WAVE ENERGY FLUX VARIABILITY ALONG CONTINENTAL SHELVES OF THE PACIFIC NORTHERN ANDES: A WAVE CLIMATE DATA-DRIVEN APPROACH

By

Esteban Ramos-Chavarriaga

Research work presented to the Department of Earth Sciences and the School of Sciences of Universidad EAFIT in partial fulfillment of the requirements for the degree of

Master in Earth Sciences (*Magíster en Ciencias de la Tierra*)

Universidad EAFIT

2021

To my advisor Juan F. Paniagua Arroyave “Pani”, for his guidance, care, and freedom.

To my mentor Daniel Velásquez P., for planting the seed of rigorous research.

A mi madre Mónica y padre Juan Diego siempre, infinitas gracias.

When despair for the world grows in me
and I wake in the night at the least sound
in fear of what my life and my children's lives may be,
I go and lie down where the wood drake
rests in his beauty on the water, and the great heron feeds.
I come into the peace of wild things
who do not tax their lives with forethought
of grief. I come into the presence of still water.
And I feel above me the day-blind stars
waiting with their light. For a time
I rest in the grace of the world, and am free.

Wendell Berry

Acknowledgments

The author would like to acknowledge the Vice presidency of Discovery and Creation at EAFIT University for funding this research project and master thesis. All the academic and personal training and support received in becoming a transparent, pedagogic, and passionate scientist has been possible thanks to EAFIT University. For that, I will always be extremely grateful.

I would like to specially acknowledge my advisor Juan Felipe Paniagua-Arroyave, for supporting me and my ideas through our journey together. Additionally, his preference on research groups based on transdisciplinary ideas and backgrounds in the Earth Sciences is a fundamental paradigm he brings to our institution and plants in our students. It is also something I am particularly thankful for.

This research project could not be possible without the care, help and guidance of Juan Darío Restrepo, Olga L. Quintero, José F. Duque, Marcela Jaramillo, Juan J. Restrepo B., and many others whose conversation enriched widely this document.

Content

List of figures

List of tables

Glossary

P – wave energy flux (W/m)

Q_s – sediment transport (kg/s)

WW3 – WAVEWATCH III®

R^2 – determination coefficient

r – Pearson or correlation coefficient

NRMSE – normalized root mean square error (%)

ENSO – El Niño southern oscillation

ONI – oceanic Niño index (°C)

SVD – singular value decomposition

PCA – principal component analysis

EOF – empirical orthogonal functions

DMD – dynamic mode decomposition

ROM – reduced-order model

PCS – principal component space

MMN – monthly mean normalization

PDE – partial differential equation

ODE – ordinary differential equation

Introduction

Research proposal

Dominant dynamics of available wave energy dictate coastal gradients in alongshore sediment transport over wave-dominated continental shelves, where large-scale coastal evolution depends upon wave-induced sediment fluxes. We propose to study wave energy flux variability (P in W/m) from high-fidelity simulations of wave climate, to classify the fundamental dynamic patterns of P into morphodynamical regimes based on pioneering classifications, serving as a data-driven indicator for large-scale, and wave-induced coastal evolution. Additionally, we describe and predict the variability of P in a complex system framework with a dimensionality reduction approach to produce a reduced-order model (ROM) of the system of interest. Where we use spatio-temporal data to discover, reduce, and solve a symbolic expression that forecasts the coastal variability of P along compressional continental shelves of the Pacific Northern Andes.

In particular, the problems we tackle in this document are: **i**) the classification of P into low-rank regimes based on wave climate metrics, pioneering morphodynamical and geological characterizations, as a low-dimensional dynamic representation; **ii**) the description of temporal patterns using spectral and modal analysis on time series data to identify climate-driven periodicities (i. e. El Niño Southern Oscillation); and finally, **iii**) the discovery and evaluation of a ROM of the dynamical system, from finding the nonlinear governing terms to numerically solving the model for future states. Overall, we pretend to extract physical intuition from the dynamical system and PDE, to gain morphodynamical insights and interpretable results. In our view, modeling mesoscale changes in the amount of energy available in ocean waves to transport sediment along continental shelves would complement previous studies, and move forward our understanding of large-scale coastal evolution in the Pacific Coast of the Northern Andes, outlining the importance of leading-order variables in such complex systems.

Background and context

Ocean surface gravity waves exert decisive control on the morphodynamics of nearshore environments (Komar & Holman, 1986), with the shoreface being defined as the transitional zone between the continental shelf and the shoreline in which long-period waves ("ordinary waves") interact with the seabed (Hamon-Kerivel et al., 2020). Their origin allows a separation into two regimes: locally wind-generated waves, called local waves, and waves that outrun their generating wind, called swell (Holthuijsen, 2007; Thompson et al., 1996). These wave environments, or wave fields carrying energy, interfere with each other and scatter over ocean bedforms, creating long-term complex wave variability commonly referred to as "wave climate" (Hallermeier, 1980).

The use of operational wave climate predictions (based on phase-averaged properties) provides an alternative to modeling fully nonlinear wave dynamics, given the complexity of wave transformations over variable bathymetry (Athanassoulis & Belibassakis, 1999; Sheremet et al., 2016). This application has proven useful along the coastlines of developing countries that lack proper instrumentation (Osorio et al., 2016). Using reanalysis models such as WAVEWATCH III® (WW3), an approximated solution is obtained by empirically applying source terms that represent wave evolution, along with wind and temperature datasets to output statistically significant wave parameters (Montoya et al., 2013; Morim et al., 2020; Portilla et al., 2015). Furthermore, hindcast datasets are in good agreement with buoy measurements (Bromirski et al., 2013) and allow quantifying long-term coastal morphodynamics from wave climate (Nienhuis et al., 2020). As a complex dynamical system, wave-induced coastal morphodynamics can be analyzed from models of dominant dynamic patterns in wave climate.

Recent work on dynamical systems and big data analytics has provided new data-driven tools, algorithms, and graphical representations of complex, multivariate, and high-dimensional systems. Through robust statistics and computational learning tools, data-driven models produce accurate representations and parsimonious models of nonlinear behavior (Brunton et al., 2020). Furthermore, the use of data-driven models to reduce computational expense, solvent analytical complexity, and generate physical intuition on the fundamental dynamical system (Thomson & Emery, 2014), has had significant results over the last two decades (Bai et al., 2020; Köhler et al., 2010; Rudy et al., 2017; Zou et al., 2006).

We use the philosophical and scientific framework of complex systems science, where information from parts of the system and their non-trivial interactions are more important than information within the parts themselves (Sayama, 2015). Additionally, we owe to consider the Anthropocene thesis that we live in a new epoch of human-driven climate forcing (Syvitski et al., 2020), where the study of Earth systems is presented with a new paradigm that climate balance is no longer only driven by planetary (Milankovitch) cycles. Consequently, changes from 1980 to 2010 would represent both anthropic and Natural forcing factors. In this case, assessing wave energy variability requires considering the coupled wave-atmosphere system controlled by synoptic climate drivers, nonlinear interactions, and complex leading-order dynamics (Christakos et al., 2020), to develop a model which helps understanding potential climate change impact and future coastal evolution in the Pacific Northern Andes (Figure 1).

Specific objectives

1. Research the scientific background and context regarding 1) oceanographic and morphodynamic studies on P dissipation, distribution, and impact along continental shelves; as well as 2) the applied mathematics developed to model such dynamical systems from high-fidelity data.
2. Acquire and explore reanalysis wave parameters of hindcast wave climate from an operational wave prediction model (WW3) to numerically calculate P from time-series data (from 1980 to 2010).
3. Apply pre-processing and processing stages to compute time series outlining seasonal, anomaly, and spectral components in the variability of P along continental shelves.

4. Apply classification and clustering techniques to coastal values of P associated to wave climate metrics and coastal morphodynamics, in a reduced-order representation of hierarchical modal behavior.
5. Develop and evaluate a modeling architecture, based on combining data-driven techniques, analytic frameworks, and numerical schemes, to build a mathematical model that predicts accurate future states of reduced-order dynamics of P.
6. Generate physical intuition and interpretability discussions from the developed model, to better understand wave-induced morphodynamics and coastal evolution.

About the structure of this document

This thesis is divided into two chapters. Chapter 1 focuses on the exploratory analysis of wave climate and its relation to shoreface morphodynamics through the variable wave energy flux, together with the temporal identification of climate-driven periodicities in the data. Chapter 2 focuses on the applied mathematics used to obtain a reduced-order model of wave energy flux from data-driven methods and linearization schemes. Both chapters are presented in the form of the final manuscripts to be submitted for peer review. Therefore, some context may be repetitive in the introductory sections of both chapters. Nevertheless, both chapters present separate approaches with relevant discussions, while the conclusions of the research project are presented lastly encompassing both chapters.

1. Chapter 1: Dynamics of Wave Energy Flux along the Pacific Northern Andes Compressional Shelf

Abstract

The shelf morphology along the Northern Andes Pacific coast exerts control on wave energy flux alongshore variability. In turn, available wave energy dictates gradients in alongshore sediment transport, shaping coastline morphology at various spatio-temporal scales. Here, we analyze wave energy flux variability from wave climate simulations along the Northern Andes Pacific coast to classify the fundamental patterns into morphodynamical regimes. We identify three regimes in a low-rank, optimal coordinate system from our data-driven modeling architecture. Further, we associate these regimes with shoreface morphodynamics and coastal morphology classifications, connecting our modeling architecture with wave-induced sediment transport metrics along continental shelves. Our results provide a modeling architecture of wave climate, illuminating the importance of hidden leading-order variables in complex systems, such as mesoscale wave-induced sediment transport along continental shelves.

1.1. Introduction

Wave climate variability, along with sea-level rise, will be dominant factors impacting shoreline evolution in the coming decades (Syvitski et al., 2020). Particularly considerable impacts are found at tropical coasts, such as South America's Northern Andes compressional margins and continental shelves (Bender & Dean, 2003; Mortlock & Goodwin, 2015; Restrepo & Kjerfve, 2002; Salazar et al., 2018). Wave energy variability, driven by climate, tectonics, wave-wave interactions, and dissipation mechanisms, has been proven to relate significantly with seafloor topography at the nearshore (Ardhuin et al., 2003, 2009). The almost stochastic interplay of oceanographic, hydroclimatic, and morphodynamics factors creates a complex dynamical system of nearshore wave energy, in which nonlinear processes regulate, balance, and produce the resultant wave behavior and observed coastal impact.

Understanding wave climate spatio-temporal variability plays a fundamental role in forecasting littoral feedbacks along wave-dominated coasts. These feedbacks include wave-driven sediment transport at wave-dominated deltas (Almar et al., 2021; Nienhuis et al., 2015), shoreline evolution under unstable wave regimes (Ashton et al., 2001), and wave transformation over-complicated nearshore bathymetry (Paniagua-Arroyave et al., 2019). In that sense, hindcasting dominant dynamics of wave energy associated with sediment transport allows forecasting mesoscale wave-induced morphodynamics and climate-driven patterns in coastal evolution.

On the other hand, the theoretical-empirical description used in most operational wave prediction models applied in the ocean waves modeling community (WW3 included) is currently challenged by a new view of sea-wave physics based on resonant nonlinear wave interactions controlling the mechanism of wave growth rather than wind forcing (Zakharov et al., 2015, 2017). Although both frameworks result in good approximations of observed wave behavior, they differ mostly on the physical phenomena driving the sea-wave evolution. Overall, simulated wave energy dissipation is not fully operational at variable bathymetry, but significant values can be found at the boundary between the open ocean and nearshore environments, applicable to our implementation. This boundary is defined as the shoreface by Hamon-Kerivel et al (2020). Where the upper shoreface is mostly affected by wave hydrodynamics, and the lower shoreface is more heavily influenced by geological factors, climate change, and sea-level rise (Hapke et al., 2016).

As an emergent complex system, mesoscale wave-induced coastal changes are dominated by wave energy spatio-temporal evolution. Consequently, high-fidelity simulations of wave climate provide means for understanding wave energy variability on a global scale, from mesoscale to synoptic changes, along unmonitored coasts (Nienhuis et al., 2020). Which combined with data-driven techniques, these high-fidelity results allow for unraveling wind-wave and wave-induced dynamic patterns (Brunton et al., 2020).

Here, we analyze wave climate data from a coastal morphodynamics standpoint to develop a data-driven modeling architecture that assesses wave energy flux's (P) nonlinear dynamics along continental shelves. We classify the dominant patterns and predict future dynamical states of the system. We apply this architecture as a study case in the Pacific Coast of the Northern Andes compressive margin.

1.2. Background: Modeled wave climate

Since WW3 accounts for wave scattering and dissipation, we can explore correlations between the spatial variability in P and seabed morphology. WW3 uses source terms based on bottom friction coefficients obtained from swell observations in shallow water by the Joint North Sea Wave Project (JONSWAP) for swell and local wind waves. This value represents sea bottoms such as shelf seas, tidal regions, and lakes with sand, gravel, and fine clay materials (Zijlema et al., 2012).

In the WW3 framework, surface gravity waves are described by amplitude and phase parameters, i.e., wavenumber (k or $\boldsymbol{k}$ in vector form), direction (θ), intrinsic and angular frequencies (σ and ω, respectively). The model assumes linear wave theory such that the scale variations in depth and currents are much larger than those of an individual wave perturbation. To include mean current effects, a distinction is made between relative and absolute frequencies, (f_r and f_a, respectively).

The model also assumes incompressible fluids and irrotational flows to solve the wave action balance equation by incorporating source terms that account for wind input, wave transformation (including nonlinear evolution) (Ardhuin et al., 2003), bottom friction, and scattering (Hasselmann et al., 1980), and energy dissipation (Chawla et al., 2011; Phillips, 1977). The model simplifies the wave description by using the variance density spectra denoted as $F(k,\sigma,\text{f};x,y)$, where and the wave action density spectrum is defined as $N(k,\theta;x,t) = F(k,\theta\colon x,t)/\sigma$, within a wave action balance supported by the linear theory (i.e. Chiang & Stiassnie, 2005, section 3.1). WW3 uses the balanced equation for the wave action spectrum $N(k,\theta;x,t)$ such that:

$$^{dN}/_{dt} = {}^{S}/_{\sigma} \tag{1}$$

where the total derivative includes variability in time, space, wavenumber, and direction, and S represent the effect of various processes in the wave variance density as sources and sinks of wave energy. See Chawla, Spindler, and Tolman (2011) for a detailed mathematical description.

1.3. Background: The Pacific coast of the Northern Andes

The study area where we apply the developed architecture is the Northern Andes Pacific basin and the Colombian Pacific coastline (Figure 1). From the border with Ecuador (1.3°N, 79°W) in the westernmost branch of the South American Andes (the Cordillera Occidental, in Spanish) to the border with Panama (7°N, 78.5°W). The coastline length is approximately 864 km with a linear distance of 629.3 km. It is characterized by heavy rainfall, numerous rivers, and luxuriant vegetation (Correa & Morton, 2010; Morton et al., 2000) showing overall tropical meteorological conditions (Portilla-Yandún et al., 2019).

The geological configuration of the littoral zone is characterized by a narrow continental shelf varying in dimensions of tens of kilometers (≈ 8 km at Solano Bay and ≈ 70 km at Mira delta) before a slope zone leading into the deep ocean. The resultant variety of beaches along the coast is due to de complex geological history of the Northern Andes; a combination of big rivers with dense vegetation and mangroves, rainfall, as well as high tectonic and seismic activity (Correa & Gonzalez, 2000; Latrubesse & Restrepo, 2014; West, 1956). According to the Colombian Pacific Basin Oceanographic Compilation II (Dimar, 2020) the Pacific region can be classified into two distinct morphodynamical regimes related to the main hydrodynamic, geographic, and oceanographic conditions: The Northern Pacific and Southern Pacific. The Northern Pacific is defined from Panama to Cabo Corrientes with an approximate distance of 375 km, characterized by high sloping coast with cliffs, pillars, islands, caverns, and littoral arcs, associated with the Baudó ridge (Ingeominas, 1996). To the south, between Cabo Corrientes and the border with Ecuador, the coast is framed within what is known as the plain Pacific coastline (Dimar, 2020), characterized by an almost flat relief, formed by large delta plains and extensive complexes of mangrove marshes, aligned parallel to the coastline, generated under a macro-tidal regime (Restrepo & Kjerfve, 2002), with gently sloping platforms cut into Tertiary sedimentary rocks (Correa & Morton, 2010).

Predominant swell environments arrive from the southwest with an average wave power of $\approx 16\,x\,10^{6}\,ergs\,s^{-1} \approx 1.6\,W$ at 9 m depth (Restrepo & Kjerfve, 2002; Restrepo & López, 2008). Significant swells and coastal currents as well as strong El Niño Southern Oscillation (ENSO) events causing sea level rise are reported to affect littoral morphodynamics (Portilla et al., 2015). The average wave power obtained by field measurements in 3 deltaic systems (San Juan, Mira, and Patía) is roughly 1.6 W (Restrepo, 2008). Spectral wave conditions on the Pacific coast of Colombia are reported using a method to identify wave features and group them into regimes employing a 2D spectral indicator associated with the occurrence probability of spectral partitions (Portilla et al., 2015). They conclude that wave characterization by the proposed indicator does not include information on wave energy. Hence, regional studies of wave conditions do not focus on wave energy variability and its relationship to morphological factors of the nearshore continental shelf and Pacific basin. We present this report to broaden the view on wave energy variability, availability, and dissipation, as well as its relation to coastal morphodynamics in the Northern Pacific coast of South America.

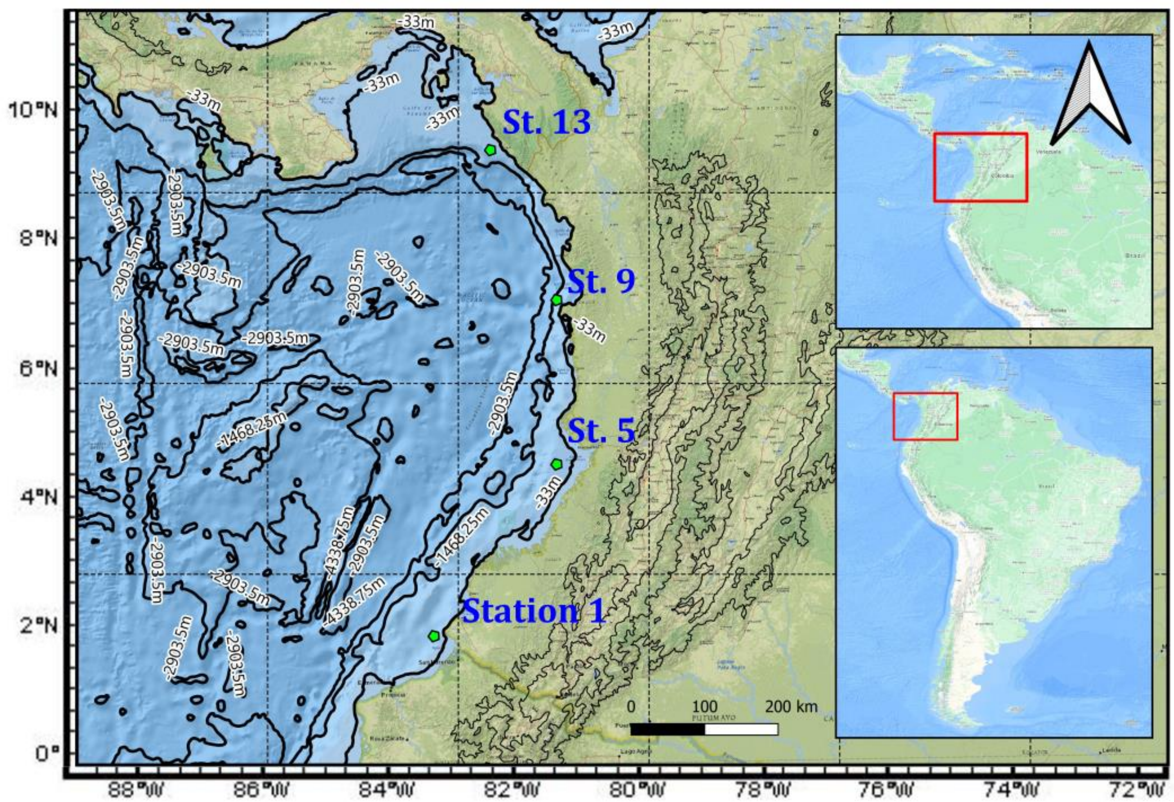

Figure 1. The Pacific oceanic basin and coastline of the Northern Andes at different spatial scales. Some of the study sites (red stroke points) shown in the figure represent significant results of each morphodynamically classified regime. The stations are located along the continental shelf around the contour line of 15 km of depth (specific values of depth for stations 1, 5 9, and 13 are: -224, -62, -1200, -1300 meters, respectively).

1.4. Methods: Wave energy flux from wave climate data

We calculated P from data derived from the reanalysis dataset WAVEWATCH III® 30-year Hindcast Phase 2. We obtained the wave data from the Earth Engine App Wave-Tide (https://jhnienhuis.users.earthengine.app/view/changing-shores), which outputs coastal wave parameters from WW3, as well as wave climate exploratory analysis and tidal constituents. We used time series of H_s, T_p, and θ_p from 1980 to 2010 with spatial resolution of approximately 50 km (0.5° x 0.5°) and temporal resolution of 3 hours.

Using the wave parameters from WW3 of H_s, T_p we calculated the wave number (k) to further compute the wave energy flux/transport (P in units of W/m) as (Holthuijsen, 2007 Eq. 5.5.12 therein),

$$P = Ec_g, \tag{2}$$

where $E = g\rho H_s/16$ (Lentz and Fewings, 2012), $c_g = 0.5\omega/k\left[1 + \frac{2\cdot k\cdot h}{\sinh(2\cdot k\cdot h)}\right]$ is the group velocity, $\omega = 2\pi/T_p$, and the wavenumber (k) and radian frequency (ω) follow the well-known dispersion relation (i.e. Mei & Black, 1969 Eq. 1.4.11):

$$\sigma^2 = gk\tanh(kh), \tag{3}$$

where h is the water depth, σ is the relative radian frequency, $\omega = \sigma + k\ \cdot U$, and U is the time- and depth-averaged current velocity. The exploratory analysis includes wave roses of P vs D_p, $H_s - T_p - D_p$ diagrams of swell and local waves, computation of morphodynamical parameters from wave climate statistics, and latitudinal variability of mean P values (P_{mean}) and variability coefficients (COV). The bathymetric data were obtained from the Global Multi-Resolution Topography (GMRTMapTool at https://www.gmrt.org/GMRTMapTool/). Additionally, we provide a supplementary diagram of the architecture developed as a pipeline of pre-processing, processing, descriptive, and predictive stages, attached as an appendix (Figure 13).

1.5. Methods: Data-driven analytics of wave climate

1.5.1. *Dimensionality-reduction*

We applied the Principal Component Analysis (Pearson, 1901) based on the Singular Value Decomposition (SVD) algorithm to obtain a hierarchical coordinate system that captures the maximum data variance (Yule, 1938). Based on the SVD, we computed a more robust technique called the Dynamic Mode Decomposition (DMD) (Tu et al., 2014). These algorithms use the Koopman operator theory that advances the system in time, linearizing the solution by calculating a data-driven optimal basis set. The new coordinate system of reduced dimensionality constitutes a parsimonious model of reality (Brunton & Kutz, 2017).

The PCA, and further the DMD, constitutes one of the fundamental numerical matrix decomposition techniques in the computational era (Rudy et al., 2017). The goal is to reduce dimensionality into the most significant correlation structures (or dominant patterns) representing a non-square data matrix as the product of three other matrices. According to the SVD, we can define our collection of snapshot measurements of P in time as a data matrix (X_P), such that it can be expressed as the multiplication of three other matrices as:

$$X_P = U\Sigma V^T, \tag{4}$$

where U is a unitary matrix with orthogonal columns associated with the spatial realizations of P (rows), thus called left singular vectors or spatial principal components in this case (also denoted as φ_r when truncated into the most critical components r). The Σ represents a matrix with real, non-negative values in the diagonal and zeros off the diagonal. These values represent the loadings, or "weights" (σ), which indicate how much of the original variance is explained by each principal component. The V matrix represents the right singular vectors, or temporal principal components in this case, as their rows are associated with the temporal realizations of P (columns).

We used the matrix V^T from the PCA to compute an optimal coordinate system that best represents (in the statistical regression sense) the temporal variability of each continental shelf station $(x = V^T \cdot X_p(r = 1),\ y = V^T \cdot X_p(r = 2),\ z = V^T \cdot X_p(r = 3)\ )$. By multiplying each truncated temporal mode or eigen-process (V^T) by each station's time series of X_p, we obtain a unique coordinate point representing each station in terms of the first three PCA temporal modes in the optimal coordinate system. This representation corresponds to the Principal Component Space (PCS), where we statistically model and classify each station's unique "fingerprint". The PCA is often referred to as Empirical Orthogonal Functions (EOFs) and is applied to understand various coastal geomorphological phenomena (Conlin et al., 2020; Miller & Dean, 2007).

More sophisticated algorithms, such as the DMD algorithm tries to find the best linear operator (A) to advance the data matrix in time, allowing a linear approximation as $X_P{}' = A \cdot X_P$ (Schmid, 2010). DMD goes further from PCA, in the sense that the variability of each DMD spatial mode (denoted with φ as well) is not orthonormal anymore, reproducing only specific frequencies that oscillate in a sinusoidal manner like the Fourier transform. This representation works as an advantage in physical systems where the modes represent specific periodic behavior with a typical growth rate. Nevertheless, PCA's temporal modes are more parsimonious, which is advantageous in applications like Principal Spectral Components (PSC) and Reduced-Order Models (ROMs). Thus, we use DMD to identify periodic behavior and the complete PCA to classify and describe fundamental patterns in the data.

In Chapter 2, we use φ_r from the DMD algorithm to obtain a new basis set that optimizes the spatial variability in the reduced-order modeling framework. We discover the nonlinear terms using sparse regression over a library of candidate terms and obtain a predictive ROM by approximating a linear behavior of the fundamental dynamics using Koopman theory and Galerkin projections.

1.5.2. *Classification and clustering*

To classify the spatial distribution of P in terms of the PCS, we applied the Naïve Bayes classifier (Langley, 1993; Langley & Sage, 1994). This technique consists of calculating the possible output based on the input. In other words, it can compute the posterior probability $P(c\,|\,x)$ from a class predictor $P(c)$, the probability $P(x)$, and the likelihood $P(x\,|\,c)$. The most important aspect of the technique is that it does not consider relationships between features of a given class (Langley, 1993). In our case, it means there is no initial preference between the stations (in the PCS) regarding the morphodynamical regime they should belong to.

We then applied the k-means (or Lloyd) algorithm in the PCS to cluster our observations into a user-defined number of clusters (k) (Camus, Cofiño, et al., 2011; Camus, Mendez, et al., 2011; Duda et al., 1995). The algorithm outputs the centroid of each cluster that minimizes the distance between the observations and centroids (Brunton & Kutz, 2017). We implemented both algorithms in MATLAB® with the "k-means" function and the "fitcnb" function from the Machine Learning Toolbox. We use both learning algorithms to (1) corroborate the physical intuition of previous morphodynamical classifications and (2) provide statistical models to classify new data. The developed data architecture encompasses a previous morphological classification, supervised and unsupervised learning, and a new optimal coordinate system to represent the data. This architecture aims to elucidate the relation between coastal morphodynamical regimes and oceanographic regimes based on the fundamental behavior of P.

1.6. Methods: Wave-induced coastal morphodynamics

We quantified the alongshore sediment transport, Q_s (in units of kg s^{-1}), using the CERC equation (Komar & Holman, 1986) modified to deep-water wave properties (Ashton et al., 2001) as:

$$Q_s = K_1 \cdot \rho_s \cdot (1-p) \cdot H_s^{\frac{12}{5}} T_p^{\frac{1}{5}} \cdot \cos^{\frac{6}{5}}(\phi_0 - \theta) \cdot \sin(\phi_0 - \theta) \tag{5}$$

where

$$K_1 = 5.3 \times 10^{-6} \cdot K \left(\frac{1}{2n}\right)^{\frac{6}{5}} (\sqrt{g\gamma_b}/2\pi)^{1/5} \tag{6}$$

is an empirical constant based on $K = 0.46\rho g^{3/2}$, where sediment and water density are denoted by ρ_s and ρ (kg m^{-3}), respectively, the dry mass void fraction is p, g is the gravitational acceleration (m s^{-2}), γ_b is the ratio of breaking wave height and water depth ($\gamma_b = 0.78$), and n is the ratio of group velocity to phase velocity of the breaking waves (1 in shallow water), ϕ_0 is the deep-water wave approach angle, and θ is the local shoreline orientation. Both angles are azimuth, even though the metrics used are in terms of the relative incoming wave angle ($\phi_0 - \theta$).

Plan-view coastal change depends on wave-induced (direction controlled) sediment reworking along the coastline (Nienhuis et al., 2016). For example, high-angle wave instability in shoreline shapes results in naturally occurring coastal landforms such as flying spits and capes (Ashton et al., 2001). Waves approaching the shore from different angles over time contribute to Q_s, either to the left or right flank. Integrated over time, the relative contribution of each wave direction to the alongshore sediment transport is given by the wave energy probability density distribution as (Nienhuis et al., 2015 Suppl. Information):

$$E(\phi_0) = \frac{H_s^{\frac{12}{5}}(\phi_0) \cdot T_p^{\frac{1}{5}}(\phi_0)}{\sum_{\phi_0} H_s^{\frac{12}{5}}(\phi_0) \cdot T_p^{\frac{1}{5}}(\phi_0)} \tag{7}$$

Additionally, we calculated the inner and outer depth of closure (DoC$_i$ and DoC$_o$) according to Valiente et al. (2019) but defined by Hamon-Kerivel et al. (2020) as:

$$DoC_i = 2.28H_s - 68.5({H_s}^2 / g{T_p}^2) \tag{8}$$

$$DoC_o = (H_s - 0.3\sigma_H)T_p(g / 5000D)^{0.5} \tag{9}$$

where g is the acceleration due to gravity, σ_H is the standard deviation of the significant height, and D the sand diameter; to find the seaward limit of wave-induced morphological change at mesoscale for each coastal station wave climate and geologically controlled continental shelves. Since WW3 produces an offshore regime of wave parameters, we compared and correlated each station's maximum width of the continental shelf with the DoC to evaluate the sediment accommodation availability of the shoreface at each station. We aim to evaluate the significant sediment exchange in high energy wave patterns associated with morphodynamical regimes.

1.7. Methods: Climate-driven periodic variations

The time series analysis consists of a Butterworth filter to obtain seasonal to decadal behavior of interest. The signal processing technique is also referred to as a maximally flat magnitude filter introduced by Stephen Butterworth (1930). We compare ordinary (or raw) data with the filtered time series to identify noise and frequencies of interest.

Since the teleconnection indices (i.e., ONI for ENSO) represent anomalies in variables of interest, we first normalized P to represent anomalies around a central statistic by applying a modified detrending process (Vega et al., 2020). We refer to these data as the Monthly Mean Normalization (MMN) time series. We calculated the mean energy flux value of every month for a decade and then subtracted it from every same month. After removing the monthly mean, we applied a normalization to represent the variability with zero means. We then calculated the wavelet coherence and correlation between the Anomalies, PCA modes, ordinary time series, and the Oceanic Niño Index (ONI), to analyze the atmosphere-ocean coupling driving wave energy variability.

Wavelet transform (Torrence & Compo, 1998), or multiresolution spectral analysis, is the most robust technique to obtain the frequency content present in a signal (Kumar & Foufoula-Georgiou, 1997; Vega et al., 2020). It has significantly changed the way data science deals with compressing sensing and representing time series in the digital era (Brunton et al., 2020). Like the Fourier transform and the PCA, the wavelet basis is an orthogonal decomposition that includes a multiresolution graphical representation of the signal, assuming that large frequencies do not require time resolution. We used the wavelet transform as the primary spectral analysis technique and the Fourier transform as a corroboration of the results obtained in the wavelet analysis. We used the MATLAB® implementation of the wavelet analysis provided by Grinsted et al. (2004) and Torrence and Compo (1998), which includes a Monte Carlo test for statistical significance. We focus our results on the wavelet coherence between the ONI and the ordinary variability of P, specifically on the quantification of the phase-lag (or time-lag). We use the arrow convention of 0° in the eastern direction. Each frequency resolution of the wavelet coherence (y-axis in Figure 8) would be equal to 2π, to compute the time-lag from the phase-lag.

In our case, the natural frequency of our physical system, according to the number of data elements (87,600) and the sampling period (3 h), is $f_0 = 0.033\ cpy$, meaning we can resolve a maximum frequency of ~3 cycles per century (1 cycle every ~30 years) according to the Shannon-Nyquist sampling theorem (Thomson & Emery, 2014). Similarly, the Nyquist frequency is $f_N = 4\ cpd$. Since we are interested in periodic variations related to ENSO, we focused on the amount of energy within 2-to-4-year periods for the warm El Niño phase and 4-to-8-year periods for the cold La Niña phase. This way, we identify specific events of intense ENSO phases related to P dynamics along the Pacific coast of the Northern Andes.

1.8. Results and discussion: Statistical trends and correlations

The decadal average inner and outer depth of closure or DoC (DoCi and DoCo, respectively in Table 1) are computed using an average sand diameter of 0.001 meters, and hourly time series of H_s and T_p for 30 years. We observe that latitudinal changes explain 83% and 85% of the variability in DoCi and DoCo, respectively. Assuming the simulated wave parameters reach the shoreface with similar values, the upper shoreface starts around 2 meters in depth and ends around 10-to-20 meters in the continental shelf. Comparing these results with the morphological parameters in Table 1, we assume that the wave climate present at the inner shoreface is approximately the same as in the outer shoreface Meaning wave hydrodynamics can be considered in our morphodynamic implementation.

Table *1* presents wave and geomorphological parameters from the 13 stations along the Northern Andes Pacific coast. Observations of these statistical parameters indicate that H_s, T_p increases with latitude (from 0.8 to 1.2 m and 11.1 to 14.6 s, respectively), suggesting a S-N rise in wave energy. A strong negative correlation between the continental shelf width and average P ($r_{Pearson} = -0.74$, $r_{Spearman} = -0.86$ and $RMSE = 24.07$) supports the idea that shoreface morphology, largely controlled by the compressional geologic setting (Correa

& Morton, 2010), exerts control on P variability, likely by bottom friction and dissipation of swell (Ardhuin et al., 2003).

The decadal average inner and outer depth of closure or DoC (DoC_i and DoC_o, respectively in Table 1) are computed using an average sand diameter of 0.001 meters, and hourly time series of H_s and T_p for 30 years. We observe that latitudinal changes explain 83% and 85% of the variability in DoC_i and DoC_o, respectively. Assuming the simulated wave parameters reach the shoreface with similar values, the upper shoreface starts around 2 meters in depth and ends around 10-to-20 meters in the continental shelf. Comparing these results with the morphological parameters in Table 1, we assume that the wave climate present at the inner shoreface is approximately the same as in the outer shoreface Meaning wave hydrodynamics can be considered in our morphodynamic implementation.

Table 1. Wave and geomorphological parameters used in the present study, shown as latitudinal changes with a spatial resolution of 0.5°N x 0.5°W. Data include the average significant height (H_s), average peak period (T_p), depth at buoys location, average wave energy flux (P), variability coefficient, and maximum width of the continental shelf.

Station	*Coord. (lat, lon)*	*Mean H_s [m]*	*Mean T_p [s]*	*Depth [m]*	*Mean P [W/m]*	*COV [%]*	*Shelf width [km]*	*DoC_i [m]*	*DoC_o [m]*
1	(1.5°N, -79°W)	0.7	11.1	223	0.46E+4	62	96	1.38	8.42
2	(2°N, -79°W)	0.9	12.2	494	0.92E+4	57	87	1.89	13.09
3	(2.5°N, -78.5°W)	0.9	10.6	80	0.62 E+4	62	67	1.57	9.22
4	(3°N, -78°W)	0.7	10.8	127	0.57E+4	64	116	1.54	9.06
5	(3.5°N, -77.5°W)	0.7	10.2	62	0.45E+4	63	67	1.38	7.53
6	(4°N, -77.5°W)	0.8	12.2	267	0.68E+4	60	47	1.64	10.98
7	(4.5°N, -77.5°W)	0.8	13.1	618	0.82E+4	60	40	1.73	12.60
8	(5°N, -77.5°W)	0.9	13.8	924	1.07E+4	64	31	1.91	14.94

9	(5.5°N, -77.5°W)	0.9	14.0	1193	1.13E+4	64	15	1.95	15.55
10	(6°N, -77.5°W)	0.9	14.1	1759	1.05E+4	64	21	1.87	14.94
11	(6.5°N, -77.5°W)	0.9	14.4	651	1.29E+4	68	29	2.04	16.92
12	(7°N, -78°W)	1.1	14.5	2487	1.89E+4	72	19	2.45	20.87
13	(7°N, -78.5°W)	1.2	14.6	1374	2.03E+4	70	10	2.55	21.81

From directional histograms, also known as wave roses of P (Figure 2A), we observe the angular distributions of incoming waves from different directions (between $\sim 240° N$ and $\sim 300° N$) at lower latitudes (from station 1 to 5). In contrast, at higher latitudes or Northern stations (stations 11 to 13), most of the energy arrives at a similar incoming angle ($\sim 220° N$). Interestingly, middle latitudes (stations 6 to 10) present a steady transition between both regimes (Northern and Southern stations) showing different morphologies. Suggesting the need to model the transition (Middle stations) to better understand the wave dissipation mechanisms associated with sediment fluxes, geologic characteristics, shoreface bedforms, and wave climate metrics. From exploratory results and statistical models applied on wave and geomorphologic parameters, we observe a clear trend of negative correlation between continental shelf width and wave energy flux. Suggesting that long-term coastal variability related to mesoscale morphodynamics appears to be mostly driven by swell energy, while local behavior requires further measurements to accurately assess its variability and forcing factors. Consequently, we focus on modeling hidden and leading-order variables in swell energy driving mesoscale changes in the continental shelf's morphology, to further evaluate long-term wave-induced coastal evolution.

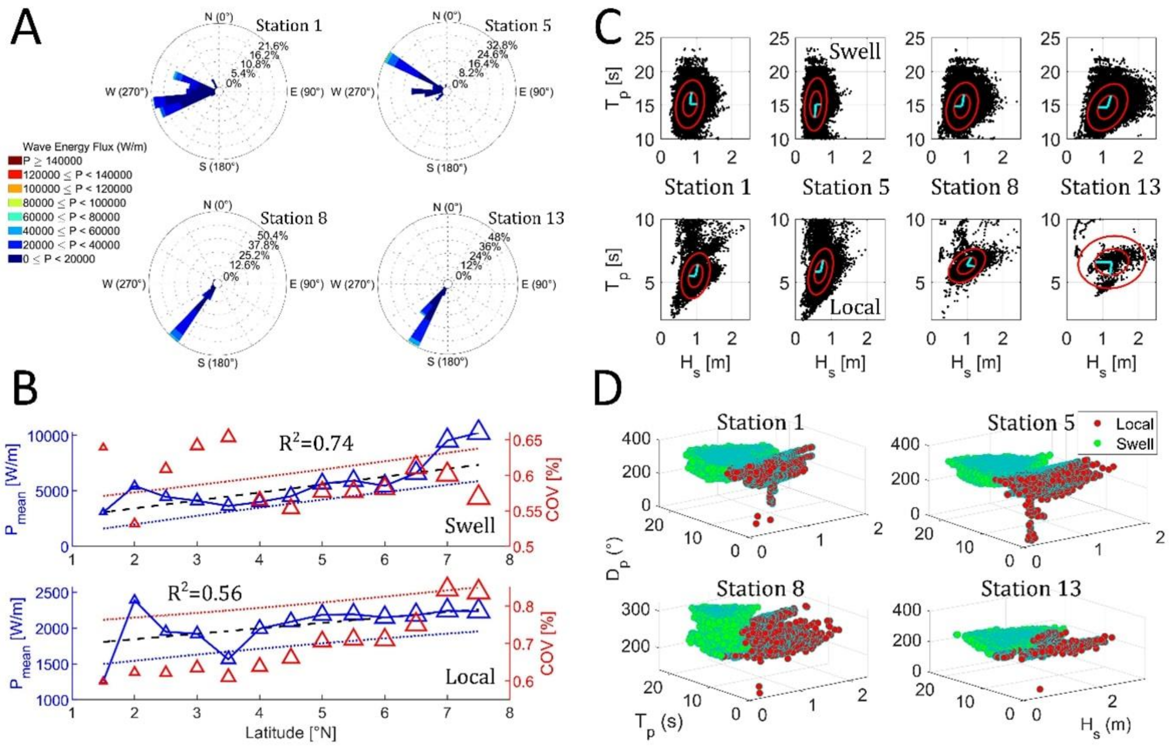

Figure 2. Exploratory analysis on wave climate reanalysis data at the Northern Andes Pacific coast. (A) Directional histograms of P angular distribution in stations 1, 5, 9, and 13 showing directional changes of the incoming waves. (B) Latitudinal changes of statistical parameters (mean and variability coefficient) from swell and local P, with a linear regression model for both wave environments (R^2=0.74 for swell and R^2=0.56 for local waves). (C) $H_s - T_p$ scatter diagrams for wave environment characterization at different stations along the coast, including the main directions of variance from the PCA (cyan lines). (D) $H_s - T_p - D_p$ diagrams highlighting swell and local wave evolution trends.

We classified swell and local waves in terms of T_p (local waves: $0 < T_p < 10s$ and swell: $10 < T_p < 20s$), as suggested by Holthuijsen (2007) in Chapter 3 Section 1.3, to observe the general dissipation mechanics of both short-wave types as latitude increases (Figure 2B). We find that swell carries most of the energy in the 30-year evolution of P along the Pacific coast of the Northern Andes (almost one order of magnitude more than local waves). We also corroborate that morphological changes of the continental shelf do not influence local wave evolution nearly as much as swell. As shown before, swell energy appears to dissipate in three distinct morphodynamical regimes: Stations 1 to 5 (Southern regime), stations 6 to 9 (Middle regime), and stations 10 to 13 (Northern regime).

Swell arrives mainly undisturbed to the nearshore (Portilla et al., 2015). Thus, the dissipation (and consequent reduction of swell energy) increases over the transition from narrow continental shelves at higher latitudes to larger widths at lower latitudes. In the Northern Pacific basin, the Galapagos Island serves as an energy dissipation landform that produces wave diffraction, creating interference of new wave fronts with the undisturbed swell (Dimar, 2020 Chapter IV). Bragg scattering also plays a fundamental role since it is included as a source term in WW3, accounting for resonant triad interactions with the bottom component that produces energy exchanges between waves with similar radian frequency (Ardhuin et al., 2003). We argue that both mechanisms contribute to the observed latitudinal reduction in swell energy. As latitude decreases, larger widths of the continental shelf dissipate swell energy, and the temporal variability increases to almost 65%. Swell energy dissipation by bottom friction occurs more significantly over large continental shelves in the Southern and Middle regimes. From these exploratory analyses, we observe the expected behavior of morphologically driven wave energy dissipation. However, swell, and local waves seem to be reduced substantially around Buenaventura Bay (station 5), possibly due to destructive phase interference caused by the Galapagos island's dissipative effects.

Table 2. Analysis of variance (ANOVA) applied to the linear regression model of swell and local energy flux latitudinal variability. The statistical parameters used to assess the analysis of variance and obtain the goodness of fit are the coefficient of determination (R^2), the root mean squared error, the p-value, t stat value and Fisher test value, as well as the correlation coefficient (r), the model's equation ($y = f(x)$) and the degrees of freedom.

Statistics/Variables	*Swell Energy vs Latitude*	*Local Energy vs Latitude*
R^2	0.74	0.56
RMSE	1.296e+03	2.648e+02
SST	37204568	368063.4
p-value	0.1352	3.75e-06
t stat slope	1.61	8.47
F-fisher	22.1356	5.2481
Pearson coefficient (r)	0.81	0.71
Regression equation	$y = 1510.17 + 904.258\,x$	$y = 1621.69 + 89.9406\,x$
Degrees of freedom	12	12

We further applied a linear regression regularizing least squares to model P_{mean} in terms of latitudinal changes (Figure 2B and Table 2). We assessed both exploratory and model evaluation analyses using the goodness of fit metrics between simulated and measurement results. The statistical model shows that latitudinal changes explain 56% of local and 74% of swell variability. From the ANOVA applied to the linear regression in (The decadal average inner and outer depth of closure or DoC (DoC_i and DoC_o, respectively in Table 1) are computed using an average sand diameter of 0.001 meters, and hourly time series of H_s and T_p for 30 years. We observe that latitudinal changes explain 83% and 85% of the variability in DoC_i and DoC_o, respectively. Assuming the simulated wave parameters reach the shoreface with similar values, the upper shoreface starts around 2 meters in depth and ends around 10-to-20 meters in the continental shelf. Comparing these results with the morphological parameters in Table 1, we assume that the wave climate present at the inner shoreface is approximately the same as in the outer shoreface Meaning wave hydrodynamics can be considered in our morphodynamic implementation.

Table *1*), we can obtain statistical significance based on the null hypothesis that both distributions share a constant slope. In swell energy, the regression model presents a value of 1.61 that satisfies a 95% significance confidence for the slope for 12 degrees of freedom. The p-value of 0.13 and F-fisher value of 22.13 also support the null hypothesis and decision rule that the model explains 74% of the variability. In other words, since there is a significant linear relationship between the variables, we can statistically model the overall spatial variability trend of P using latitudinal changes (Figure 2B) in the Northern Andes Pacific coast.

From the wave roses of P and the $H_s - T_p - D_p$ diagrams (Figure 2D), we could argue that the local wave energy diminishes as latitude increases. Rather, the latitudinal changes of statistical parameters in Figure 2C show that local wave energy moves around the same values along the continental shelf. Therefore, the decreasing number of observations appears from either an artifact of the model time-stepping scheme or the distributions only show a more substantial presence of undisturbed swell arriving at higher latitudes (Northern regime). All data visualization results suggest considerable local wave energy at lower latitudes, decreasing as swell energy increase at higher latitudes.

1.9. Results and discussion: Morphodynamic classification of coastal regimes

Since our goal is not to reconstruct the specific temporal and morphodynamical behavior, but rather to classify the different patterns in the evolution of the morphodynamical system, we apply a dimensionality reduction to adequately assess the hierarchical patterns in the incoming wave energy variability. Figure 3 shows the obtained dimensionality reduction of the 13 stations along the coastline associated with each morphodynamical regime highlighted. The first three PCs show predominance over the other singular values from the hierarchical decomposition (Figure 3B). Explaining 95% of the total variability, according to the cumulative percent of the variance of the singular values.

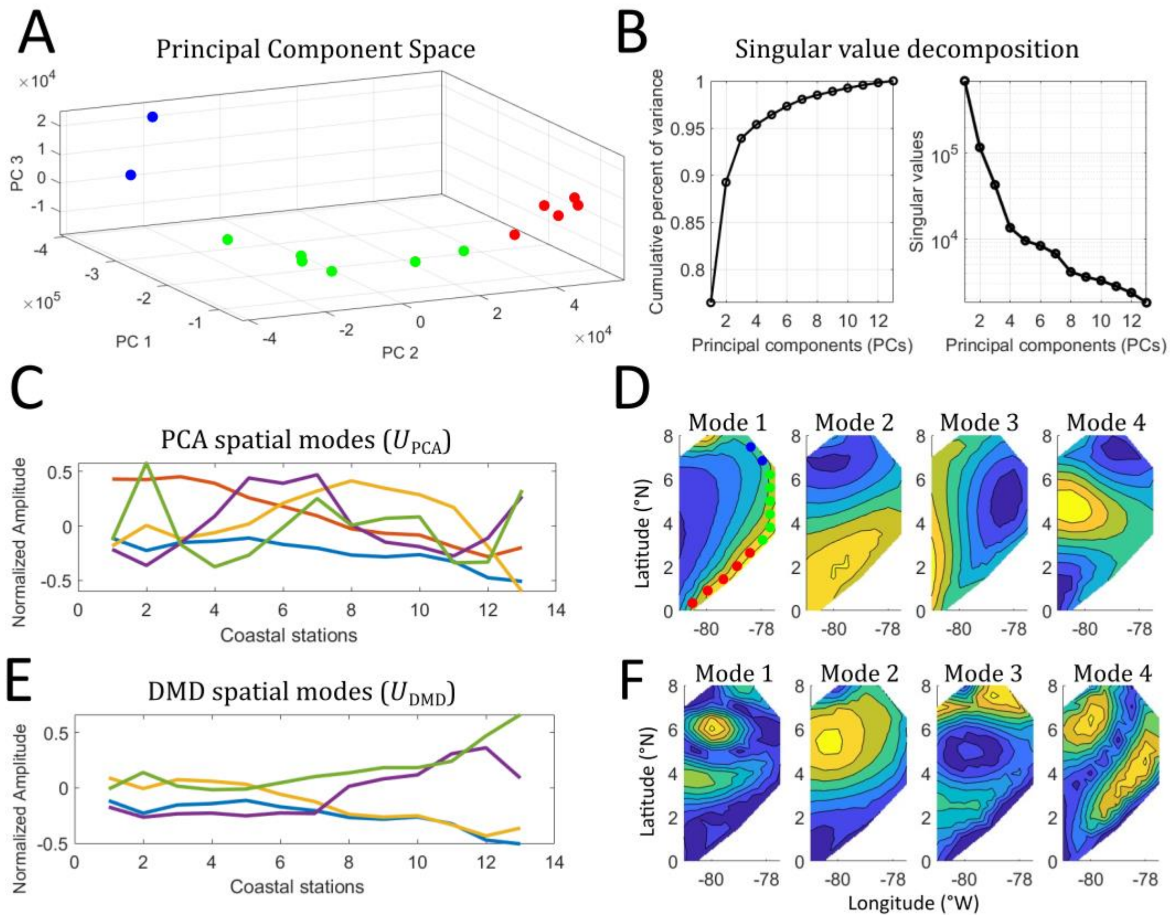

Figure 3 Data-driven modal decomposition of wave climate data at the Northern Andes Pacific basin and coastline. The singular value decomposition (SVD) from both PCA and DMD algorithms produces a dimensionality reduction space (A) to express the fundamental variability trends in the data. From the singular values (B) we observe that most of the energy is captured by the first four PC, explaining almost 95% of the

total variability. PCA and DMD modes (C) represent dynamic patterns in spatio-temporal variability of P in the basin's seafloor topography (D), showing fundamental variabilities related to physical processes.

This result indicates that we can effectively reduce the dynamical system from 13 partial differential equations (PDEs) to 3 ordinary differential equations (ODEs) and represent 95% of the total variability. Thus, the model composed of the first modes represents fundamental patterns or hidden variables in the system. Here, we use this as an advantage to classify the coastal stations in the PCS (Figure 3A) to compute statistical and machine learning models for the morphodynamical regimes.

We computed the first four spatial modes (both PCA and DMD) for the whole Pacific basin of the Northern Andes and coast, as illustrated in Figure 3C and E, and Figure 3D and F. We found similar spatial variability between PCA mode 1 and DMD mode 2, PCA mode 4 and DMD mode 2, and PCA mode 2 with DMD mode 4. The critical aspect of these similarities is the differences between the dimensionality reduction algorithms and their associations to the temporal modes. DMD spatial modes represent better the physical processes (Tu et al., 2014), but its temporal information can be correlated and represented as a spectrum of defined frequency values. Therefore, since the temporal modes of the PCA are not correlated (orthonormal), and the spatial DMD modes represent better physical processes, we use the temporal modes of the SVD associated with the spatial modes of the PCA and DMD to compute the PCS and the Reduced-Order Model (ROM).

We observed the incoming swell of the basin reaching the coast in PCA mode 1. The PCA mode 2 appears to represent a latitudinal bimodal variability that could be related to the Intertropical Convergence Zone (ITCZ). The PCA mode 3 could be related to the energy from the reflected waves on the continental shelf. On the other hand, the DMD mode 1 suggests the effects of a source/sink of variability around the longitudinal trenches found near 5°N. The Malpelo island and ridge could also influence this behavior in the DMD modes (modes 1 and 2) as a dissipative source of the basin. All modes appear to represent wave dynamics related to the continental shelf morphology.

From the PCS, we generated three modeling approaches for classifications of the morphodynamical coastal regimes: empirical (Figure 3A), supervised, and unsupervised (Figure 4A and B, respectively). From previous characterizations of wave-induced morphodynamics along the Pacific coast (Correa, 1996; Correa & Morton, 2010) we propose three regimes, i.e., Southern, Middle, and Northern, to label the data in the supervised classification by the Naïve Bayes algorithm. We observe that the Middle regime presents the largest statistical variance centroid, which we interpret as representing the boundary between Southern and Northern conditions. To obtain more statistical significance around the computed centroids, we also applied the unsupervised learning algorithm k-means clustering. This way, we regularized around three clusters -or regimes- with $k_c = 3$ to then confront the previous classification with our computational clustering. We observe similar behaviors, interpreted as robust corroboration of the proposed regimes.

The final classification considers both centroids to compute the statistical model. The main difference between the supervised and unsupervised techniques is that the Southern regime (red cluster) should include stations 6 and 7, enlarging the boundary evolution of the Middle regime (green cluster) away from the Southern regime (red cluster). This result confirms the statistical significance of the pioneer geologic and morphodynamic characterization of wave-driven coastal evolution (Correa & Morton, 2010). It does so by grouping the stations into morphodynamical regimes obtained from the spatio-temporal variability of P.

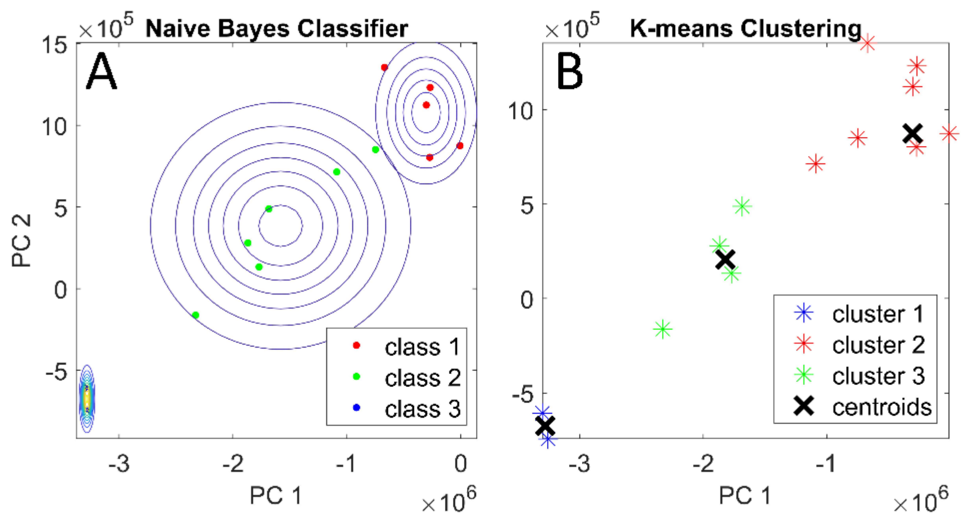

Figure 4. (A) Supervised and (B) unsupervised learning algorithms applied to the PCS of coastal P. The Naïve Bayes Classifier uses the previous expert classification of empirical coastal regimes provided as labels to obtain statistical significance on three distinct regimes. The K-means clustering algorithm provides a data-driven classification of each station based on regularizing over the input of clusters (k=3) and returning the best fit class of each station. This approach aims to corroborate the morphodynamic classification.

From Figure 5A, we observe patterns in the PCS for the whole Pacific basin of the Northern Andes. The fundamental pattern emerging in the PCS illustrates trajectories in the basin represented as spatio-temporal coherent structures, which we can model to observe dissipation trends and wave evolution across several stations. Using both a multivariate linear regression or Sparse Identification of Non-linear Dynamics - SINDy, by Brunton et al. (2016), we can obtain a linear model and partial differential equations to predict the dynamic trajectory of the incoming P within the basin. This approach would produce an interpolation model to obtain wave-induced morphodynamics from new time series data.

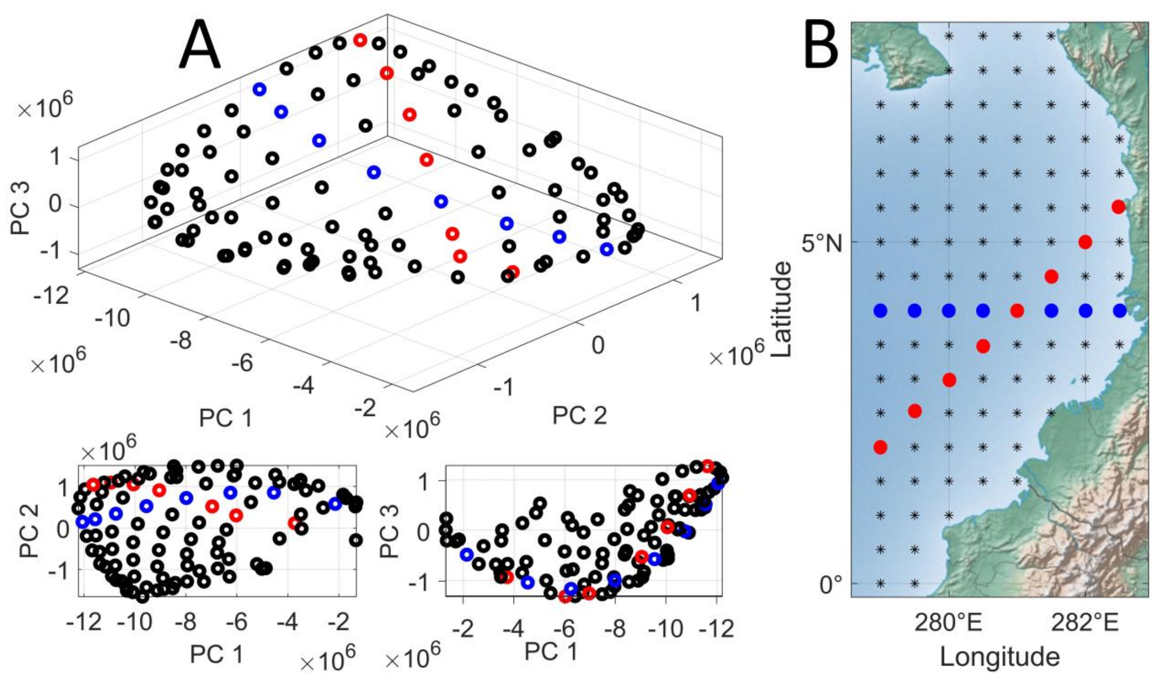

Figure 5. Principal Component Space of 101 stations of wave energy flux (P) variability along the Pacific oceanic basin of the Northern Andes. We highlight the longitudinal variability in blue, and the average incoming wave angle ($\phi_0 \sim 300°$ Azimuth) in red. Panel A presents all 101 locations on the Pacific basin as a structural dynamic pattern of the system, including the highlighted trajectories. The PCS conveys a well-defined pattern representing the fundamental variabilities explained by the first three PCs (90% of the total variability). In this optimal, reduced-order coordinate system, we can model and classify P changes along the basin and observe the dissipation behavior along the continental shelf

Assuming that the fundamental patterns in P dynamics dictate coastal gradients in alongshore sediment transport, we classify the low-dimensional representation (PCS) into the three, previously mentioned morphodynamical regimes. We argue that this classification would serve as a basis for assessing future wave-induced coastal evolution scenarios. Figure 6 presents the proposed association of morphodynamical classification, sediment transport, and wave-climate metrics of coastal features to wave climate. The satellite images of the Pacific coast of the Northern Andes are free to access by the USGS Earth Explorer platform (https://earthexplorer.usgs.gov/).

We recall that the Northern morphology, on the western flanks of the Baudó Range (Serranía del Baudó), is composed predominantly of oceanic basalts, diabase, and associated cherts and radiolarites. Also, south of the Baudó Range, the relief of the Pacific Coast exhibits 20–100 m high hills cut into Tertiary sedimentary sequences (Correa & Morton, 2010). Our results on wave-induced morphodynamics along the Northern regime show high sediment transport values and stable coastal response. However, we highlight that the geological setting and sediment characteristics, not considered by the applied metrics, would likely overestimate sediment transport for the Northern regime. In contrast, coastal relief on the Central regime includes the three major Plio-Quaternary deltaic prisms of the San Juan, Patía, and Mira rivers. Which suggests a more accurate representation of sediment transport and shoreline instability shape formation on the Middle and Southern regimes.

Extensive, highly unstable, sandy barrier islands and muddy tidal flats are common at all the main tidal inlets of the Pacific Coast (Morton et al., 2000). All the barrier islands in the Middle and Southern Pacific coast have experienced critical morphological changes in the last century that implied erosional and accretional events at specific locations (Restrepo & López, 2008). For example, at the San Juan delta (Figure 6, lower part of the Middle regime), significant morphological changes have occurred during the last 30-year period, including shoreline retreat, barrier islands narrowing, and breaching (Restrepo & Cantera, 2013; Restrepo & López, 2008). Previous evidence indicates periodic coastal features as beach ridge systems along the Pacific coast. We corroborate these observations with the instability of wave climate computed for the Southern and early Middle regimes (Figure 6). We observe that 180° coasts with 130° to 180° incoming wave climate produces high instability metrics associated with periodic morphological changes. This is the case for most of the central coast of the Northern Andes South American basin.

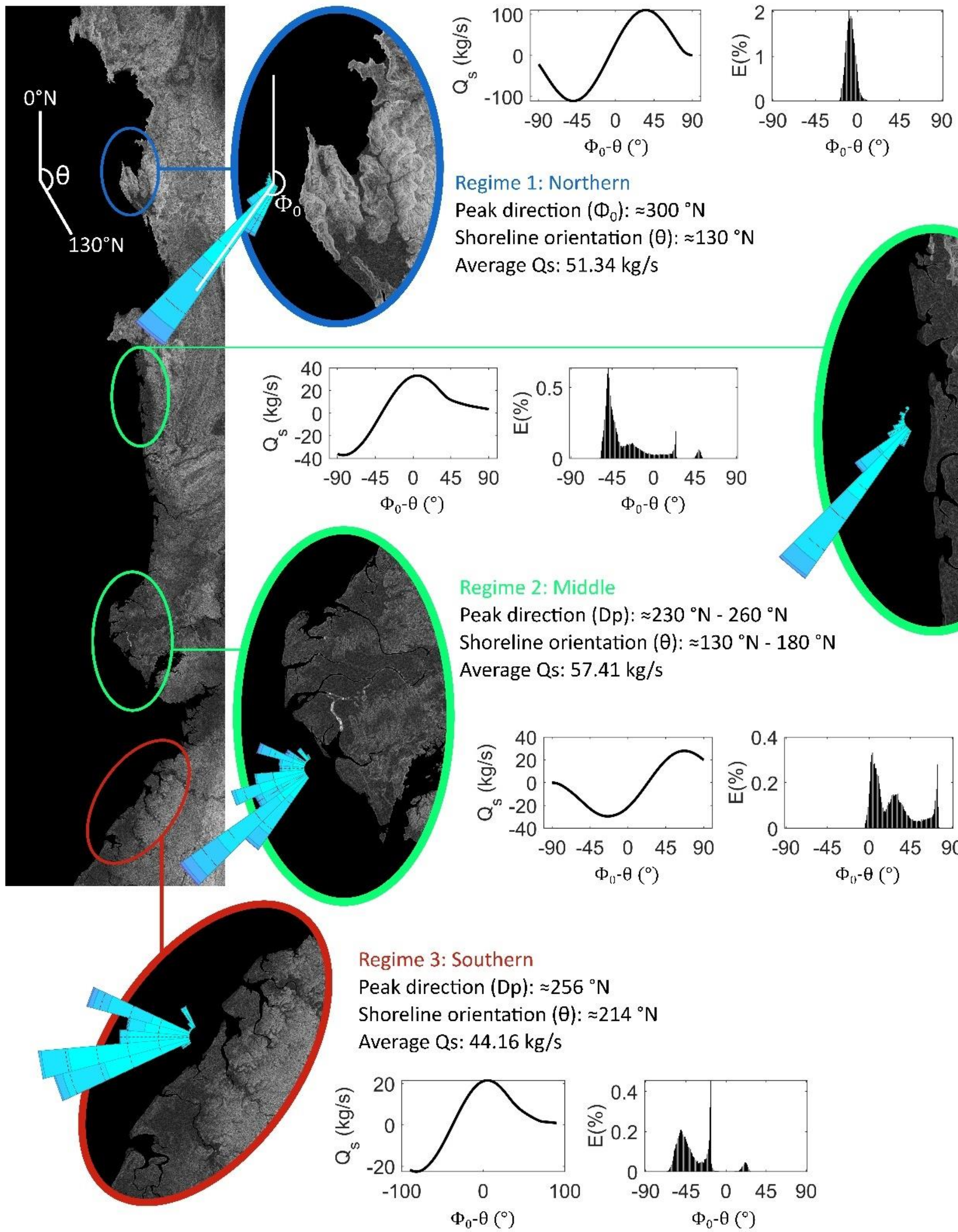

Figure 6. Morphodynamical classification of coastal regimes using wave climate metrics of alongshore sediment transport (Q_s) and directional energy (E) versus relative incoming wave angle (Φ_0-θ). In colors, we represent the distinct regimes classified to relate specific morphodynamics conditions. Stable morphodynamics (Northern regime in blue), the transition between stable and unstable morphodynamics (Middle regime in green), and unstable morphodynamics (Southern regime in red).

From wave climate metrics, we show that Colombia's Pacific coast presents instabilities on the Southern and early Middle regimes. The specific shoreline orientation of the Pacific coast of South America ($\theta \geq 180$) and the high breaking wave angles ($\Phi_0 - \theta \geq \pm 45$) produce gradients in alongshore sediment transport that result in unstable geomorphological features, such as spits, dynamic barrier islands and swamps, inside very complex littoral cells. We find that this unstable behavior slowly decreases as latitude increases, where the shoreface is controlled by the geological setting.

Additionally, we observe that the compressional margin that produces a deep trench with a small shoreface also results in a small transitional zone where sediment accommodation is generally limited, suggesting that seasonal variations in wave climate and storm events likely interact with the small surf zone or upper shoreface sediment to produce the resultant local trends in the morphodynamical regimes. While decadal variations of P-driven sediment fluxes are mostly produced by the tectonic setting of the compressional margin and its deep trench formations, explaining the resultant swell dissipation trends and latitudinal variability.

1.10. Results and discussion: Climate-driven variability related to the ENSO

To adequately assess the temporal variability of P, we apply spectral, data-driven, and time-series techniques to identify the periodic behavior related to climate forcing. Figure 7 shows significant coherence in seasonal and intra-annual variability, as well as a strong La Niña phase around 1997. The computed average phase lag of $\pi/2$ (or time lag of 8 months) between both time series at around 32 months suggests that P responds to strong ENSO events, as expected (Caicedo-Laurido et al., 2019). Furthermore, wavelet results show that the phase-lag varies in time, exhibiting different behaviors for both ENSO events. La Niña evidences an average phase-lag of $\sim 90°$ at ~ 3 $-$year periods, corresponding to a time-lag of 9 months. In contrast, for El Niño events, the phase varied from $\sim 90°$ to $\sim 270°$, corresponding to time-lags of 4 and 12 months for ~ 16 $-$month periods.

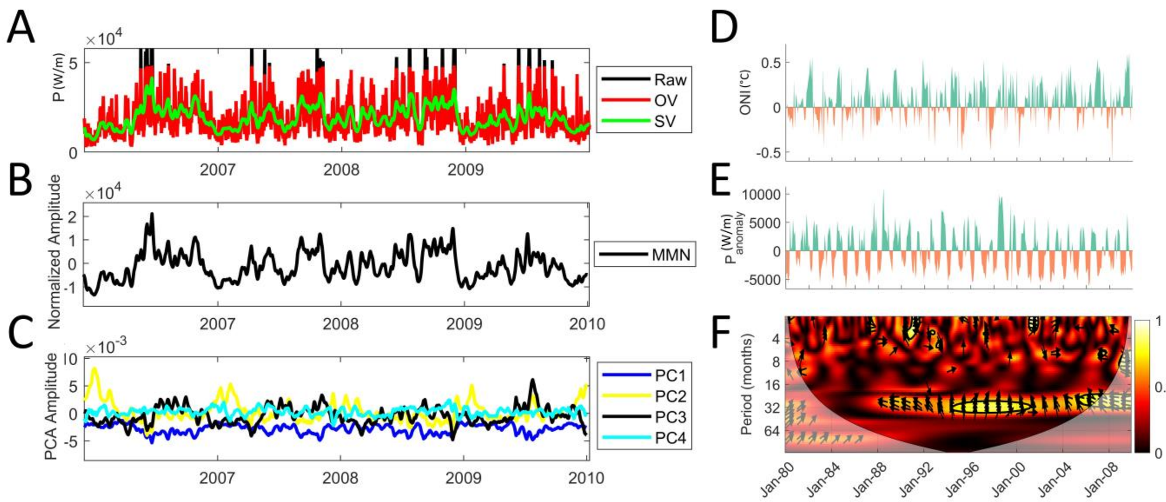

Figure 7. Time series of *P*, ONI, and wavelet coherence between them. (A) Temporal evolution of filtered variability of *P* (Ordinary Variability: OV and Seasonal Variability: SV), anomalies from monthly mean normalization (MMN), and the first four PCA modes from 2007 to 2010. (B) Wavelet coherence between ordinary variability of *P* and the ONI.

Figure 8 presents the 95% significant (segment-averaged) and ordinary Fourier spectrums (panel A and B, respectively). We computed the significant spectrum to obtain robust spectral trends, while the ordinary spectrum was computed to identify peaks of present periodic behavior in the complete interval of spectral resolution. We observe strong diurnal tidal behavior (360 cycles per year) related to tidal constituent k1, with decreasing presence of semi-diurnal tides (720 to 1080 cycles per year) related to tidal constituent m2 and n2. From the Averaged Fourier Spectrum (Figure 8A) we deduce that lower frequency (from 12 to 0.05 cycles per year) presents low resolution due to swell superposition over 30 years. Nevertheless, we observe a clear increase of spectral density in the low-frequency regime as latitude increases, showing less tidal presence (360, 720, and 1080 cycles per year) and stronger low frequencies at higher latitudes (Middle and Northern regimes).

We also computed the DMD spectrum (Figure 8C). We obtained significant frequencies in both spectrums (DMD and Fourier) of 1 and 2 cycles per year (seasonal variability). Fourier, Wavelet, and DMD results show a range of low frequencies from 0.5 to 0.1 cycles per year (2 to 10 years every cycle), meaning wave climate variability shows the presence of synoptic behavior likely related to ENSO phases and Madden-Julian oscillations. The overall spectral analysis performed using the SVD, Wavelet, and Fourier basis present similar behaviors, meaning the time series data was accurately pre-processed and processed, and the results are statistically significant.

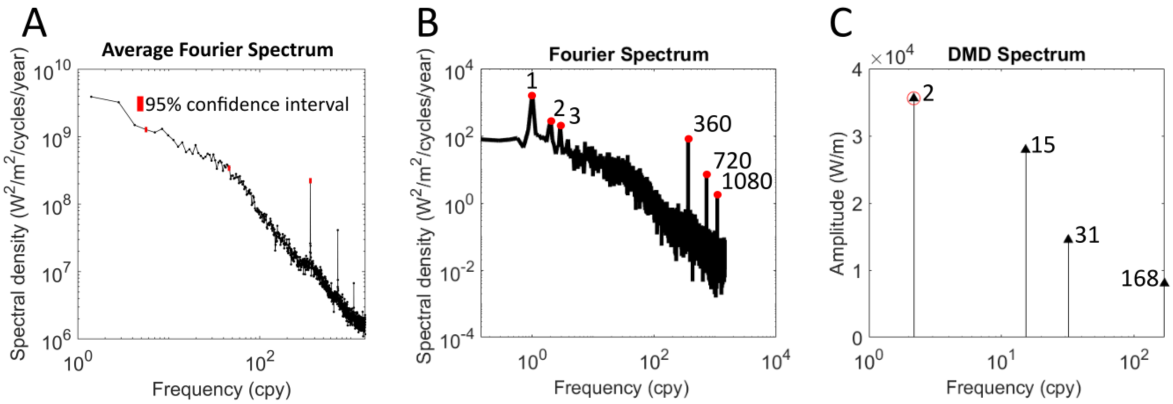

Figure 8. Spectral decompositions in the Fourier basis (A and B) and DMD basis (C). Fourier transform of *P* presents substantial intra-annual variability, with significant frequencies related to seasonal trends (1, 2, and 3 cycles per year) and tidal behavior (360). The DMD spectrum also presents similar results: monthly variability (15 and 31 cycles per year), and seasonal variability (2 cycles per year).

Figure 9 summarizes the Wavelet analysis performed on the latitudinal stations by showing the main differences in Wavelet transforms (column 1), coherence (column 2), and correlation (column 3) between stations 1 and 13, and ONI time series, representing each extreme regime (Southern and Northern, A and B respectively). We also present the Wavelet analysis performed on the PCA temporal modes (Figure 10) to relate and identify the fundamental dynamics of P related to climate-driven periodic behavior driven by the ENSO.

Wavelet transform (Figure 9 – Column 1), computed with a significant threshold of frequencies from the Monte Carlo test, supports the idea that the fundamental variability and spectral components in wave climate are driven by seasonal, and synoptic or mesoscale variability patterns. We identify substantial seasonal (4 to 6 months) and moderate La Niña-related periodicities (3 to 5 years) in the Southern regime. In comparison, we observe significant annual variability (12 months) and moderate El Niño-related periodicities (12 to 16 months) in the Northern regime. By following the Wavelet transform of the anomalies obtained from the MMN, we also identify significant La Niña anomaly events present in the Northern regime. Both regimes present similar average time lags (approximately 90°, or 8 months) in significant periods related to La Niña events, showing a more considerable Wavelet coherence (Figure 9– Column 2) at higher latitudes.

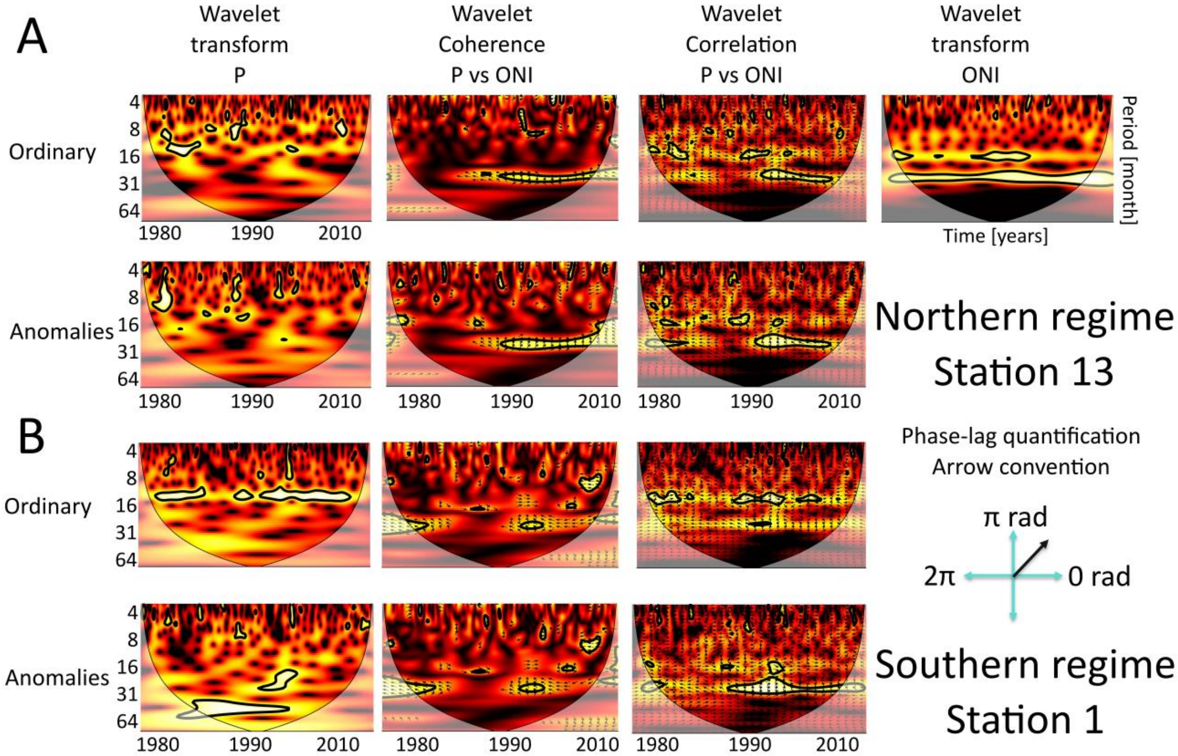

Figure 9. Table format to visualize wavelet transform, coherence, and correlation between time series of P and the ONI, with Monte Carlo statistical significance. Ordinary variability and Anomalies are compared between stations 1 (Panel A) and 13 (Panel B) to identify changes in response between lower and higher latitude wave energy.

From the Wavelet transform of the ordinary variability of P (first and third row), we observe seasonal periodicities from 6 to 8 months in both regimes. Nevertheless, station 1 shows a more significant and almost constant annual variability (around 12 months). This result suggests that local waves, which are more present in the Southern regime, appear to be driven mainly through annual and inter-annual variabilities (6-to-12-month periods), while swell seas are modulated mostly by synoptic periodicities (2-to-10-year periods).

Figure 10 shows the first five (5) PCA modes, illustrating the presence of similar periodicities compared to El Niño phase with a significant correlation between 12 to 16 months. PCA modes present coherence with El Niño phases, with a time lag of 24 months.

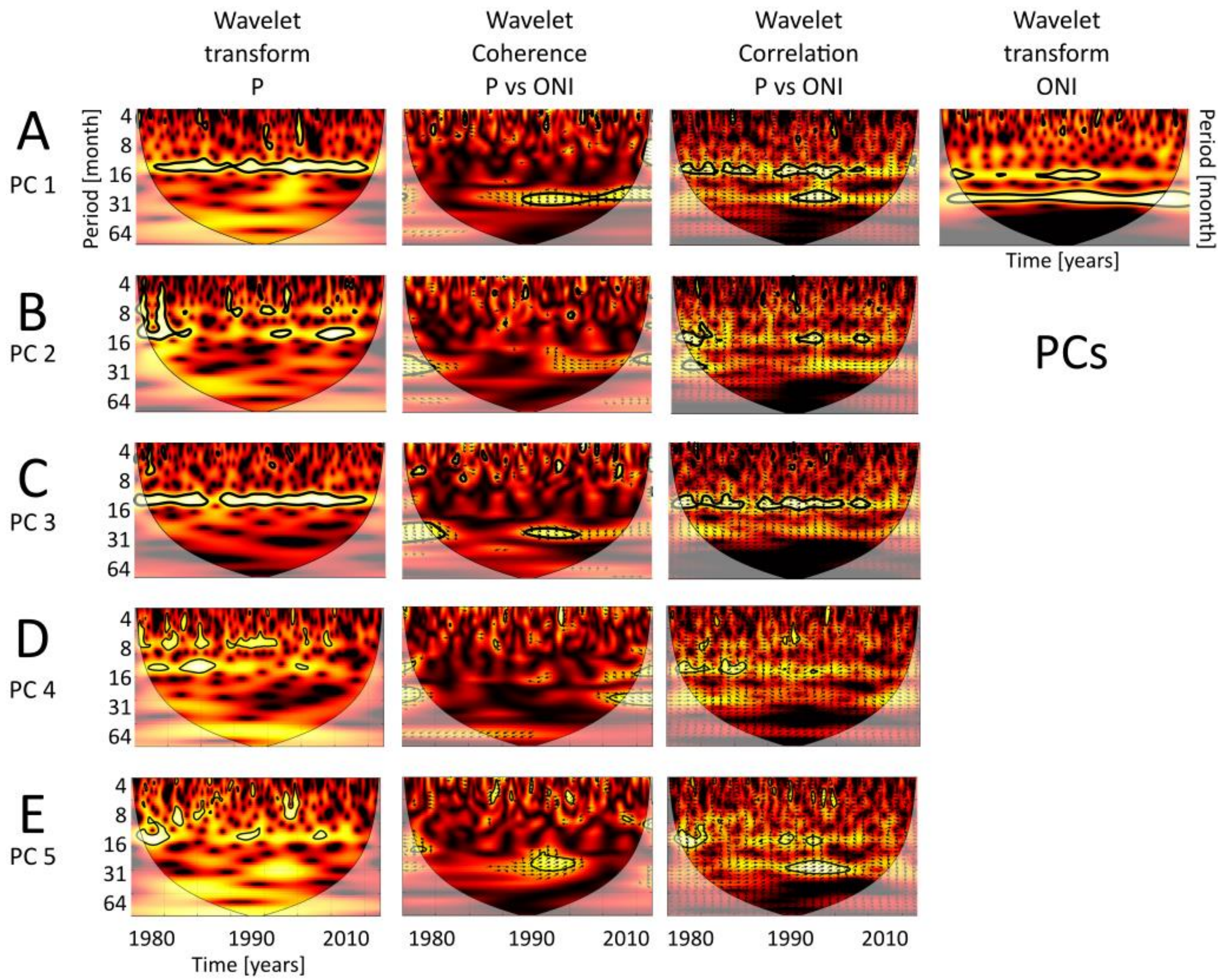

Figure 10. Table format to visualize wavelet transform, coherence, and correlation between temporal PCA modes of *P* and the ONI, with Monte Carlo statistical significance. Modal variability is compared between stations 1 and 13, showing no significant difference. Consequently, the temporal modes correlation and coherence with the ENSO index shows climate-driven behavior related to warm and cold phases, together with interannual feedback responses.

1.11. Conclusions

We find evidence that the spatial variations of *P* along the continental shelf of the Pacific Northern Andes show three apparent morphodynamical regimes, divided into Southern, Middle, and Northern latitudes. This result agrees with previous morphodynamical and hydroclimatic classifications (Correa & Morton, 2010; Dimar, 2020 Chapter I and IV), yet we propose the addition of the Middle regime to express and model the changing boundary using sediment transport metrics and shoreface evolution between the two distinct behaviors in the Northern and Southern regimes. We observe an almost linear trend of increasing sediment transport and average *P* as latitude increases, and we find that continental shelf width decreases linearly as well. The same behavior is found between latitudinal changes and depth of closure, with an increasing trend of DoC as latitude increases. This dissipation behavior appears mostly driven by the deep trenches along the compressive continental shelf, suggesting that the tectonic activity controls most synoptic and mesoscale wave energy variability in the Pacific coast of Northern South America. Likely illustrating a decreasing trend in wave-induced morphological effects on coastal evolution as latitude increases.

Since global reanalysis datasets like WW3 do not accurately represent nearshore wave environments, rather representing lower shoreface incoming waves corresponding to deep water waves from nonlinear wave-wave interactions. Therefore, the morphodynamical and wave climate metrics we apply to compute the sediment transport and high-angle wave instability for large-scale coastal systems fail to quantify accurately local processes. This allows a general morphodynamical relation, based on large-scale coastal feature and lower shoreface evolution, but fails to represent medium and short-scale wave-induced coastal morphodynamics. Nevertheless, our focus is to understand mesoscale variability trends along the continental shelf and the spatio-temporal, wave-induced coastal feature evolution scenarios at mesoscale in the Pacific basin and coast of the Northern Andes compressional margin.

On the data-driven analysis, physical processes can be observed in the DMD and PCA modal analysis, such as the dissipation of *P* by longitudinal trenches in the oceanic basin, and the ITCZ latitudinal variability. Both processes are equally identified as periodic behavior in spectral analysis. We observe stronger La Niña-related coherent periodicities within P around the Southern regime. While the Northern regime shows a stronger presence of El Niño-related periodicities. This hints that local wave energy might be more affected by La Niña phases in the ENSO, contrary to swell energy that seems to be more coherent with El Niño phases on the Pacific coast of the Northern Andes. The reported *P* response to ENSO climate variability may contribute to the development of early-warning systems of high energy wave events along the Pacific coast of Northern South America.

2. Chapter 2: Data-Driven Model from Reduced-Order Dynamics of Wave Energy Flux along Compressional Shelves

Abstract

We present a data-driven modeling architecture of wave climate variability along compressional continental shelves based on a dynamical system definition for the wave energy flux, P in W/m, with $\{P \in \mathbb{R} | P \geq 0\}$. We use wave energy flux data calculated from wave statistics, a reanalysis model of wave dynamics from discrete spatial locations and time realizations. We determined the functional form of the governing equation from sparse regression methods, numerical differentiation schemes, and model selection metrics to predict its reduced-order dynamics in time and space. Further, we used a dimensionality reduction architecture to train the algorithm and numerically solve a linear, low-rank, parsimonious representation of the nonlinear system of interest. We applied the reduced-order model, trained with data from 1980 to 2000 and several distributions of P along the continental shelf of the Northern Andes Pacific compressional margin. We forecast significant results to 2010 evaluated against high-fidelity simulations ($NRMSE = 15.14\%$). We finally apply our data-driven architecture to predict scenarios of alongshore distribution of P, corroborating the wave-induced morphodynamics regimes proposed by pioneering observations.

2.1. Introduction

The dynamical systems approach to study Earth systems provides a rigorous mathematical framework to describe and understand natural phenomena (Holmes, 2005 and references therein). Earth systems exhibit properties that limit their predictability and constrain their modeling, such as chaos and self-organization (Lorenz, E. N., 1963; Murray et al., 2009; Poincare, H., 1893, and many others). Thus, modeling efforts aim to unravel such constraints to provide useful information to society (Broomhead & King, 1986; Limber et al., 2017; Murray, 2007).

We propose a modeling architecture based on complex systems theory and a reduced-order framework (Brunton & Kutz, 2017; Sayama, 2015) to study and predict wind-wave energy flux variability over continental shelves. We focus on wave data with high spatial and temporal resolutions obtained from reanalysis studies (Chawla et al., 2011). Our framework deals with the system's complexity, nonlinearity, self-organization, high-dimensionality, and emergent properties by: (1) discovering the governing symbolic expressions, (2) selecting the ideal model for the defined system, and (3) solving it to specific conditions and boundaries. We apply our approach to wave data from the Northern Andes Pacific continental shelf.

2.2. Background and context

The PDE-FIND algorithm (also known as Sparse Identification of Nonlinear Dynamics – SINDy, (Rudy et al., 2017)) allows the discovery of nonlinear terms from time-series measurements of a dynamical system. This discovery is possible by applying a sparse regression on $Ax = b$, with a library of possible candidate terms as A and the temporal derivate as b (Brunton & Kutz, 2017). The discovery of a partial differential equation (PDE) describing the system also depends on the creativity to propose candidate terms, and the use of effective numerical differentiation techniques (Brunton et al., 2016). Symbolic regression, together with recent developments on graphic neural networks and genetic programming, also allows obtaining a simple mathematical representation of the data. PDE-FIND works better when assuming that the PDE is a sum of known functional forms. In contrast, symbolic regression with graph neural networks and genetic programming is better suited for nested expressions with rare and specific functional forms such as $e^{x/2}$ or $e^{-dx/dx}$ (Cranmer et al., 2020). Our implementation uses the PDE-FIND algorithm to discover the mathematical representation of the system assuming the nonlinearities are expressed as different functional forms of spatial derivates of the system's state (x).

Pioneer works proposed the Kullback–Leibler divergence (KL divergence) or relative entropy as a model selection approach. Comparing both simulated and actual probability distributions aid in evaluating how much the simulation explains the observations (Kullback & Leibler, 1951). This idea later evolved into the Akaike and later Bayes Information Criteria (BIC), including several models to find which one is more representative of the data (Akaike, 1974). Both techniques depend on the Pareto front, although only BIC assures a convergence to the best-fitted model (Dziak et al., 2020). Consequently, we use the Pareto analysis to evaluate the model, balancing accuracy (low error) and complexity (few terms and data elements). In essence, to contrast the different time series measurements used to train the model, we evaluate the errors and deviations to select a parsimonious description and optimal training data for our system of interest. This approach is possible by forecasting each PDE and trained model in a reduced-order framework.

The reduced-order modeling framework, or Galerkin projections, embeds the dynamics into an orthonormal basis set. This framework provides the construction of a new coordinate space that linearizes the fundamental variability of the system (Rapún & Vega, 2010; Rowley et al., 2004). The problem lies in finding the orthonormal basis that best fits the actual data. In this vein, the Singular Value Decomposition (SVD) is the fundamental dimensionality reduction technique, usually described as a data-driven generalization of the factor analysis such as the Fast Fourier Transform algorithm (Brunton, et al 2019). This technique extracts fundamental correlation patterns in the data, producing an ideal basis set to embed the dynamics. The SVD algorithm, applied as the Principal Component Analysis (PCA) and Dynamic Mode Decomposition (DMD), are some of the fundamental numerical matrix decomposition techniques in the computational era (Brunton & Kutz, 2017). For example, the DMD effectively offers a tool for extracting dynamic information from a sequence of uniformly sampled measurements in Earth systems (J. M. Zhang et al., 2020). The resulting modes represent the relevant variability structures that contribute most to the overall evolution captured in the measurement sequence (Schmid, 2010). The DMD theoretical framework depends on the Koopman operator that advances the system in time, allowing a linearization with a more straightforward solution (Brunton et al., 2016; Rudy et al., 2017). We apply the PDE-FIND algorithm in the model discovery, the Pareto analysis in the model selection, the PCA formalism in the reduction and simulation, and the goodness of fit in the model evaluation, defined as the main steps to assess the spatio-temporal variability of the system of interest.

2.3. Methods: Data-driven discovery of the dynamical system

We defined the dynamical system by analyzing the temporal variability (30 years) of wave energy flux (P) simulations at nearshore stations (13 virtual buoys) along the compressional continental shelf. We describe P computation in the previous chapter. The assumptions that apply to our implementation are: (1) the variable P represents the oceanographic conditions interacting with coastal morphodynamics, (2) we do not account for wave dissipation at depths shallower than one typical wavelength of $O(100\ m)$, (3) empirical parametrizations represent swell variability more accurately, especially on sandy coastlines, and (4) the model uses a heuristic formalism (source terms) to quantify wind input and nonlinear quartet interactions, which are still subject of scientific scrutiny.

We computed the measurement matrix of the system $X_P(i,n) \approx \bar{P}(x,t)$, assuming the measured subsystem (or sample) explains most of the defined system, containing a discrete collection of 87663 temporal realizations (n) of $\bar{P}$ at 13 coastal stations (i). We compute X_P, for simplicity referred to as X, as the matrix containing discrete measurements of $\bar{P}$. By definition, the reanalysis data do not represent actual measurements. Rather, they represent a parametrization of climate-driven oceanic phenomena. Thus, we do not have access to exact PDEs describing $\bar{P}$ system. With these premises, we first enforce a general description of the dynamical system based on a nonlinear PDE, such as:

$$\frac{\partial X}{\partial t} = X_t = N(X, X_x, X_{xx}, \dots, x, t, \beta) \tag{10}$$

where the subscripts represent partial differentiation in time "t" and space "x". We define the matrix $N(\cdot)$ to represent the unknown right-hand side terms dependent on $X(x,t)$, its derivatives, and other parameters that do not depend on X included in β. We then pretend to construct $N(\cdot)$ from time-series data at a fixed number of coastal locations, using the PDE-FIND algorithm developed by Rudy et al. (2017). The central assumption of the PDE-FIND algorithm is that $N(\cdot)$ is sparse relative to a library of possible nonlinear term candidates for the list of N factors (Θ). In other words, only a selection of nonlinear terms (that N depend on) contributes to the fundamental behavior of the system. Here, we choose the functional form of the nonlinear terms in the library to be polynomial nonlinearities and higher dimensional spatial domains as spatial derivates:

$$\Theta(\mathrm{X}) = [1\ X\ X^2\ X^3 \dots X_x, X_{xx}, X_x^2,] \tag{11}$$

as we support this choice by examples in canonical models of mathematical physics and dynamical systems (Cranmer et al., 2020; J. M. Zhang et al., 2020; Z. Zhang & Liu, 2021).

The optimization framework of the PDE-FIND algorithm is based on the solution to $Ax = b$, where A is the library Θ of nonlinear spatial derivates, and b is equal to the temporal derivates of the system, P_t. The goal is to obtain a sparse x vector that indicates which candidate terms best fit the data. This solution produces a linear equation representing the PDE:

$$X_t = \Theta(\mathrm{X})\mathrm{\alpha} \tag{12}$$

with each column of Θ representing the candidate nonlinear terms, which are dependent on spatial derivates. To solve the linear system, we apply a conventional least-squares regression ($A\backslash b$), as well as the convex relaxation of a sparse regression, referred to as the LASSO method (Rudy et al., 2017), which regularizes over α as a loss function in the form:

$$\alpha = \arg\min_{\hat{\alpha}} \|\Theta\hat{\alpha} - X_t\|_2^2 + \lambda\|\hat{\alpha}\|_1 \tag{13}$$

where λ represents the L1 norm regularization coefficient, the tilde indicates the optimization variable and the function $\arg\min\|\cdot\|$ represents minimization of the loss function. The α quantification assures that the sparse terms in the derived PDE would appear only if their effect on the error $\|\Theta\hat{\alpha} - X_t\|$ outweigh their addition to $\|\hat{\alpha}\|_1$. Allowing a sparse solution of the least-squares problem exhibiting good performance (Brunton & Kutz, 2017). The "backslash" method for solving $Ax = b$ is based on a factorization using the spectral decomposition, therefore it also finds a sparse solution.

Nevertheless, the optimization procedure is meaningless if the nonlinear terms defined as spatial numerical derivates are not accurate. Since wave reanalysis simulations produce relatively clean data, we can apply a Finite Difference numerical method (based on centered differences) to compute the temporal derivates of X (Tu et al., 2014). We can also compute the spatial derivates (X_x, X_{xx}) as a differentiation matrix multiplication of centered differences. We evaluate these ideas below. Finally, we implemented the algorithm in MATLAB® to solve the PDE-FIND methodology and discover the governing symbolic expressions of the system (Rudy et al., 2017 our Figure 11).

To evaluate the discovered terms, we pre-processed the matrix (X) into four different training datasets (Ordinary, Random, Uniform, and Seasonal) to input in the PDE-FIND algorithm. We then computed optimal basis sets (see below), for various distributions and amounts of data. This sequence allows a Pareto analysis framework, where the balance between model complexity and accuracy can be measured and optimized (Veldhuizen & Lamont, 1998).

We look for a balance between model complexity and accuracy by comparing measured and simulated data. We contrast the ordinary variability (raw time series data) of P against seasonal or interannual variability (Butterworth filtered data), principal spectral components (Fourier spectrums identified significant periodicities), and uniform and random distributions from the data. In that sense, we used the different training datasets to obtain nonlinear terms of the measurement matrix and later a reduced-order linearization to further compare the models using the determination coefficient (R^2), mean percentual error ($error$), and the normalized root mean square error ($NRMSE$), following Occam razor's idea that the best model would be the simplest between equally agreeable results.

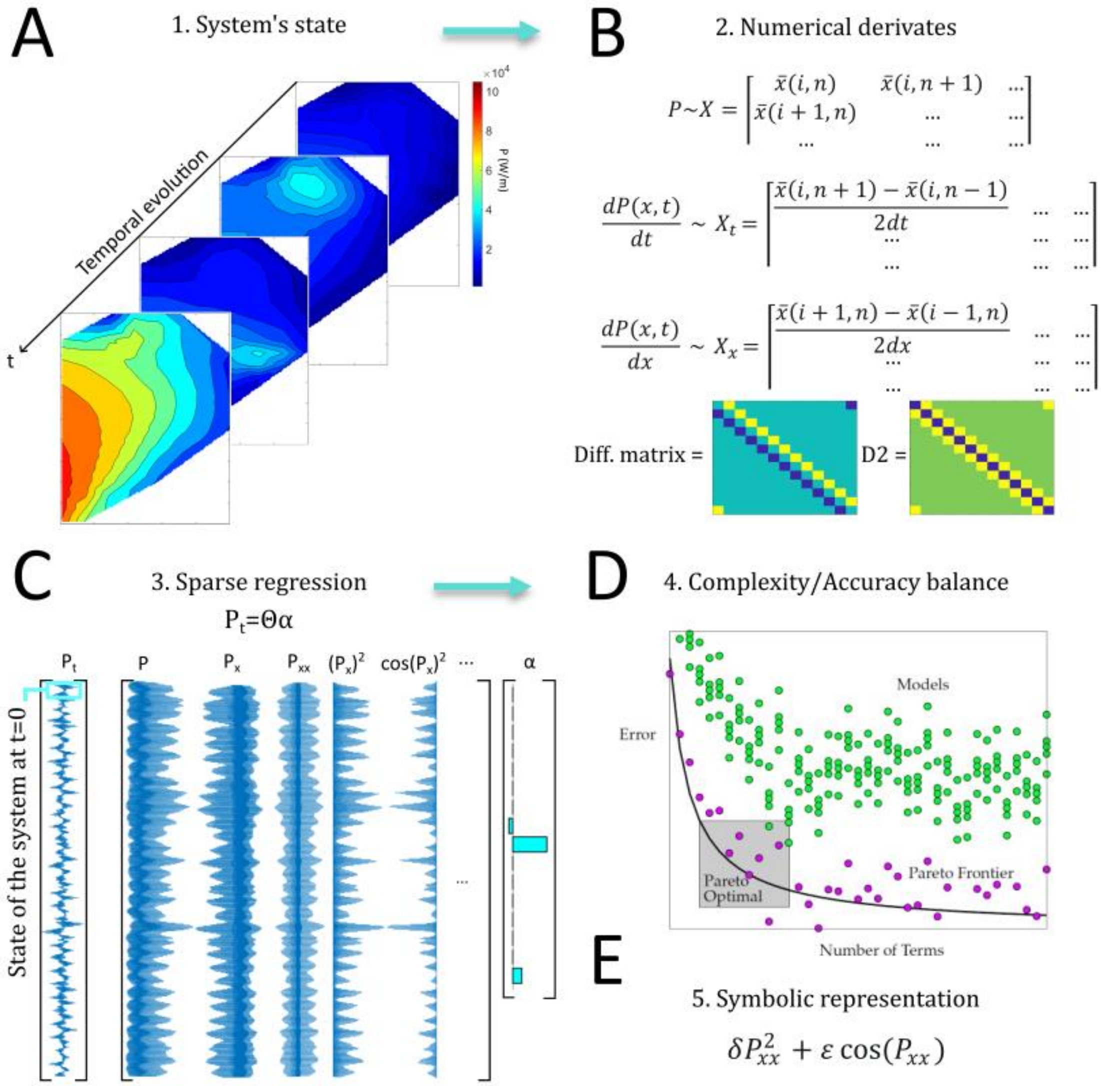

Figure 11. Data-driven discovery of PDEs, showing the model discovery and selection with the PDE-FIND algorithm (Brunton et al., 2016; Rudy et al., 2017). This methodology allows for a model selection that optimizes the mathematical description of the system's nonlinear behavior. Figure adapted from Brunton and Kutz (2017).

2.4. Methods: Dimensionality reduction

We applied the PCA (Pearson, 1901) based on the SVD algorithm to obtain a hierarchical coordinate system to capture the maximum variance in the data (Yule, 1938). Based on the SVD, we further compute the more robust DMD technique (Tu et al., 2014). The goal is to reduce dimensionality into the most significant correlation structures (or dominant patterns) representing a non-square data matrix, which would allow an economical (i.e., accurate and optimal) representation of the system. According to the SVD, we can represent our collection of snapshot measurements of P in time as a data matrix (X), such that can be expressed as:

$$X = U\Sigma V^T, \tag{14}$$

where U is a unitary matrix with orthogonal columns associated with the spatial realizations and correlations of P (rows), thus called left singular vectors or spatial principal components (also denoted as U_{PCA} when truncated into the r most essential components). The Σ represents a matrix with real, non-negative values on the diagonal and zeros off the diagonal. These values represent the loadings, or "weights" (σ), which indicate how much of the original variance is explained by each principal component. The V matrix represents the right singular vectors, or temporal principal components, as their rows are associated with the temporal realizations of P (columns). Both PCA and DMD are based on the SVD, which solves a linear system of the form:

$$X_t = AX, \tag{15}$$

where $X \in R^n \ \forall n \gg 1$. If we apply the solution $X = ve^{\lambda t}$, we obtain the eigendecomposition or spectral decomposition. Similarly, the idea is to get a solution expressed in the basis set that optimizes the spatial variability of the data in a reduced-order modeling framework. The DMD algorithm looks for the best linear operator (A) to advance the data matrix X in time (X_{t+1}) (Schmid, 2010). We find A by multiplying the time derivative by the pseudo-inverse of X expressed as the singular value decomposition $A = X_t X^\dagger = X_t V \Sigma^{-1} U^T$. We can project A into the spatial modes of U and compute the spectral decomposition as $AW = W\Lambda$. The W modes are then used to compute the U_{DMD} such that:

$$U_{DMD} = X' V \Sigma^{-1} W \tag{16}$$

In theory, DMD goes further from PCA, in the sense that the variability of each DMD spatial mode (denoted with φ as well) is not orthonormal anymore, reproducing only specific frequencies that oscillate in a sinusoidal manner like the Fourier transform. This representation works as an advantage in physical systems where the modes relate to specific periodic behavior with a typical growth rate. Nevertheless, PCA's temporal modes are orthonormal and more parsimonious, which is advantageous in applications like Principal Component Space (PCS) and Reduced-Order Models (ROMs). Here, we apply the DMD spectrum to identify periodic behavior. We also use DMD and PCA to classify, describe and predict fundamental patterns in the data as a ROM.

2.5. Methods: Reduced-order dynamical system

In this paper, proper orthogonal decomposition or POD is the main methodology to obtain low-dimensional dynamical patterns to build a ROM. We evaluate and compare two POD architectures using the PCA and DMD basis set ($\varphi_{r1} = U_{PCA}$ and $\varphi_{r2} = U_{DMD}$). We assume that the defined nonlinear system $P_t \sim X_t$, discovered using PDE-FIND, can be solved by separation of scalar variables or modal expansion in space and time, so that:

$$X(x,t) = \sum_{r=1}^{n} a_r(t)\varphi_r(x), \tag{17}$$

where $a(t)$ represents the time dependence of the nonlinear behavior, φ_r represents the spatial dependence assumed to be embedded in the matrix decomposition, and r represents the truncated modes (from 1 to R). To ensure that an analytical separation of variables can occur, we assume Eq. 5 with constant coefficients. This expansion produces a low-rank set of ODEs that approximates the true solution of the high-dimensional PDEs. If we plug Eq. 17 into Eq. 10, represented by scalar values, we obtain:

$$a_r(t)_t\varphi_r(x) = N[a_r\varphi_r, a_r(\varphi_r)_x, a_r(\varphi_r)_{xx}, \dots, x, t, \beta] \tag{18}$$

Now, the form of $N[\cdot]$ determines the nonlinear mode-mixing that occurs between the r modes. By obtaining the most effective basis set (φ_r) and selecting the most important principal components (r), we could model the nonlinear effects of the system.

To compute the ROM of the variable $a(t)$, we find the PDE form of the nonlinear terms and calculate the optimal basis set, which multiplied by φ_r' (given that $\varphi_r'\varphi_r = I$ by orthonormality in Galerkin projections) would yield (Rowley et al., 2004):

$$a_r(t)_t = \varphi_r' N(a_r, \varphi_r, (\varphi_r)_x, (\varphi_r)_{xx}, \dots, x, t, \beta) \tag{19}$$

From this mathematical framework, we present the discovery of nonlinear PDEs with constant coefficients, the analytical development of a reduced-order description, and the computational algorithms that allow to simulate and solve the spatio-temporal dynamics of the system.

2.6. Results and discussion: Exploratory analysis

Figure 12-A shows the statistical trends of P along the coast by contrasting the mean values of P as well as the variability coefficient (COV) at each nearshore station. This latitudinal representation relates to the average picture of the system of interest. The latitudinal trend suggests an S-N, alongshore rise in wave energy, supporting the idea that continental shelf morphology, controlled mainly by the compressional geologic setting (Correa & Morton, 2010), exerts control on P variability, mostly by bottom dissipation (Ardhuin et al., 2009). We observe that a wide continental shelf (Northern stations) presents higher energy values with low temporal variability than a narrow shelf (Southern stations). Lower latitudes (Southern stations) might be more susceptible to wave dissipation mechanisms, increasing the nonlinear interactions of different wave environments and producing more stochastic behavior. Furthermore, we cannot identify the temporal evolution by observing the statistical behavior in Figure 12-A as the mean state of the system in 30 years.

Figure 12-B shows the classical characterization of wave environments as a function of statistical wave parameters as dimensions in coordinate space. This representation permits a clear and intuitive classification of swell and local waves. We define local waves as exhibiting peak periods <10 seconds, whereas swell periods range from 10 to 20 seconds. We observe that most wave energy comes from swell environments arriving between 230° and 340° azimuth, with wave heights indicating moderate-to-high energy nearshore environments (Restrepo & Kjerfve, 2002).

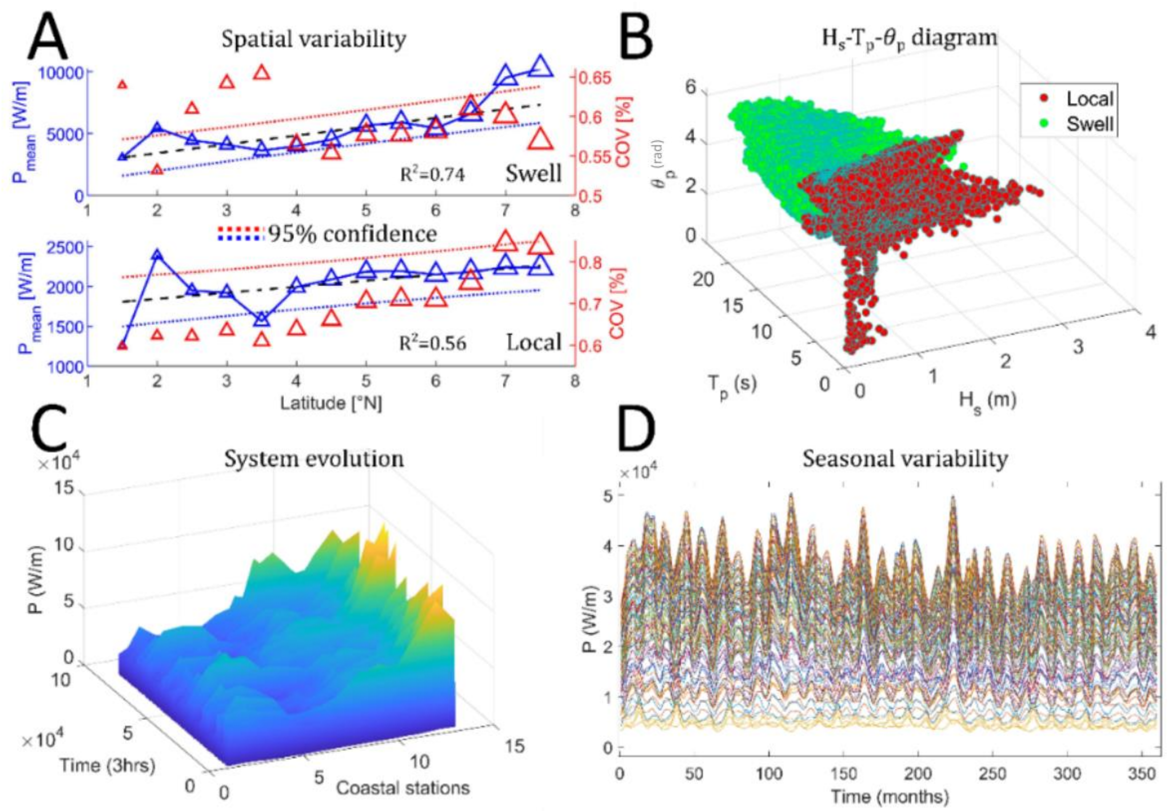

Figure 12. Experimental results illustrating statistical trends in the spatial variability of P along the continental shelf (A), characterization of wave climate into swell and local environments by wave parameters (B), raw spatio-temporal evolution (C), and seasonal variability of the system (D). Statistical trends and dominant patterns provide practical information and intuition about the system, which is essential to generate the possible symbolic terms that describe the data.

2.7. Results and discussion: Model discovery and selection

Since the interest of the reduced framework is to solve a system in the form of $Ax = b$, we define our system as an overdetermined, as we have few unknowns (13 alongshore stations) and many constraints (87,663 snapshots). This definition is necessary for obtaining a detailed spatial variability with a low temporal resolution, dictating the type of regression we should apply.

Figure 13 presents the PDE discovery process from different training datasets, the obtained regression, and the model selection to evaluate errors and get the simplest model. We use these datasets to assess how other distributions in the data impact model errors in predicted results. Using random and uniform distributions extracted from the data, we evaluate how statistically robust each proposed model is, as the seasonal and ordinary models explain the behavior of interest in the system.

To solve $Ax = b$ in each of the training datasets, we applied two methods: "backslash" ($A\backslash b$) and LASSO. Overall, we observe that ordinary and seasonal variability render the same mathematical representation as random and uniform distributions. The PSC is the only time series with different solutions in both regression methods. Interestingly, the "backslash" solution applied to PSC resulted in the same terms as almost all the LASSO solutions for all different training datasets. This result suggests that the Fourier basis representation could encode information like the one identified by the LASSO approach.

To select the ideal terms from the PDE and evaluate the complexity/accuracy balance, we reduced each individual term into the optimal basis to linearize each symbolic representation (ROM of each term), as well as adding all the terms to obtain a 6-term PDE and later 6-term ODE (ROM of the complete PDE). We linearized the system embedding the terms in a low-rank structure that reproduces the most important dynamic evolution. By simulating each term and the complete PDE, allowed us to find the Pareto front, defined as the inflexion point of the curve where the error starts increasing and the number of terms is decreasing (Figure 11-D, as well as Figure 13-C). We find that the error using only the first term (P_x) is $NRMSE = 36.63\%$ and the error using four terms ($P_x + P^2{}_x + PP_x + {P_{xx}}^2 + \cos{(P_{xx})}$) is $NRMSE = 37.16\%$. Thus, we must select the PDE with one term and conclude that the error obtained from the model is almost entirely dependent on the basis set that embeds the dynamics of the system, the ROM. Therefore, the error analysis performed in the next sections is based on evaluating different training datasets (seasonal, ordinary, random, and uniform) on reduced-order algorithms (PCA and DMD).

After running the model with each dataset and symbolic term, as well as the model containing all terms added together, the Pareto analysis showed that the error of using only one reduced-order term ($a_x = \varphi_r'\varphi_x a$) was 21% ($NRMSE = 38$) while the error using all the five discovered terms ($a_x = \varphi_r'\varphi_x a + \varphi_r'{\varphi^2}_x a^2 + \varphi_x a^2 + \varphi_r'{\varphi_{xx}}^2 a^2 + \varphi_r'\cos(\varphi_x a)$) was 18% ($NRMSE = 36.6\%$). This indicates that the ROM's robustness lies in the computation of the optimal basis set to embed the data and the number of truncated modes to expand the data, rather than the symbolic expression used to describe it. The ROMs developed on our system do not seem to depend much on the nonlinear terms, but rather on the basis functions that linearize the system. Consequently, we conclude that the Pareto analysis cannot perform correctly assuming the model's complexity as a function of the number of nonlinear terms, which is not as efficient as assuming the model's complexity or computational expense as a function of the number of elements in the training data. We observe that the errors depend upon the information represented on the time series signals rather than the amount of data, and we also show that the moving average filtering technique applied does not affect the error, but the opposite is found for ordinary training datasets where the filtered data improved the error.

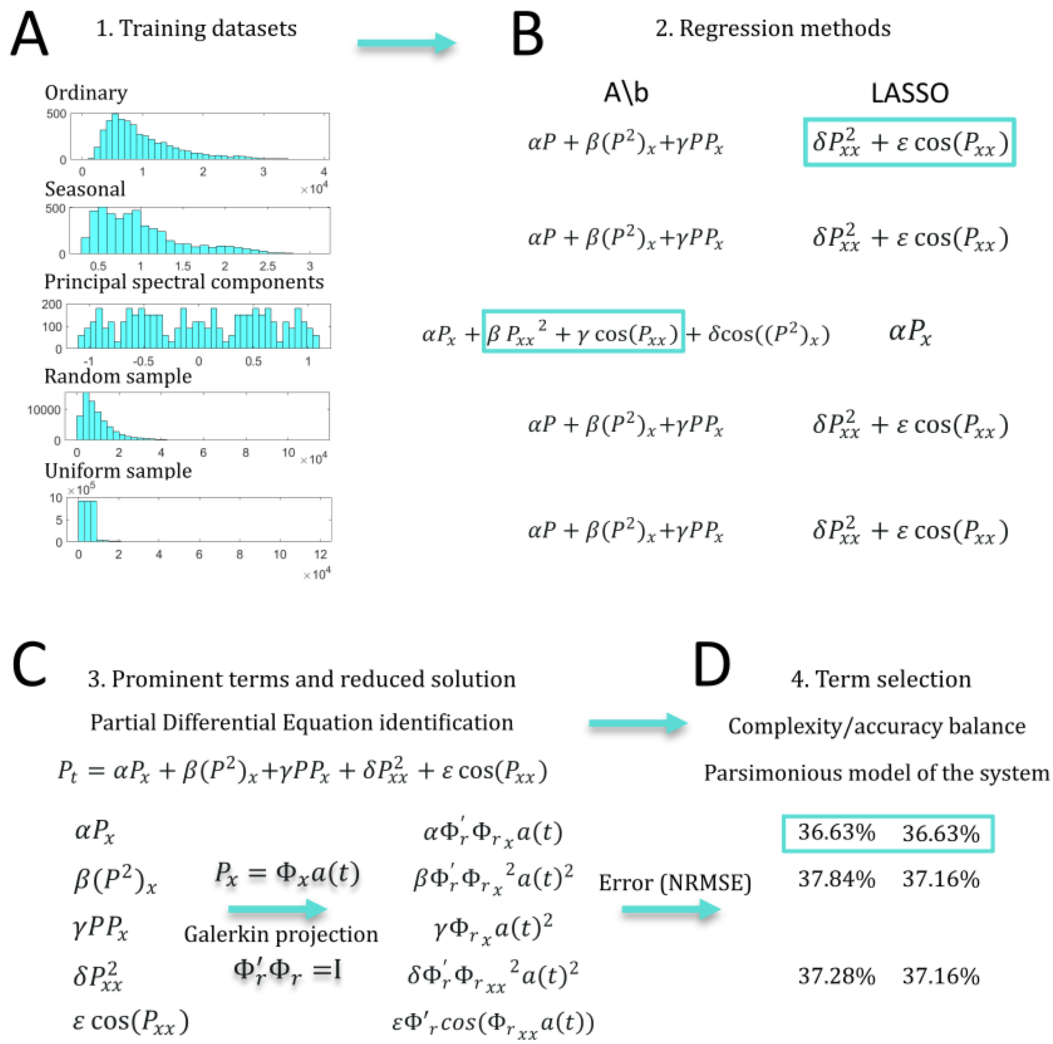

Figure 13. Methodological diagram proposed in the present study. We applied it to different training datasets (step 1) to discover the possible nonlinear terms (step 2), select the appropriate reduced model (step 3), and evaluate the simulated results versus actual data (step 4). This result illustrates the developed algorithm and conceptual method to adequately obtain a parsimonious PDE description of the fundamental spatio-temporal variability of P along the continental shelf of a compressive tectonic margin.

2.8. Results and discussion: Model reduction and solution

After discovering the ideal nonlinear term that best represents the data, we find the low-rank structure that allows linearizing the system. Thus, we compute the SVD algorithm applied to normalize training data to generate the PCA modes along the basin and the coastline (Figure 14-A). Further, we use the Koopman operator to produce the DMD modes (Figure 14-B). We contrast different basis sets with three and five-mode expansions to solve the reduced-order dynamical system and find that the error using a three-mode expansion, which results in equally significant values as the five-mode expansion. We observe the modes in the basin to identify the wind-wave arriving trends on each mode. Therefore, the spatial modes computed represent the alongshelf variability of P (Figure 14-C and D). From the SVD decomposition, the cumulative percent of variance and the singular values show that the first three modes represent almost 95% of the total variability for the seasonal time series (Figure 14-F). The ordinary training data set shows that the first three modes explain 70% of the total variability. This contrast highlights that we can model the hierarchical processes that drive most of the energy in the system using a three-mode expansion of the ROM.

We used the matrix V^T from the SVD to compute an optimal coordinate system that represents the temporal variability of each station ($x = {V^T}_{r=1} \cdot X_{i_{r=1}}$, $y = {V^T}_{r=2} \cdot X_{i_{r=2}}$, $x = {V^T}_{r=3} \cdot X_{i_{r=3}}$). By multiplying each truncated temporal mode (V^T) by each station's time series (X_i), we obtain a unique coordinate point representing wave energy flux variability in terms of the first three temporal modes in the optimal coordinate system. This representation corresponds to the Principal Component Space, where we statistically model and classify each station's unique "fingerprint" as a function of nearshore morphodynamic regimes. The three colors in Figure 14-E correspond to the morphodynamic regimes we identified (i.e., Northern: cyan, Middle: red, Southern: blue).

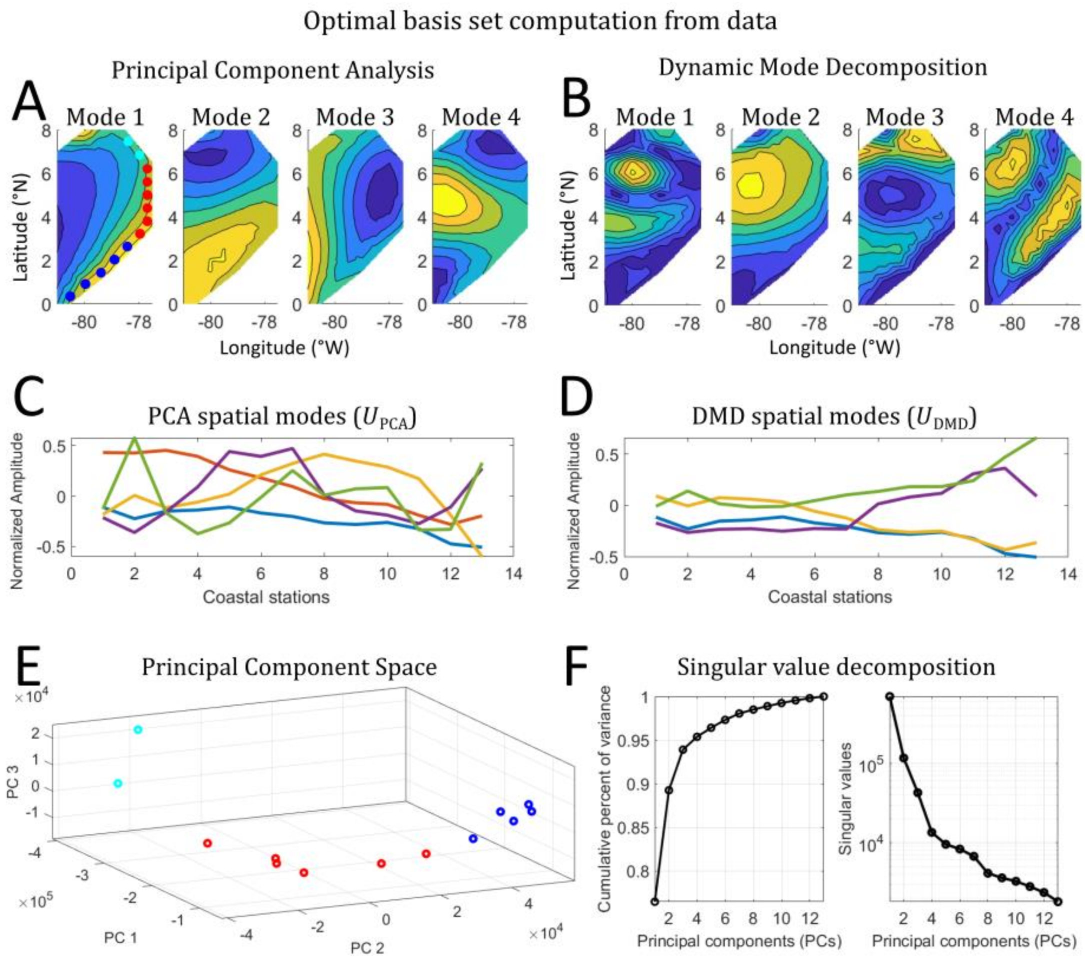

Figure 14. Optimal basis set computation from data in the form of two algorithms: principal component analysis or PCA (Panel A), and dynamic mode decomposition or DMD (Panel B). Each algorithm produces a basis set of hierarchical spatial variability patterns used to embed the dynamics and linearize the system. Panel C shows the spatial modes of the PCA, and Panel D shows the DMD spatial modes. Panel E represents the coordinate representation known as the principal component space (PCS), where each station presents a unique coordinate as a function of the principal components (PCs). Panel F illustrates important information for both techniques, since it shows the hierarchical information in the singular value decomposition used for both approaches.

We found the selected ideal model of the dynamical system of interest to be:

$$P_t \approx X_t = \alpha X_x + \beta \tag{20}$$

where α is the coefficient for the nonlinear term, and β represents the parameter spa ce. By assuming the separation of variables $X(x,t) = \varphi_r(x)a_r(t)$ given independence between time and space, we linearize the system as a three-mode expansion, which now depends on the coefficients $a(t)$ of the basis functions $\varphi_r(x)$, such as:

$$\varphi(x)a(t)_t \approx \alpha \left(\varphi_{1_x}(x)a_1(t) + \varphi_{2_x}(x)a_2(t) + \varphi_{3_x}(x)a_3(t)\right) + \beta \qquad (21)$$

Since the time dynamics $a(t)$ are assumed to change linearly, we define $a_1(t) + a_2(t) + a_3(t) = a(t)$. Additionally, due to the orthogonality properties of the SVD algorithm, we assume $\varphi_r{}'\varphi_r = \mathrm{I}$ by the Galerkin projection, so we can multiply Eq. 21 by $\varphi_r{}'$ to obtain

$$a(t)_t = \alpha \left(\varphi_1{}'(x)\varphi_{1_x}(x) + \varphi_2{}'(x)\varphi_{2_x}(x) + \varphi_3{}'(x)\varphi_{3_x}(x)\right) a(t) + \beta \qquad (22)$$

Eq. 22 represents the ODE of the reduced-order dynamics of the wave energy flux P. We numerically solve Eq. 22 using the ODE45 algorithm in MATLAB®, which outputs the dynamical solution $a_{sim}(t)_t$. We finally reconstruct the system as

$$X_{sim} = \varphi_r(x)a_{sim}(t) \qquad (23)$$

2.9. Results and discussion: Model evaluation

We evaluate the model by contrasting several training datasets regarding the error and the number of measurements to find the balance between model accuracy and complexity, like the Pareto front. The number of terms reflects on the computational expense of the model and the robustness of the dimensionality reduction techniques. These considerations usually relate to the number of terms in the PDE. Both measures dictate the model's computational complexity. Since we found that one term is equally significant to six terms, we apply the number of measuring samples to quantify model complexity.

We compute the error and the determination coefficient of each dataset to evaluate against different amounts and components of time series training data. Figure 15 presents the best-fitted model of both ROMs, i.e., PCA (Panels A-B) and DMD (Panels D-E). We include the Pareto analysis in Figure 14C as the balance between the determination coefficient and the number of training samples. We observe that the random training dataset computed with training samples from 100 to 100,000 elements shows R^2 values from 0.6 to 0.9, indicating a relatively strong linear relation between simulated and observed data. We additionally present the results of the model evaluation applied to all training datasets. We include their respective root mean square error (RMSE) analysis, linear regression, and mean error (Figure 16).

The Pareto analysis shows that the ordinary training datasets from 100 to 1,000 elements present a relatively significant decrease in R^2 values as the model complexity increases. This result suggests the opposite behavior expected from the Pareto analysis using the number of terms, which would over-adjust the model as the complexity increases. We observe that the seasonal training data maintains virtually the same R^2 values as model complexity increases. Overall, we find that the Butterworth filter applied to reduce the number of elements in the training data (model complexity reduction) improves modeled results. Thus, the results show good agreement between extrapolated and observed data in the cases of seasonal and ordinary variability. Nevertheless, the random training datasets evaluation shows relatively good performance of the ROM for any number of training elements.

We solved the dynamical system using the discovered equation, but since the first term in Burger's wave equation is equal to our found functional form (αX_x), we added the diffusion term of the wave equation (βX_{xx}). Similar to previous results, the error does not change. We still obtain a 15.5% error and NRMSE of 38%, Thus, the system's linearization by the optimal basis functions no longer requires specific coefficients and complex nonlinear functional forms. In other words, we can produce a robust ROM by averaging time series and using first-order spatial derivates to compute a linearization that captures the fundamental spatial variability of P along continental shelves.

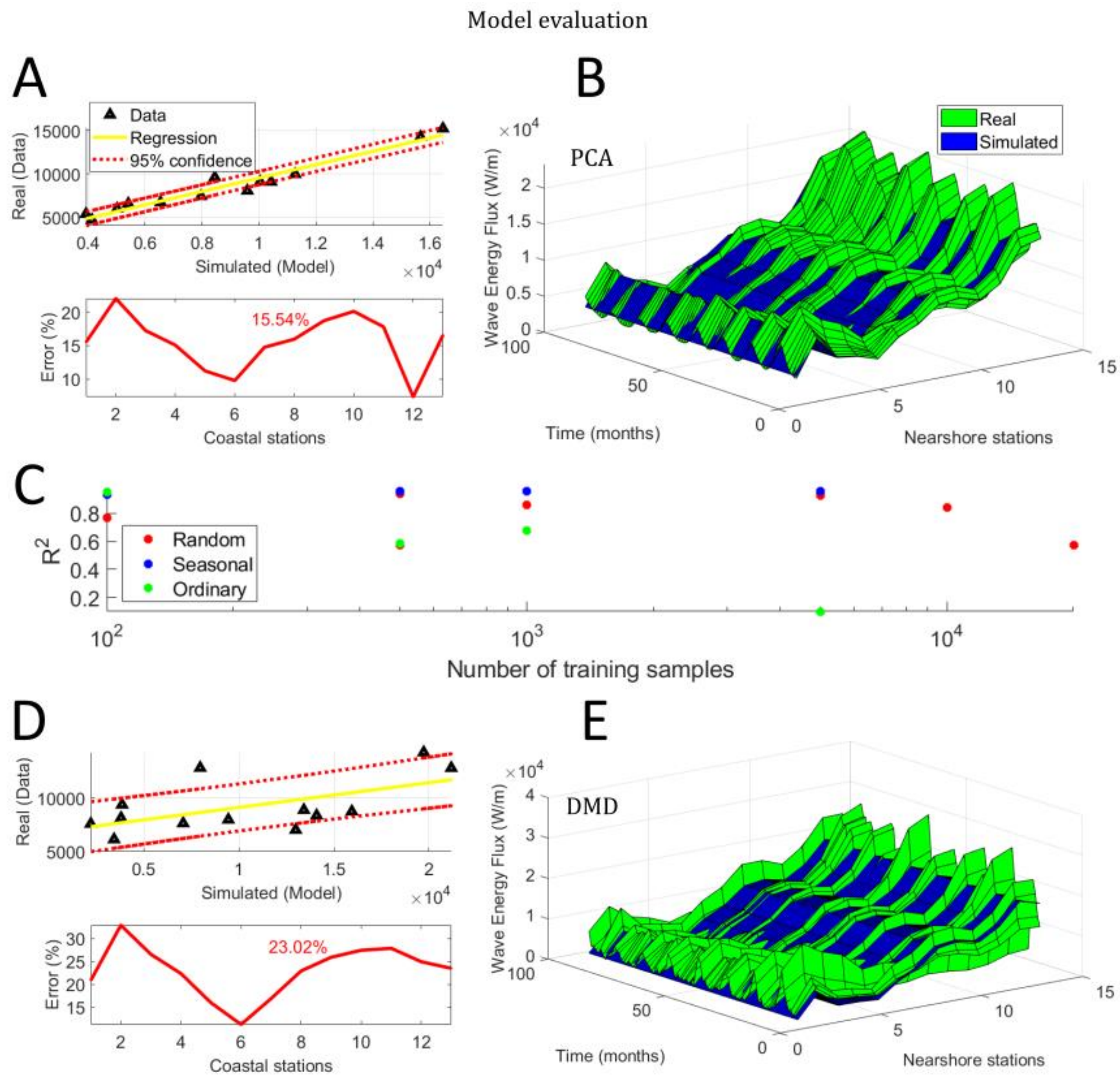

Figure 15. Model evaluation of the best fitted PCA (Panel D and E) and DMD (Panel A and B) ROMs, together with the graphical representation of the Pareto analysis (accuracy vs complexity plot) applied to the different training datasets (Panel C).

In Figures 15 and 16, the seasonal ROM shows over-estimations on the Northern stations and under-estimations on Southern stations, with an overall good performance (15.5% of error). The ordinary ROM also presents the same behavior, reducing the performance (23.7%. error). In contrast, the uniform distributed data shows overall good performance (18.5% error), with similar behavior to seasonal and ordinary ROMs in over and under-estimated values along the coast. More importantly, the random trained ROM presents the more significant result, with a 10.5% error and relatively little (and almost constant) over-estimations at the nearshore stations.

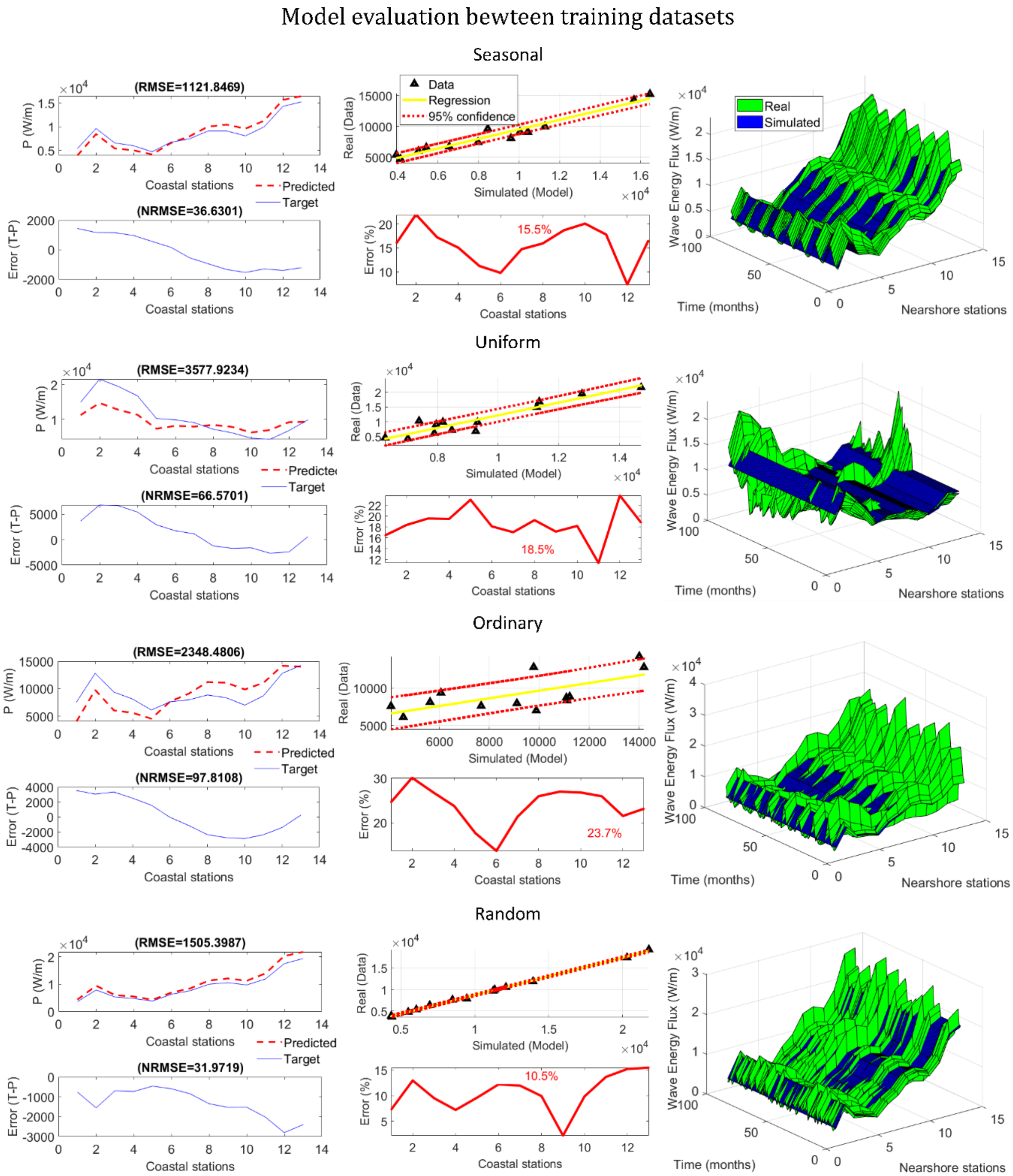

Figure 16. Model evaluation for different training datasets. From top to bottom: seasonal, uniform, ordinary, and random. The left column shows the RMSE analysis, the center column the linear model between simulated and real values with the mean standard and percentual error values, and the right column shows the simulation and real evolution of the spatio-temporal system.

By forecasting 100 years using the seasonal, ordinary, and uniform ROMs, we found specific trends for each station, and more importantly, morphodynamical regime. We observe that the Northern regime appears to decrease in energy flux as time progresses, while the Middle regime increases drastically with time and latitude. The Southern regime also presents a clear and accelerated decrease in wave energy flux in time. These trends provide an important future state assessment of coastal morphodynamics driven by mesoscale wave climate in the Northern Andes Pacific coastline.

Overall, we accurately find the analytical expression that best represents the defined system's evolution data, and then we reduce it into its fundamental modes, obtaining a proxy linear model for future predictions. The model was compared to itself due to the lack of real measurements along the area, which is a recurrent problem in underdeveloped countries. The model was evaluated using different training data varying in size and distribution, so we could obtain a parsimonious relation between accuracy and complexity. We obtained the best results using filtered time series of seasonal variability with a temporal reduction in data elements of 95% and a spatial reduction of 75%. As well as reducing a $13\ x\ 87663$ size PDE to a $3\ x\ 360$ size ODE, and still simulating a $13\ x\ future\ time\ elements$ system.

Despite the efforts to find the ideal symbolic representation of the system, the linearization resulting from the reduced-order framework is the most important step in accurately modeling the system's behavior. Nevertheless, symbolic regression, which can also find the analytical expression of the data without assuming the creativity to generate the possibilities, is a remarkable technique that combined with graph neural networks, encouraging sparse latent representations to distill symbolic representations, could be another important method worth exploring in future work.

3. Conclusions and future work

The proposed framework and modeling architecture identifies trends and dominant dynamic processes driving P variability, which, as shown before, can effectively be associated with morphodynamical conditions and sub-aerial profiles along the continental shelf. New measured time series of wave climate from coastal environments can be interpolated in the PCS to classify and statistically model recent wave-induced coastal morphodynamics. Since we could not have access to new measurements, we test our model with the latest available reanalysis data to corroborate the significance of the model. This conclusion pretends to illustrate that, even though we live in an era of overwhelming amounts of data, both data scientists and data-mining tools are still more available than scientific measurements in most developing countries of the South American continent.

Nevertheless, we apply the proposed data-driven analytics of wave climate to elucidate coastal morphodynamics along the continental shelf of the compressive tectonic margin in the Northern Andes Pacific basin. We obtain and propose three distinct morphodynamical regimes associated with unstable feedback processes in shoreline shape formation in a low-rank, optimal coordinate system. Which provides relevant information on wave-induced long-term coastal evolution and helps further understanding the complex system of wave-induced sediment transport in local bathymetry.

On the other hand, the discovered ROM was trained using data from 1980 to 2000, to further predict from 2000 to 2010, as well as 2100. We find significant agreement between extrapolated values and high-fidelity global simulations (R2=0.86 and NRMSE=15.14%), for both training datasets of identified periodic behavior and random distributions for the 2010 predictions. More importantly, we found decadal or mesoscale behavior by predicting coastal values of P into 2100 associated with wave-induced coastal morphodynamics. Our results provide a modeling architecture of wave climate, illuminating the importance of hidden leading-order variables in complex systems such as wave-induced sediment transport along continental shelves. The found mesoscale trends of each proposed morphodynamical regime represent an important assessment of wave-induced coastal evolution. Additionally, we aim to work further on the bifurcation analysis of the parameter beta, this would allow the characterization and modeling of different scenarios of the system based on empirical adjustments.

We define a complex system from high-fidelity simulated wave climate data. Meaning, we assume that the selected variable (wave energy flux) represents the system of interest to model wave-induced morphodynamics along continental shelves. We also assume that the measured system is the solution of a set of PDEs in space and time and that it can be solved as a linear problem by the separation of variables, such that the dynamic patterns of interest are embedded in each variable. We numerically solve the model for different training datasets and dimensionality reduction algorithms to evaluate the parsimony of the model and obtain a low-rank predictive model. We conclude that our approach is efficient and produces a parsimonious model of reality.

4. Supplementary methods (appendix)

We present the project data architecture proposed to assess wave energy by a variety of data-driven techniques (Figure 16). The pre-processing stage defines the system and computes the wave energy flux (P) from WW3 wave parameters. The processing of time series of P into seasonal, multi-year, spectral components and anomaly variability allows evaluating different training datasets for the ROM to produce a parsimonious model. The data-driven descriptive analytics were classification and identification of swell and local waves, using linear regression and ANOVA, supervised and unsupervised learning algorithms, as well as wavelet and Fourier analysis. The data-driven predictive analytics consists of the PDE-FIND algorithm applied to both the Principal Component Space and training datasets to discover the parsimonious expression for the system, while the ROM is applied to reduce dimensionality and computational expense in solving future states of the system of P.

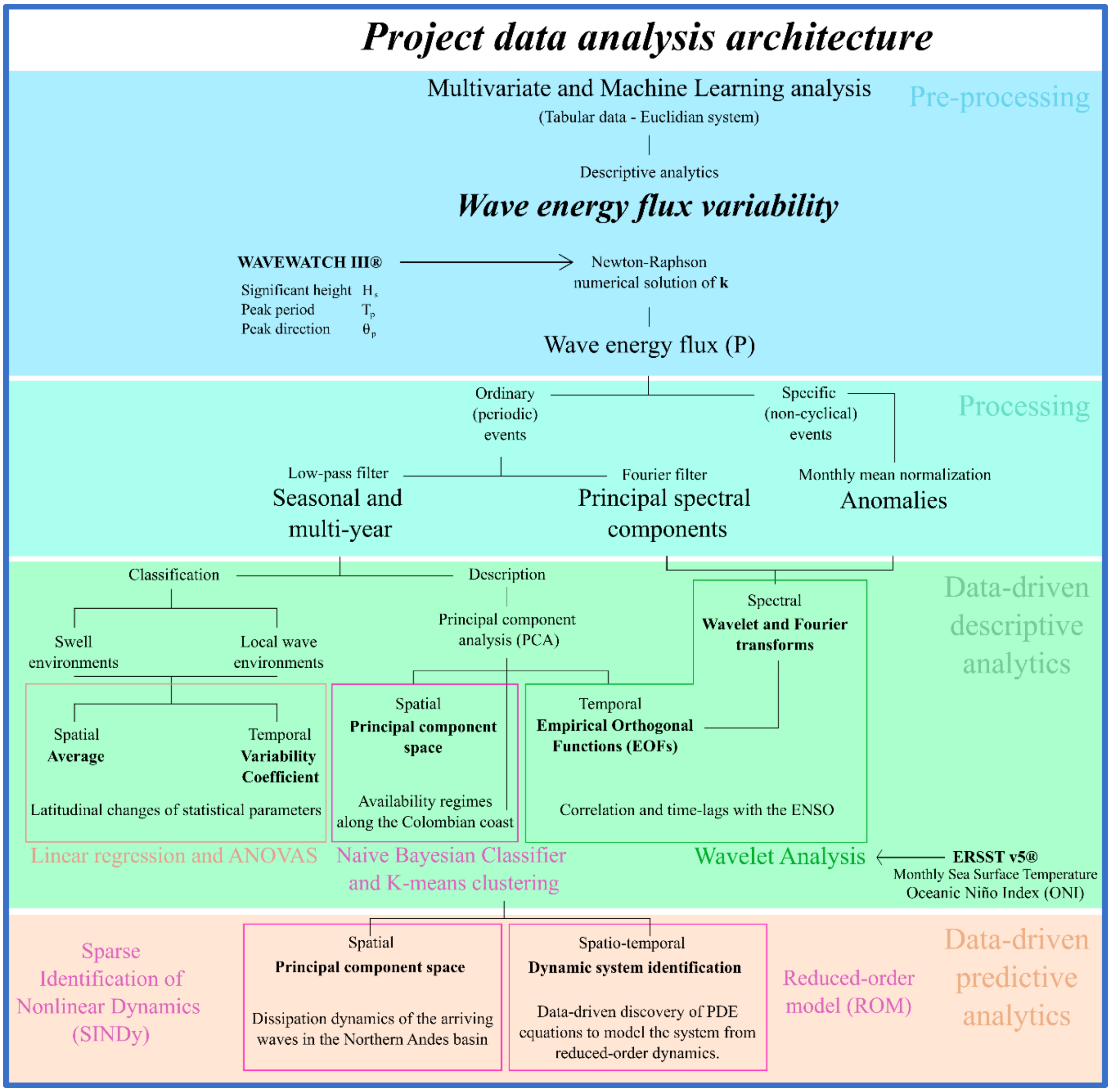

Figure 17. Project data analysis architecture and modeling framework proposed to assess wave climate data, with a focus on shoreface morphodynamics and coastal evolution. The architecture stages are pre-processing, processing, data-driven descriptive analytics, and data-driven predictive analytics. The pre-processing stage defines the system and computes the wave energy flux (P) from WW3 wave parameters. Processing P time series data into seasonal and multi-year variability, the principal spectral components, and the anomalies in the data allowed the evaluation of different training data for the ROM. The data-driven descriptive analytics were applied by classification and identification of swell and local waves, using linear regression and ANOVA, supervised and unsupervised learning algorithms, as well as wavelet and Fourier analysis. The data-driven predictive analytics consists of the PDE-FIND algorithm applied to both the PCS and time-series datasets to discover the parsimonious expression for the system, while the ROM is applied to reduce dimensionality and computational expense in solving future states of the system of P.

5. References

Almar, R., Ranasinghe, R., Bergsma, E. W. J., Diaz, H., Melet, A., Papa, F., Vousdoukas, M., Athanasiou, P., Dada, O., Almeida, L. P., & Kestenare, E. (2021). A global analysis of extreme coastal water levels with implications for potential coastal overtopping. *Nature Communications*, *12*(1), 3775. https://doi.org/10.1038/s41467-021-24008-9

Ardhuin, F., Chapron, B., & Collard, F. (2009). Observation of swell dissipation across oceans. *Geophysical Research Letters*, *36*(6), L06607. https://doi.org/10.1029/2008GL037030

Ardhuin, F., O'Reilly, W. C., Herbers, T. H. C., & Jessen, P. F. (2003). Swell Transformation across the Continental Shelf. Part I: Attenuation and Directional Broadening. *JOURNAL OF PHYSICAL OCEANOGRAPHY*, *33*, 19.

Ashton, A., Murray, A. B., & Arnoult, O. (2001). Formation of coastline features by large-scale instabilities induced by high-angle waves. *Nature*, *414*(6861), 296–300. https://doi.org/10.1038/35104541

Athanassoulis, G. A., & Belibassakis, K. A. (1999). A consistent coupled-mode theory for the propagation of small-amplitude water waves over variable bathymetry regions. *Journal of Fluid Mechanics*, *389*, 275–301. https://doi.org/10.1017/S0022112099004978

Bai, Z., Kaiser, E., Proctor, J. L., Kutz, J. N., & Brunton, S. L. (2020). Dynamic Mode Decomposition for Compressive System Identification. *AIAA Journal*, *58*(2), 561–574. https://doi.org/10.2514/1.J057870

Bender, C. J., & Dean, R. G. (2003). Wave field modification by bathymetric anomalies and resulting shoreline changes: A review with recent results. *Coastal Engineering*, *49*(1–2), 125–153. https://doi.org/10.1016/S0378-3839(03)00061-9

Bromirski, P. D., Cayan, D. R., Helly, J., & Wittmann, P. (2013). Wave power variability and trends across the North Pacific: WAVE POWER ACROSS THE NORTH PACIFIC. *Journal of Geophysical Research: Oceans*, *118*(12), 6329–6348. https://doi.org/10.1002/2013JC009189

Broomhead, D. S., & King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, *20*(2–3), 217–236. https://doi.org/10.1016/0167-2789(86)90031-X

Brunton, S. L., & Kutz, J. N. (2017). *Data Driven Science & Engineering*. 572.

Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine Learning for Fluid Mechanics. *Annual Review of Fluid Mechanics*, *52*(1), 477–508. https://doi.org/10.1146/annurev-fluid-010719-060214

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932–3937. https://doi.org/10.1073/pnas.1517384113

Butterworth, S. (1930). On the Theory of Filter Amplifiers. *Experimental Wireless*, *The wireless engineer*, 536–541.

Caicedo-Laurido, A., Marriaga, L., Iriarte, J., Cabeza, L., Raynero Zurita, Grados, M., Tello, E., García, W., Pizarro, L., Sánchez, S., Bernales, A., Elcira Delgado, Chang, F., Jacobo, N., Hernan Reyes, Grendi, C., Perugachi, C., Recalde, S., Vera, L., …

Gabriel, R. (2019). *Informe Final del XVIII Crucero Regional Conjunto de Investigación Oceanográfica en el Pacífico Sudeste: Condiciones físicas y químicas del océano Pacífico tropical oriental en agosto-octubre 2015 y enero 2016. Comisión Permanente del Pacífico Sur - CPPS. Guayaquil, Ecuador.* https://doi.org/10.13140/RG.2.2.34051.04649

Camus, P., Cofiño, A. S., Mendez, F. J., & Medina, R. (2011). Multivariate Wave Climate Using Self-Organizing Maps. *Journal of Atmospheric and Oceanic Technology*, *28*(11), 1554–1568. https://doi.org/10.1175/JTECH-D-11-00027.1

Camus, P., Mendez, F. J., Medina, R., & Cofiño, A. S. (2011). Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering*, *58*(6), 453–462. https://doi.org/10.1016/j.coastaleng.2011.02.003

Chawla, A., Spindler, D., & Tolman, H. (2011). *WAVEWATCH III R Hindcasts with Re-analysis winds†. Initial report on model setup*. 100.

Chiang, C. M., & Stiassnie, M. (2005). *Theory and Applications of Ocean Surface Waves (Advanced Series on Ocean Engineering)*.

Christakos, K., Varlas, G., Cheliotis, I., Spyrou, C., Aarnes, O. J., & Furevik, B. R. (2020). Characterization of Wind-Sea- and Swell-Induced Wave Energy along the Norwegian Coast. *Atmosphere*, *11*(2), 166. https://doi.org/10.3390/atmos11020166

Conlin, M. P., Adams, P. N., Jaeger, J. M., & MacKenzie, R. (2020). Quantifying Seasonal-to-Interannual-Scale Storm Impacts on Morphology Along a Cuspate Coast With a Hybrid Empirical Orthogonal Function Approach. *Journal of Geophysical Research: Earth Surface*, *125*(12). https://doi.org/10.1029/2020JF005617

Correa, I. D. (1996). *LE LITTORAL PACIFIQUE COLOMBIEN: INTERDEPENDANCE DES AGENTS MORPHOSTRUCTURAUX ET HYDRODYNAMIQUES. TOME 1: TEXTE*. https://doi.org/10.13140/RG.2.1.2154.3444

Correa, I. D., & Gonzalez, J. L. (2000). Coastal erosion and village relocation: A Colombian case study. *Ocean & Coastal Management*, *43*(1), 51–64. https://doi.org/10.1016/S0964-5691(99)00066-6

Correa, I. D., & Morton, R. (2010). Pacific Coast of Colombia. In E. C. F. Bird (Ed.), *Encyclopedia of the World's Coastal Landforms* (pp. 193–198). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8639-7_29

Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., & Ho, S. (2020). Discovering Symbolic Models from Deep Learning with Inductive Biases. *ArXiv:2006.11287 [Astro-Ph, Physics:Physics, Stat]*. http://arxiv.org/abs/2006.11287

Dimar. (2020). *Compilación Oceanográfica de la Cuenca Pacífica Colombiana II* (Vol. 9). Dirección General Marítima Editorial Dimar.

Duda, R. O., Hart, P. E., & Stork, D. G. (1995). Part 1: Pattern Classi cation. *John Wiley & Sons, Inc.*, 69.

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, *21*(2), 553–565. https://doi.org/10.1093/bib/bbz016

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, *11*(5/6), 561–566. https://doi.org/10.5194/npg-11-561-2004

Hallermeier, R. J. (1980). A profile zonation for seasonal sand beaches from wave climate. *Coastal Engineering*, *4*, 253–277. https://doi.org/10.1016/0378-3839(80)90022-8

Hamon-Kerivel, K., Cooper, A., Jackson, D., Sedrati, M., & Guisado Pintado, E. (2020). Shoreface mesoscale morphodynamics: A review. *Earth-Science Reviews*, *209*, 103330. https://doi.org/10.1016/j.earscirev.2020.103330

Hapke, C. J., Plant, N. G., Henderson, Rachel. E., Schwab, W. C., & Nelson, T. R. (2016). Decoupling processes and scales of shoreline morphodynamics. *Marine Geology*, *381*, 42–53. https://doi.org/10.1016/j.margeo.2016.08.008

Hasselmann, D. E., M. Dunkel, & J. A. Ewing. (1980). Directional wave spectra observed durign JONSWAP 1973. *JOURNAL OF PHYSICAL OCEANOGRAPHY*, *10*.

Holmes, P. (2005). NINETY PLUS THIRTY YEARS OF NONLINEAR DYNAMICS: LESS IS MORE AND MORE IS DIFFERENT. *International Journal of Bifurcation and Chaos*, *15*(09), 2703–2716. https://doi.org/10.1142/S0218127405013678

Holthuijsen, L. H. (2007). *Waves in Oceanic and Coastal Waters*. Cambridge University Press.

Ingeominas. (1996). *Atlas Geomorfología y Aspectos Erosivos del Litoral Caribe Colombiano.* Instituto Colombiano de Geología y Minería. Publicación Geológica Especial N° 21. Bogotá,.

Köhler, A., Ohrnberger, M., & Scherbaum, F. (2010). Unsupervised pattern recognition in continuous seismic wavefield records using Self-Organizing Maps: Unsupervised seismic pattern recognition. *Geophysical Journal International*, *182*(3), 1619–1630. https://doi.org/10.1111/j.1365-246X.2010.04709.x

Komar, P. D., & Holman, R. A. (1986). Coastal Processes and the Development of Shoreline Erosion. *Annual Review of Earth and Planetary Sciences*, *14*(1), 237–265. https://doi.org/10.1146/annurev.ea.14.050186.001321

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694

Kumar, P., & Foufoula-Georgiou, E. (1997). Wavelet analysis for geophysical applications. *Reviews of Geophysics*, *35*(4), 385–412. https://doi.org/10.1029/97RG00427

Langley, P. (1993). Induction of recursive Bayesian classifiers. In P. B. Brazdil (Ed.), *Machine Learning: ECML-93* (Vol. 667, pp. 153–164). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-56602-3_134

Langley, P., & Sage, S. (1994). Induction of Selective Bayesian Classifiers. In *Uncertainty Proceedings 1994* (pp. 399–406). Elsevier. https://doi.org/10.1016/B978-1-55860-332-5.50055-9

Latrubesse, E. M., & Restrepo, J. D. (2014). Sediment yield along the Andes: Continental budget, regional variations, and comparisons with other basins from orogenic mountain belts. *Geomorphology*, *216*, 225–233. https://doi.org/10.1016/j.geomorph.2014.04.007

Limber, P. W., Adams, P. N., & Murray, A. B. (2017). Modeling large-scale shoreline change caused by complex bathymetry in low-angle wave climates. *Marine Geology*, *383*, 55–64. https://doi.org/10.1016/j.margeo.2016.11.006

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, *20*, 130–141.

Mei, C. C., & Black, J. L. (1969). Scattering of surface waves by rectangular obstacles in waters of finite depth. *Journal of Fluid Mechanics*, *38*(3), 499–511. https://doi.org/10.1017/S0022112069000309

Miller, J. K., & Dean, R. G. (2007). Shoreline variability via empirical orthogonal function analysis: Part I temporal and spatial characteristics. *Coastal Engineering*, *54*(2), 111–131. https://doi.org/10.1016/j.coastaleng.2006.08.013

Montoya, R. D., Osorio Arias, A., Ortiz Royero, J. C., & Ocampo-Torres, F. J. (2013). A wave parameters and directional spectrum analysis for extreme winds. *Ocean Engineering*, *67*, 100–118. https://doi.org/10.1016/j.oceaneng.2013.04.016

Morim, J., Trenham, C., Hemer, M., Wang, X. L., Mori, N., Casas-Prat, M., Semedo, A., Shimura, T., Timmermans, B., Camus, P., Bricheno, L., Mentaschi, L., Dobrynin, M., Feng, Y., & Erikson, L. (2020). A global ensemble of ocean wave climate projections from CMIP5-driven models. *Scientific Data*, *7*(1), 105. https://doi.org/10.1038/s41597-020-0446-2

Mortlock, T. R., & Goodwin, I. D. (2015). Directional wave climate and power variability along the Southeast Australian shelf. *Continental Shelf Research*, *98*, 36–53. https://doi.org/10.1016/j.csr.2015.02.007

Morton, R. A., Gonzalez, J. L., Lopez, G. I., & Correaj, I. D. (2000). Frequent Non-Storm Washover of Barrier Islands, Pacific Coast of Colombia. *Journal of Coastal Research*, *16*, 6.

Murray, A. B. (2007). Reducing model complexity for explanation and prediction. *Geomorphology*, *90*(3–4), 178–191. https://doi.org/10.1016/j.geomorph.2006.10.020

Murray, A. B., Lazarus, E., Ashton, A., Baas, A., Coco, G., Coulthard, T., Fonstad, M., Haff, P., McNamara, D., Paola, C., Pelletier, J., & Reinhardt, L. (2009). Geomorphology, complexity, and the emerging science of the Earth's surface. *Geomorphology*, *103*(3), 496–505. https://doi.org/10.1016/j.geomorph.2008.08.013

Nienhuis, J. H., Ashton, A. D., Edmonds, D. A., Hoitink, A. J. F., Kettner, A. J., Rowland, J. C., & Törnqvist, T. E. (2020). Global-scale human impact on delta morphology has led to net land area gain. *Nature*, *577*(7791), 514–518. https://doi.org/10.1038/s41586-019-1905-9

Nienhuis, J. H., Ashton, A. D., & Giosan, L. (2015). What makes a delta wave-dominated? *Geology*, *43*(6), 511–514. https://doi.org/10.1130/G36518.1

Nienhuis, J. H., Ashton, A. D., & Giosan, L. (2016). Littoral steering of deltaic channels. *Earth and Planetary Science Letters*, *453*, 204–214. https://doi.org/10.1016/j.epsl.2016.08.018

Osorio, A. F., Montoya, R. D., Ortiz, J. C., & Peláez, D. (2016). Construction of synthetic ocean wave series along the Colombian Caribbean Coast: A wave climate analysis. *Applied Ocean Research*, *56*, 119–131. https://doi.org/10.1016/j.apor.2016.01.004

Paniagua-Arroyave, J. F., Adams, P. N., Parra, S. M., & Valle-Levinson, A. (2019). Observations of surface-gravity-wave scattering and dissipation by an isolated shoal related to a cuspate foreland. *Continental Shelf Research*, *173*, 43–55. https://doi.org/10.1016/j.csr.2018.12.004

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572. https://doi.org/10.1080/14786440109462720

Phillips, O. M. (1977). *The Dynamics of the Upper Ocean (Second Edition)*. Syndics of the Cambridge University Press, England.
Poincaré, H. (1893). Les méthodes nouvelles de la mécanique céleste: Méthodes de MM. *Gauthier-Villars It Fils.*, *2*(Newcomb, Glydén, Lindstedt et Bohlin.).
Portilla, J., Caicedo, A. L., Padilla-Hernández, R., & Cavaleri, L. (2015). Spectral wave conditions in the Colombian Pacific Ocean. *Ocean Modelling*, *92*, 149–168. https://doi.org/10.1016/j.ocemod.2015.06.005
Portilla-Yandún, J., Barbariol, F., Benetazzo, A., & Cavaleri, L. (2019). On the statistical analysis of ocean wave directional spectra. *Ocean Engineering*, *189*, 106361. https://doi.org/10.1016/j.oceaneng.2019.106361
Rapún, M.-L., & Vega, J. M. (2010). Reduced order models based on local POD plus Galerkin projection. *Journal of Computational Physics*, *229*(8), 3046–3063. https://doi.org/10.1016/j.jcp.2009.12.029
Restrepo, J. D. (2008). Morphodynamics of a high discharge tropical delta, San Juan River, Pacific coast of Colombia. *Marine Geology*, *25*(1), 1–21. https://doi.org/10.1016/j.jsames.2007.09.002
Restrepo, J. D., & Cantera, J. R. (2013). Discharge diversion in the Patía River delta, the Colombian Pacific: Geomorphic and ecological consequences for mangrove ecosystems. *Journal of South American Earth Sciences*, *46*, 183–198. https://doi.org/10.1016/j.jsames.2011.04.006
Restrepo, J. D., & Kjerfve, B. (2002). The San Juan Delta, Colombia: Tides, circulations, and salt dispersion. *Continental Shelf Research*, *22*(8), 1249–1267. https://doi.org/10.1016/S0278-4343(01)00105-4
Restrepo, J. D., & López, S. A. (2008). Morphodynamics of the Pacific and Caribbean deltas of Colombia, South America. *Journal of South American Earth Sciences*, *25*(1), 1–21. https://doi.org/10.1016/j.jsames.2007.09.002
Rowley, C. W., Colonius, T., & Murray, R. M. (2004). Model reduction for compressible flows using POD and Galerkin projection. *Physica D: Nonlinear Phenomena*, *189*(1–2), 115–129. https://doi.org/10.1016/j.physd.2003.03.001
Rudy, S. H., Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2017). Data-driven discovery of partial differential equations. *Science Advances*, *3*(4), e1602614. https://doi.org/10.1126/sciadv.1602614
Salazar, A., Sanchez, A., Villegas, J. C., Salazar, J. F., Ruiz Carrascal, D., Sitch, S., Restrepo, J. D., Poveda, G., Feeley, K. J., Mercado, L. M., Arias, P. A., Sierra, C. A., Uribe, M. del R., Rendón, A. M., Pérez, J. C., Murray Tortarolo, G., Mercado-Bettin, D., Posada, J. A., Zhuang, Q., & Dukes, J. S. (2018). The ecology of peace: Preparing Colombia for new political and planetary climates. *Frontiers in Ecology and the Environment*, *16*(9), 525–531. https://doi.org/10.1002/fee.1950
Sayama, H. (2015). *Introduction to the Modeling and Analysis of Complex Systems* (1st ed.). Open SUNY Textbooks, Milne Library.
Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, *656*, 5–28. https://doi.org/10.1017/S0022112010001217
Sheremet, A., Davis, J. R., Tian, M., Hanson, J. L., & Hathaway, K. K. (2016). TRIADS: A phase-resolving model for nonlinear shoaling of directional wave spectra. *Ocean Modelling*, *99*, 60–74. https://doi.org/10.1016/j.ocemod.2016.01.002

Syvitski, J., Waters, C. N., Day, J., Milliman, J. D., Summerhayes, C., Steffen, W., Zalasiewicz, J., Cearreta, A., Gałuszka, A., Hajdas, I., Head, M. J., Leinfelder, R., McNeill, J. R., Poirier, C., Rose, N. L., Shotyk, W., Wagreich, M., & Williams, M. (2020). Extraordinary human energy consumption and resultant geological impacts beginning around 1950 CE initiated the proposed Anthropocene Epoch. *Communications Earth & Environment*, *1*(1), 32. https://doi.org/10.1038/s43247-020-00029-y

Thompson, E. F., Chen, H. S., & Hadley, L. L. (1996). Validation of Numerical Model for Wind Waves and Swell in Harbors. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *122*(5), 245–257. https://doi.org/10.1061/(ASCE)0733-950X(1996)122:5(245)

Thomson, R. E., & Emery, W. J. (2014). *Data analysis methods in physical oceanography* (3. ed). Elsevier.

Torrence, C., & Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, *79*(1), 18.

Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L., & Kutz, J. N. (2014). On Dynamic Mode Decomposition: Theory and Applications. *Journal of Computational Dynamics*, *1*(2), 391–421. https://doi.org/10.3934/jcd.2014.1.391

Valiente, N. G., Masselink, G., Scott, T., Conley, D., & McCarroll, R. J. (2019). Role of waves and tides on depth of closure and potential for headland bypassing. *Marine Geology*, *407*, 60–75. https://doi.org/10.1016/j.margeo.2018.10.009

Vega, M. J., Alvarez-Silva, O., Restrepo, J. C., Ortiz, J. C., & Otero, L. J. (2020). Interannual variability of wave climate in the Caribbean Sea. *Ocean Dynamics*, *70*(7), 965–976. https://doi.org/10.1007/s10236-020-01377-1

West, R. C. (1956). MANGROVE SWAMPS OF THE PACIFIC COAST OF COLOMBIA∗. *Annals of the Association of American Geographers*, *46*(1), 98–121. https://doi.org/10.1111/j.1467-8306.1956.tb01498.x

Yule, G. U. (1938). Notes of Karl Pearson's Lectures on the Theory of Statistics, 1884-96. *Biometrika*, *30*(1/2), 198. https://doi.org/10.2307/2332238

Zakharov, V. E., Badulin, S. I., Hwang, P. A., & Caulliez, G. (2015). Universality of sea wave growth and its physical roots. *Journal of Fluid Mechanics*, *780*, 503–535. https://doi.org/10.1017/jfm.2015.468

Zakharov, V. E., Resio, D., & Pushkarev, A. (2017). Balanced source terms for wave generation within the Hasselmann equation. *Nonlinear Processes in Geophysics*, *24*(4), 581–597. https://doi.org/10.5194/npg-24-581-2017

Zijlema, M., van Vledder, G. Ph., & Holthuijsen, L. H. (2012). Bottom friction and wind drag for wave models. *Coastal Engineering*, *65*, 19–26. https://doi.org/10.1016/j.coastaleng.2012.03.002

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, *15*(2), 265–286. https://doi.org/10.1198/106186006X113430