_s^{SEA}, T_p^{SEA}, D^{SEA}, H_s^{SH}, T_p^{SH}, D^{SH}\right\}$, where H_s, T_p, and D follow the same distributions as above for each wave system and for each DWT_i. In this particular case, T_p^{SEA} also follows an empirical distribution for each DWT_i, due to its improper fit to a GEV distribution.

Figures 6a–6c show the empirical and fitted probability density function for the marginal distributions of H_s, T_p, and D, respectively, for SEA (yellow), NH (cyan), and SH (magenta) wave systems. We can identify
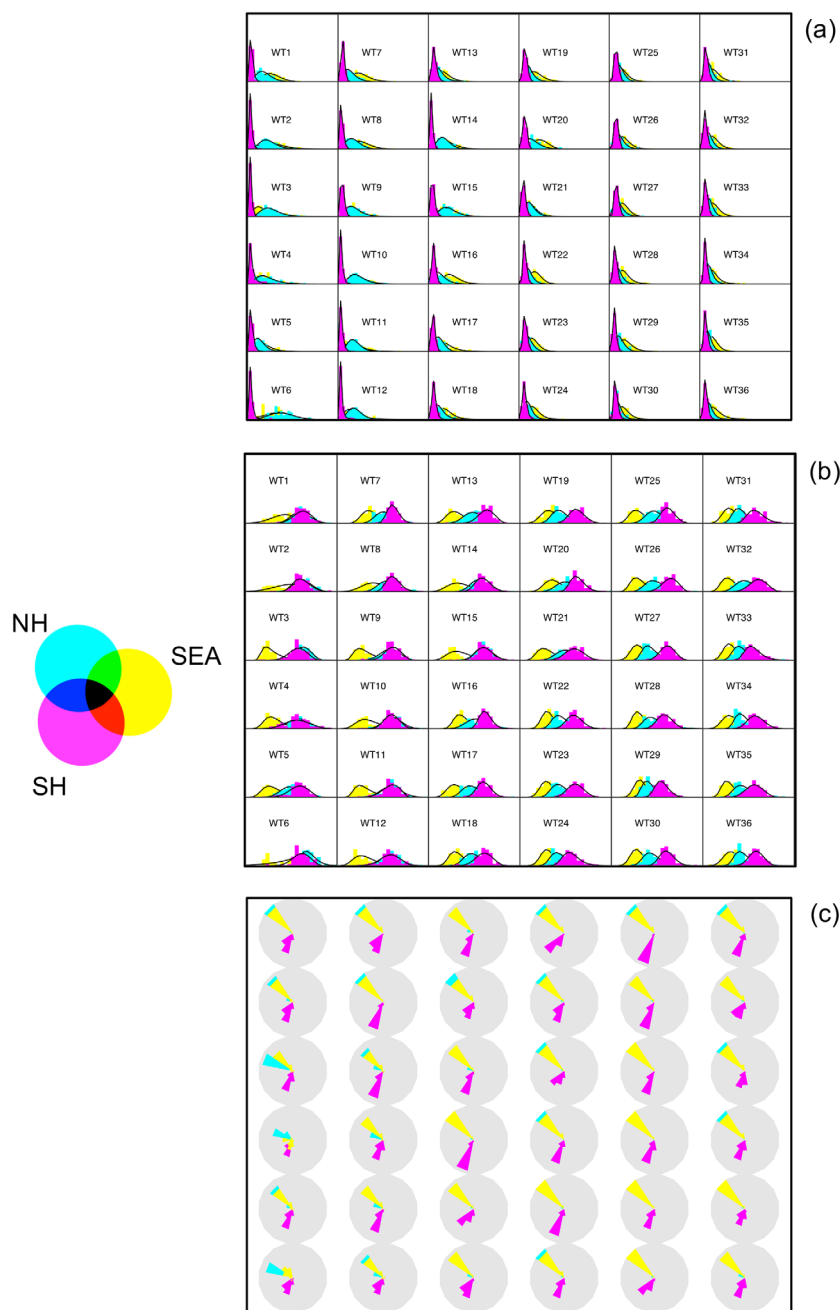
Figure 6. (a) Marginal empirical and fitted distributions for *Hs*, scale [0,11 m]; (b) marginal empirical and fitted distribution for *Tp*, scale [0,25 s]; (c) empirical distribution for *D*, scale [0,360°]. SEA in yellow, NH in cyan, and SH in magenta.

specific synoptic weather patterns such as low-pressure systems centered over the Northern Hemisphere (DWT6 in Figure 3) that produce the large NH swells (Figure 6a) in the target area. Large SEA components (Figure 6a) are associated with high-pressure systems centered over California (DWT7, DWT16, DWT20, or DWT22 in Figure 3). On the other hand, SH swells are more energetic in the DWTs that are active in the austral winter (DWT25 and DWT26). Figure 6c depicts differences in the marginal distributions of wave direction *D* associated with the position of the low- and high-pressure systems (NH swells range from NW to WNW, SH swells range from SSW to SW, and SEA waves coming predominantly from the WNW with larger variability). The variation in the intensity of each variable according to the daily weather types reflects the importance of modeling multivariate wave climate according to large-scale atmospheric predictors.
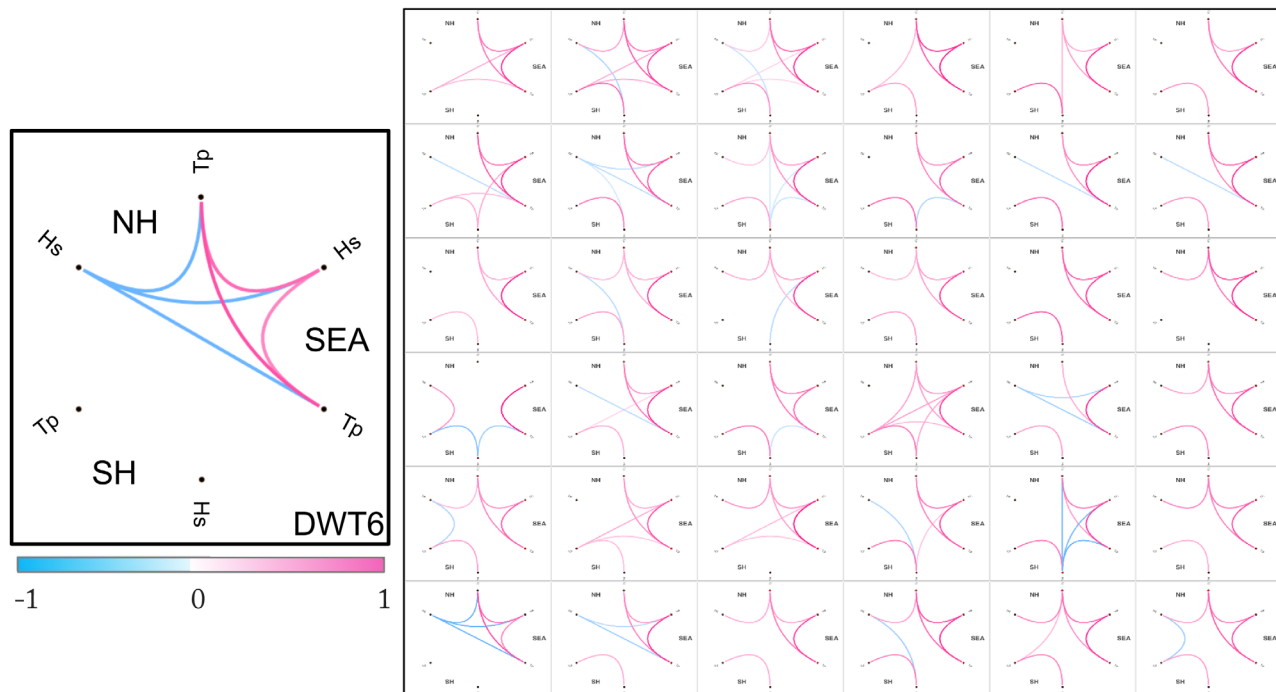
Figure 7. Graphical matrices of correlation coefficients ρ_{ab} of Hs and Tp for the SEA, NH, and SH components. Color scale is defined among -1 (blue) for negative correlation, 0 (white) for no correlation, and 1 (pink) for positive correlation.

5.3. Multivariate Gaussian Copulas

The marginal distributions for H_s, T_p, and D are transformed into a normal distribution, N(0,1), and the statistical dependence structure is modeled by a multivariate Gaussian copula, $MVN(0, \sum)$ [*Ben Alaya et al.*,
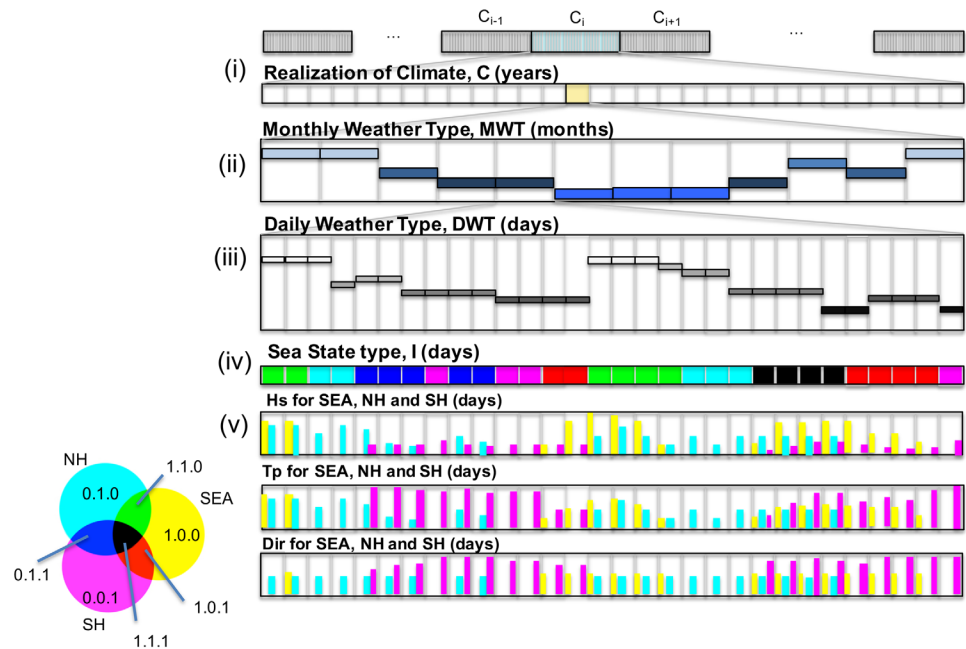

Figure 8. Sketch of the multiscale emulator of multipeak spectra in southern California, considering three wave systems, a local wind sea (SEA), and two swells (from Northern Hemisphere, NH, and from Southern Hemisphere, SH). Color represents unimodal sea states (yellow, cyan, and magenta), bimodal sea states (red, green, and blue), and trimodal sea states (black). The climate emulator works at multiple scales for the predictors (I: realization of 35 year climates; II: monthly weather types; III: daily weather types) and for the predictand (IV: sea-state type; V: sea-state parameters).
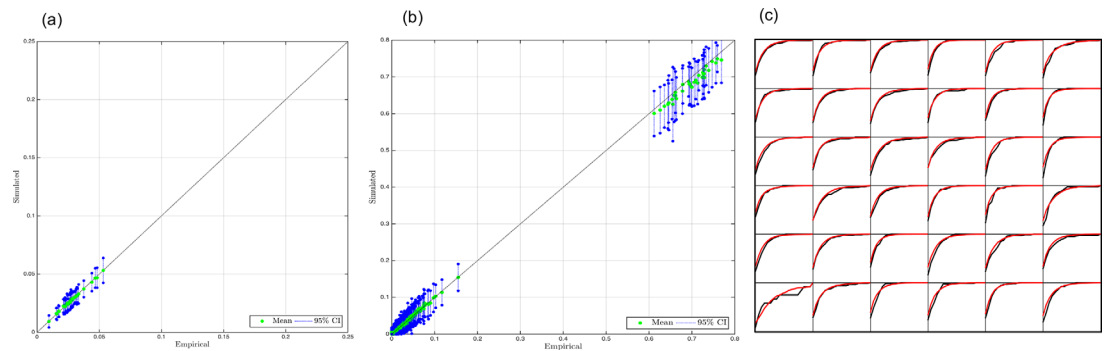
Figure 9. (a) Scatterplot of the empirical occurrence probabilities associated with the WTs versus Monte Carlo simulation results; (b) scatterplot of the empirical transition probabilities between WTs versus Monte Carlo simulation results; (c) empirical cumulative distribution for the persistence for the 36 WTs related to: (1) historical data and (2) sampled data using Monte Carlo method; horizontal axis [0 10] days, vertical axis, CDF [0,1]; black line represents empirical distribution and red the simulated one.

2014]. For example, sea-state type $\mathbf{I} = \{0,1,0\}$ has three components $\left\{ H_s^{NH}, T_p^{NH}, D^{NH} \right\}$ and a correlation matrix given by

$$\sum = \begin{pmatrix} 1 & \rho_{H_s^{NH}T_p^{NH}} & \rho_{H_s^{NH}D^{NH}} \\ \rho_{H_s^{NH}T_p^{NH}} & 1 & \rho_{T_p^{NH}D^{NH}} \\ \rho_{H_s^{NH}D^{NH}} & \rho_{T_p^{NH}D^{NH}} & 1 \end{pmatrix}$$

where ρ_{ab} is the correlation coefficient between variables a and b.

Grouping the predictand variables into daily weather types, $MVN(0, \sum |DWT_i)$, satisfies the requirement that the marginals of a Gaussian copula must be identically distributed. Variables describing multimodal spectra are often strongly correlated to each other and to the weather pattern leading to their generation. Figure 7 shows the correlation coefficients between H_s and T_p for the trimodal case $\mathbf{I} = \{1,1,1\}$. In Figure 7, strong positive and negative correlations are depicted with pink and blue lines, respectively. As expected, we observe large positive correlations between H_s and T_p for SEA for all daily weather types as evidenced by the common occurrence of the pink lines. In general, positive correlations exist for many of the DWTs; although each pattern presents particularities. Overall, the occurrence of large correlations for most sea states reinforces the advantage of splitting up the multivariate data into families (DWTs) representing homogeneous atmospheric forcing conditions.

6. Climate-Based Stochastic Simulation

Figure 8 shows a schematic representation of the multiscale emulator applied to the three wave modes (SEA, NH, and SH) at the target area offshore of southern California. The emulator works over multiple time scales for the predictor and the predictand, as explained below.

6.1. Simulation of the Predictor
We perform (i) N_C realizations of a climate (each of length $T_C = 31$ years representing the period 1979–2009); (ii) simulation of $n_\tau = N_c \times T_c \times 12$ monthly weather types, $X_\tau^m, \{\tau = 1, \ldots, n_\tau\}$; (iii) simulation of $n_t = N_c \times T_c \times 365$ daily weather types, DWTs, conditioned to the MWT, $X_t^d, \{t = 1, \ldots, n_t\}$. The first DWT, X_1^d, is generated randomly and conditioned to the first MWT, X_1^m. We perform $N_C = 32$ realizations of the climate over the period 1979–2009 to validate the Monte Carlo simulation of the predictor. Figure 9 compares the current model to the original occurrence probabilities of DWTs (Figure 9a), transition probability matrix between DWT (Figure 9b), and persistence analysis of DWT (Figure 9c), i.e., the cumulative distribution function (CDF) of the number of consecutive days in a particular DWT. In all the cases, the modeled results (occurrence probabilities and transition matrix) are close to the diagonal, which represents a perfect fit, and the average empirical and simulated CDFs show good agreement. In Figure 9b, the points with probabilities in the range 0.6–0.75 are those representing the probability of staying in the same group. These results confirm the
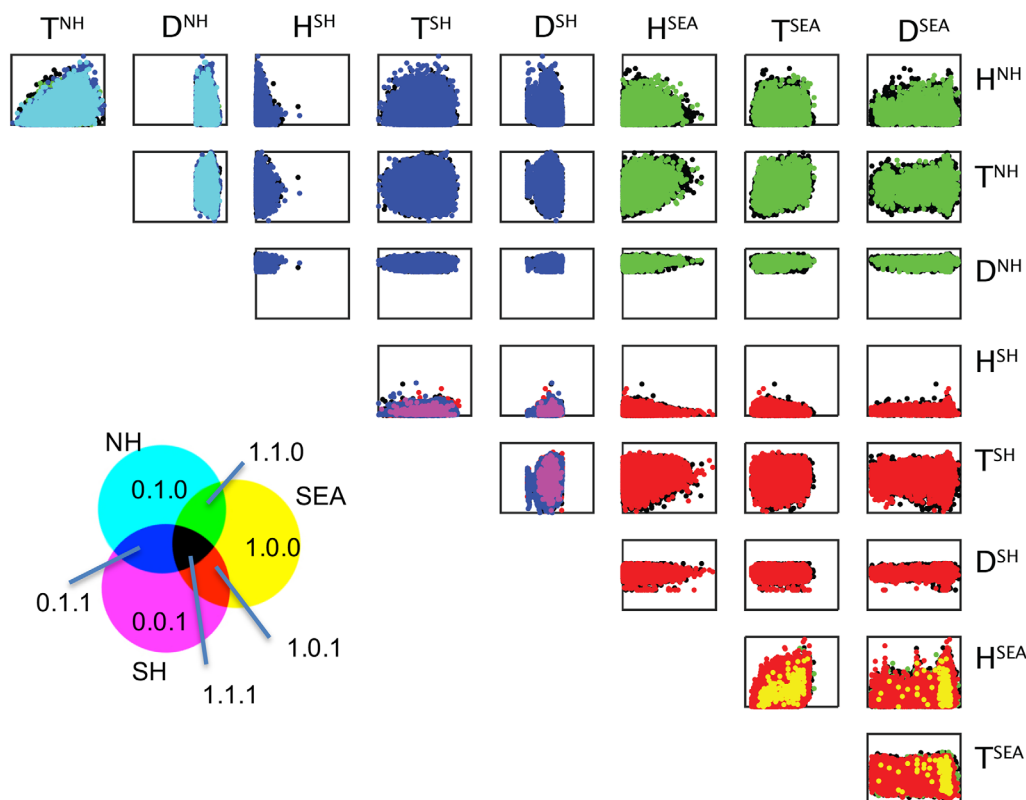
Figure 10. Scatterplots of $N_C \times T_C = 992$ years of simulation of the nine variables (Hs, Tp, and D for SEA, NH, and SH components). Color represents unimodal sea states (yellow, cyan, and magenta), bimodal sea states (red, green, and blue), and trimodal sea states (black). Scales: H [0, 11 m], T [0, 25 s], D[0, 360°].

capability of the model to reproduce synthetic sequences of DWTs in terms of occurrence probability, transition between DWTs, and persistence in each DWT.

6.2. Simulation of the Predictand

Once a long synthetic time series of the daily predictor has been constructed, we perform simulations of the sea-state type, I_t, and sea-state parameters H_s, T_p, and D for the SEA, (yellow), NH (cyan), and SH (magenta) based on the marginal fits and the Gaussian copula associated to each DWT. Figure 10 shows the scatterplots associated with a Monte Carlo simulation with $N_C = 32$ realizations of a climate of $T_C = 31$ years, obtaining 362,080 daily events. A visual inspection of the scatterplots reveals that the simulated values follow the magnitude and dependence structure quite well.

We have also performed 100 simulations with 10 years of the historical SLP data (2000–2009) in order to compare the historical and simulated histograms of the nine variables of interest (H_s, T_p, and D for SEA, NH, and SH). The results for this validation period, (shown in Figure 11a), demonstrate good agreement between the historical and modeled sea states. We also analyze the ability of the model to reproduce the chronology of the most common sea states [(p_{101}, SH + SEA), (p_{011}, SH + NH), and (p_{111}, SEA + NH + SH)] compared to the historical occurrence probabilities (Figures 11b and 5). In Figure 11b, the grey intervals represent the 95% confident intervals of the simulated occurrence probabilities, and the colored lines represent the historical occurrence probabilities. This analysis demonstrates that the emulator is able to reproduce the nonstationarity of the wave system while accounting for its inherent (intra-monthly) variability.

As it has been shown, a benefit of the proposed model is its ability to capture the nonstationary behavior of climate by accounting for the variability of weather patterns over time [see *Rueda et al.*, 2016a]. In the current Monte Carlo method, we have used the realization of the climate in 1979–2009, obtaining stationary transition probability matrices. However, we could address long-term changes associated with global
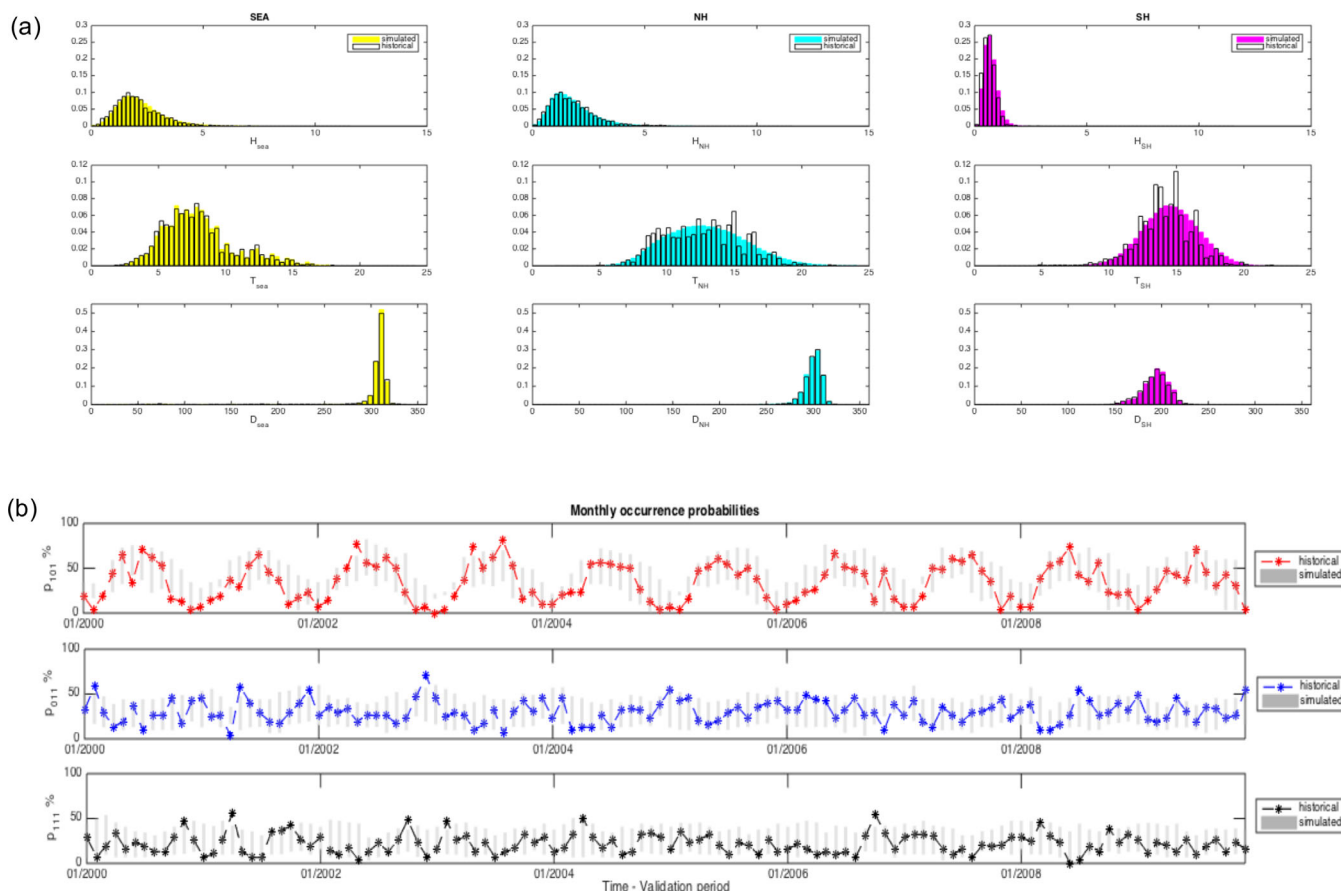
Figure 11. (a) Historical (black) and simulated (colored) histograms of the nine variables (H_s, T_p, and D for SEA(yellow), NH(blue), and SH(magenta) components). (b) Monthly occurrence probabilities of three sea states [(p_{101}, SH + SEA), (p_{011}, SH + NH), and (p_{111}, SEA + NH + SH)]. The analysis is performed during the validation period 2000–2009.

climate model scenarios simply by changing the occurrence probability of weather types and the transition probability matrices.

7. Summary and Conclusions

This paper presents a new climate-based model that stochastically simulates time series of multimodal directional wave spectra. The model is based on a predictor-to-predictand synoptic regression-guided classification model that groups daily multimodal directional wave spectra according to similar atmospheric conditions over the wave-generation regions, namely, daily weather types, which evolve following the time dependence of the historical time series. The multimodal directional wave spectra are parameterized in terms of H_s, T_p, and D for different wave systems (wind sea and swells from different geographical origins), resulting in different sea-state types (unimodal or multimodal). Stationary extreme value models for the marginals of H_s and T_p and the empirical distribution of D are applied for each wave system in each weather type, and a Gaussian copula is used to account for the statistical dependence between the variables. Here we have focused on a 9-D problem (H_s, T_p, and D for three wave systems: one wind sea and two swells), but the method is scalable to arbitrary numbers of wave systems and variables due to the flexibility of the multivariate Gaussian copula. Nonstationarity (e.g., seasonality and interannual variability) is introduced in the model through a time-dependent transition probability matrix for daily weather types.

We apply the model to a location of interest offshore of southern California, affected by local seas and swells from the Northern and Southern Hemispheres. The model identifies the weather types related to specific unimodal, bimodal, or trimodal sea-state types and generates synthetic time series of multimodal wave conditions. The multivariate model presented here accounts for the chronology associated with daily, monthly, and annual scales of the sea states by means of a time-dependent Markov chain model.

We believe that the newly developed model framework can help to characterize the stochastic behavior of the time-dependent boundary conditions needed for coastal impact studies. The statistical downscaling approach allows to expand the extent of sparse historic data sets and probabilistically model coastal flooding and shoreline evolution across time periods ranging from hindcast to seasonal forecasting or climate-change projections.

References

Antolinez, J. A. A., F. J. Mendez, P. Camus, S. Vitousek, E. M. Gonzalez, P. Ruggiero, and P. Barnard (2016), A multiscale climate emulator for long-term morphodynamics (MUSCLE-morpho), *J. Geophys. Res. Oceans*, *121*, 775–791, doi:10.1002/2015JC011107.

Bárdossy, A., G. Pegram, S. Sinclair, J. Pringle, and D. Stretch (2015), Circulation patterns identified by spatial rainfall and ocean wave fields in Southern Africa, *Frontiers Environ. Sci.*, *3*(31), doi:10.3389/fenvs.2015.00031.

Ben Alaya, M. A., F. Chebana, and T. B. M. J. Ouarda (2014), Probabilistic Gaussian copula regression model for multisite and multivariable downscaling, *J. Clim.*, *27*, 3331–3347, doi:10.1175/JCLI-D-13-00333.1.

Boukhanovsky, A., and C. Guedes Soares (2009), Modelling of multipeaked directional wave spectra, *Appl. Ocean Res.*, *31*, 132–141.

Boukhanovsky, A., L. J. Lopatoukhin, and C. Guedes Soares (2007), Spectral wave climate of the North Sea, *Appl. Ocean Res.*, *29*, 146–154.

Callaghan, D. P., P. Nielsen, A. Short, and R. Ranasinghe (2008), Statistical simulation of wave climate and extreme beach erosion, *Coastal Eng.*, *55*(5), 375–390, doi:10.1016/j.coastaleng.2007.12.003.

Callaghan, D. P., R. Ranasinghe, and D. Roelvink (2013), Probabilistic estimation of storm erosion using analytical, semi-empirical, and process based storm erosion models, *Coastal Eng.*, *82*, 64–75, doi:10.1016/j.coastaleng.2013.08.007.

Camus, P., F. J. Méndez, I. J. Losada, M. Menéndez, A. Espejo, J. Pérez, A. Rueda, and Y. Guanche (2014a), A method for finding the optimal predictor indices for local wave climate conditions, *Ocean Dyn.*, *64*(7), 1025–1038, doi:10.1007/s10236-014-0737-2.

Camus, P., M. Menéndez, F. J. Méndez, C. Izaguirre, A. Espejo, V. Cánovas, J. Pérez, A. Rueda, I. J. Losada, and R. Medina (2014b), A weather-type statistical downscaling framework for ocean wave climate, *J. Geophys. Res. Oceans*, *119*, 7389–7405, doi:10.1002/2014JC010141.

Camus, P., A. Rueda, F. Méndez, and I. Losada (2016), An atmospheric-to-marine synoptic classification for statistical downscaling marine climate, *Ocean Dyn.*, *66*, 1589–1601, doi:10.1007/s10236-016-1004-5.

Cannon, A. J. (2012), Regression-guided clustering: A semi-supervised method for circulation-to-environment synoptic classification, *J. Appl. Meteorol. Climatol.*, *51*, 185–190.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, 208 pp., Springer, London.

Corbella, A., and D. D. Stretch (2013), Simulating a multivariate sea storm using Archimedean copulas, *Coastal Eng.*, *76*, 68–78.

Crosby, S. C., W. C. O'Reilly, and R. Guza (2016), Modeling long-period swell in southern California: Practical boundary conditions from buoys observations and global wave model predictions, *J. Atmos. Oceanic Technol.*, *33*, 1673–1690, doi:10.1175/JTECH-D-16-0038,1.

Espejo, A., P. Camus, F. J. Mendez, and I. J. Losada (2014), Spectral ocean wave climate variability based on atmospheric circulation patterns, *J. Phys. Oceanogr.*, *44*, 2139–2152, doi:10.1175/JPO-D-13-0276.1.

García-Medina, G., H. T. Özkan-Haller, P. Ruggiero, and J. Oskamp (2013), An inner-shelf wave forecasting system for the U.S. Pacific northwest, *Weather Forecast.*, *28*, 681–703, doi:10.1175/WAF-D-12-00055.1.

Guanche, Y., R. Mínguez, and F. J. Méndez (2013), Autoregressive logistic regression applied to atmospheric circulation patterns, *Clim. Dyn.*, *42*, 537, doi:10.1007/s00382-013-1690-3.

Hanson, J. L., and O. M. Phillips (2001), Automated analysis of ocean surface directional wave spectra, *J. Atmos. Oceanic Technol.*, *18*, 277–293.

Hegermiller, C. A., J. A. A. Antolínez, A. Rueda, P. Camus, J. Perez, L. H. Erikson, P. L. Barnard, and F. J. Mendez (2017), A multimodal wave spectrum-based approach for statistical downscaling of local wave climate, *J. Phys. Oceanogr.*, *47*, 375–386, doi:10.1175/JPO-D-16-0191.1.

Huth, R. (2001), Disaggregating climatic trends by classification of circulation patterns, *Int. J. Climatol.*, *21*, 135–153.

Jordan, P., and P. Talkner (2000), A seasonal Markov chain model for the weather in the central Alps, *Tellus A*, *52*, 455–469, doi:10.1034/j.1600-0870.2000.00106.x.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, *25*, 1287–1304, doi:10.1016/S0309-1708(02)00056-8.

Lucas, C., A. Boukhanovsky, and C. Guedes Soares (2011), Modeling the climatic variability of directional wave spectra, *Ocean Eng.*, *38*, 1283–1290, doi:10.1016/j.oceaneng.2011.04.003.

Mártinez-Asensio, A., M. N. Tsimpli, M. Marcos, X. Feng, D. Gomis, G. Jordá, and S. A. Josey (2016), Response of the North Atlantic wave climate to atmospheric modes of variability, *Int. Journal of Climatol.*, *36*, 1210–1225, doi:10.1002/joc.4415.

Mendez, F. J., M. Menendez, A. Luceño, and I. J. Losada (2006), Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model, *J. Geophys. Res.*, *111*, C07024, doi:10.1029/2005JC003344.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Climate change: Stationarity is dead: Whither water management?, *Science*, *319*(5863), 573–574, doi:10.1126/science.1151915.

Nelsen, R. B. (2006), *An Introduction to Copulas*, 2nd ed., Springer, New York.

Perez, J., M. Menéndez, F. J. Méndez, and I. J. Losada (2014), ESTELA: A method for evaluating the source and travel-time of the wave energyreaching a local area, *Ocean Dyn.*, *64*,1025–1038.

Perez, J., M. Menéndez, P. Camus, F. J. Méndez, and I. J. Losada (2015), Statistical multi-model climate projections of surface ocean waves in Europe, *Ocean Modell.*, *96*, 161–170, doi:10.1016/j.ocemod.2015.06.001.

Rascle, N., and F. Ardhuin (2013), A global wave parameter data base for geophysical applications. Part 2: Model validation with improved source term parameterization, *Ocean Model.*, *70*, 174–188, doi:10.1016/j.ocemod.2012.12.001.

Rueda, A., P. Camus, F. J. Méndez, A. Tomás, and A. Luceño (2016a), An extreme value model for maximum wave heights based on weather types, *J. Geophys. Res. Oceans*, *121*, 1262–1273, doi:10.1002/2015JC010952.

Rueda, A., P. Camus, A. Tomás, S. Vitousek, and F. J. Méndez (2016b), A multivariate extreme wave and storm surge climate emulator based on weather patterns, *Ocean Modell.*, *104*, 242–251, doi:10.1016/j.ocemod.2016.06.008.

Saha, S. (2010), NCEP Climate Forecast System Reanalysis (CFSR) 6-hourly products, January 1979 to December 2010, https://doi.org/10.5065/D69K487J, Natl. Cen. for Atmos. Res., Comput. and Inf. Syst. Lab., Boulder, Colo.

Salvadori, G., C. De Michele, N. T. Kottegoda, and R. Rosso (2007), *Extremes in Nature. An Approach Using Copulas, Water Science and Technology Library*, vol. 56, Springer, Netherlands.

Semedo, A., K. Suselj, A. Rutgersson, and A. Sterl (2011), A global view on the wind sea and swell climate and variability from ERA-40, *J. Clim.*, *24*, 1461–1479.

Serafin, K. A., and P. Ruggiero (2014), Simulating extreme total water levels using a time-dependent, extreme value approach, *J. Geophys. Res. Oceans*, *119*, 6305–6329, doi:10.1002/2014JC010093.

Solari, S., and M. A. Losada (2011), Non-stationary wave height climate modeling and simulation, *J. Geophys. Res.*, *116*, C09032, doi:10.1029/2011JC007101.

Stockdon, H. F., R. A. Holman, P. A. Howd, and A. H. Sallenger (2006), Empirical parameterization of setup, swash, and runup, *Coastal Eng.*, *53*, 573–588.

Tolman, H. L. (2014), User manual and system documentation of WAVEWATCH III versión 4.18, *NOAA/NWS/NCEP/MMAB Tech. Note 316*, pp.194 + Appendices, Natl. Oceanic and Atmos. Admin., Silver Spring, Md.

Trulsen, K., J. C. Nieto Borge, O. Gramstad, L. Aouf, and J.-M. Lefèvre (2015), Crossing sea state and rogue wave probability during the Prestige accident, *J. Geophys. Res. Oceans*, *120*, 7113–7136, doi:10.1002/2015JC011161.

U.S. Army Corps of Engineers (2002), *Coastal Engineering Manual Part II: Meteorology and Wave Climate Engineer Manual 1110-2-1100*, Washington, D. C.

Vincent, L., and P. Soille (1991), Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Trans. Pattern Anal. Mach. Intel.*, *13*, 583–598.

Walstra, D. J. R., R. Hoekstra, P. K. Tonnon, and B. G. Ruessink (2013), Input reduction for long-term morphodynamic simulations in wave dominated coastal settings, *Coastal Eng.*, *77*, 57–70, doi:10.1016/j.coastaleng.2013.02.001.